

Andreas Oberweis ■ Christof Weinhardt
Henner Gimpel ■ Agnes Koschmider
Victor Pankratius ■ Björn Schnizler (Hrsg.)

eOrganisation: Service-, Prozess-, Market-Engineering 8. Internationale Tagung Wirtschaftsinformatik

Band 1



universitätsverlag karlsruhe

Andreas Oberweis / Christof Weinhardt / Henner Gimpel /
Agnes Koschmider / Victor Pankratius / Björn Schnizler (Hrsg.)

eOrganisation: Service-, Prozess-, Market-Engineering

8. Internationale Tagung Wirtschaftsinformatik

Band 1

eOrganisation: Service-, Prozess-, Market-Engineering

8. Internationale Tagung Wirtschaftsinformatik
Karlsruhe, 28. Februar – 2. März 2007

Band 1

Andreas Oberweis
Christof Weinhardt
Henner Gimpel
Agnes Koschmider
Victor Pankratius
Björn Schnizler
(Hrsg.)



universitätsverlag karlsruhe

Impressum

Universitätsverlag Karlsruhe
c/o Universitätsbibliothek
Straße am Forum 2
D-76131 Karlsruhe
www.uvka.de



Dieses Werk ist unter folgender Creative Commons-Lizenz
lizenziert: <http://creativecommons.org/licenses/by-nc-nd/2.0/de/>

Universitätsverlag Karlsruhe 2007
Print on Demand

ISBN: 978-3-86644-094-4 (Band 1)
ISBN: 978-3-86644-095-1 (Band 2)
ISBN: 978-3-86644-093-7 (Set)

Vorwort

Wirtschaft, Gesellschaft und öffentlicher Sektor müssen dem immer rascheren Globalisierungsprozess folgen, um wettbewerbsfähig zu bleiben. Eine immer feinere und großflächigere Aufgabenteilung stellt wachsende Anforderungen an die Koordinationsfähigkeit der entstehenden wirtschaftlichen und sozialen Netze.

eOrganisationen erlauben die Einbindung technischer Systeme und menschlicher Entscheidungsträger und erhöhen so die Anpassungsdynamik und die Qualität der Entscheidungen. Sie umfassen technische und menschliche Akteure sowie robuste Organisations- und Koordinationsregeln, die auf die Erreichung gemeinsamer Ziele ausgerichtet sind. Sie stellen vielschichtige, interdisziplinäre und neuartige Herausforderungen an die Wirtschaftsinformatik und die benachbarten Disziplinen: Es werden skalierbare und zuverlässige IuK-Systeme sowie Standards und Infrastrukturen benötigt, die den regulativen Rahmenbedingungen genügen und gleichzeitig genügend Freiraum bei ihrer Gestaltung lassen. Darauf aufbauend sind adaptive Services und Prozesse mit hohen Freiheitsgraden im Hinblick auf ihre Anwendungsmöglichkeiten erforderlich. Neben innovativen Geschäftsmodellen müssen Anreizstrukturen und dezentrale Organisationsformen, wie etwa Märkte, gestaltet und über alle Ebenen einer Anwendung durchgängig implementiert werden.

Die 8. Internationale Tagung Wirtschaftsinformatik 2007 (WI2007), die vom 28. Februar - 2. März 2007 in Karlsruhe stattfindet, greift mit dem Motto "eOrganisation: Service-, Prozess-, Market-Engineering" das weite Spektrum der Aufgaben auf, um es anhand verschiedenster Anwendungsbereiche und technischer Fragestellungen zu diskutieren. Aus 316 eingereichten Beiträgen wurden 106 in einem doppelt-blinden Begutachtungsverfahren ausgewählt und 31 Tracks aus den Bereichen

- Service-Engineering,
- Prozess-Engineering und
- Market-Engineering

zugeordnet. Zusätzlich wird das Programm durch eingeladene Vorträge herausragender Persönlichkeiten sowie Tutorien und ein Studierendenprogramm bereichert.

Für das Gelingen der Tagung möchten wir uns bei den zahlreichen beteiligten Personen und Institutionen bedanken, die uns mit Rat und Tat unterstützt haben. Bei den Autoren möchten wir uns für die hervorragenden Beiträge bedanken, die das Fachgebiet weiter voran bringen. An die Track-Chairs sowie die über 250 Gutachter, die uns durch ihre fachkundige Beurteilung der eingereichten Beiträge unterstützt haben, geht unser herzlicher Dank.

Vor allem bedanken wir uns auch bei den Sponsoren, die mit ihrer großzügigen finanziellen Unterstützung die Durchführung der Tagung ermöglicht haben. Unserem Tagungsbeirat, Prof. Dr. Dr. h. c. Lockemann und Prof. Dr. Wolffried Stucky danken wir für die Unterstützung und die wertvollen Anmerkungen während der Tagungsvorbereitung. Schließlich danken wir den beteiligten Mitarbeiterinnen, Mitarbeitern und studentischen Hilfskräften am Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB) sowie am Institut für Informationswirtschaft und -management (IISM), namentlich stellvertretend für alle Stefanie Betz, Maik Herfurth, Christian Herlambang und Anke Plog. Frau Barbara Stolzer danken wir für die grafische Gestaltung der gesamten Tagungsunterlagen.

Daneben möchten wir die Gelegenheit nutzen, auch allen hier nicht erwähnten Personen unseren Dank auszusprechen, die uns vor und während der Tagung mit Tatkraft unterstützt haben.

Bleibt uns, den Lesern eine gewinnbringende Lektüre zu wünschen....

Karlsruhe, im Dezember 2006

Andreas Oberweis, Christof Weinhardt, Henner Gimpel, Agnes Koschmider, Victor Pankratius, Björn Schnizler

Inhaltsverzeichnis

Hauptvorträge

<i>Keeping CEOs Awake At Night. Innovation auf dem globalen Markt</i> Martin Jetter	3
<i>Serious Games: The Impact of Pervasive Gaming on Business</i> Matthias Jarke	15
<i>Dezentrales Wissensmanagement bei der EnBW Energie Baden-Württemberg AG</i> Bernhardt Beck	17
<i>Flexible Prozessgestaltung als Basis innovativer Geschäftsmodelle - Von der Service-Orientierten Architektur zur Vision des Business Webs</i> Lutz Heuser, Stefan Lacher, Simone Perlmann	19
<i>Search Engine Economics</i> Hal R. Varian	29
<i>Decentralised Control of Complex Systems</i> Nicholas R. Jennings	31

Service-Engineering

Das Internet der Dinge: Betriebliche Anwendungen von RFID und Ubiquitous Computing

Track-Chairs: Elgar Fleisch, Karl Kurbel, Franz Lehner, Orestis Terzidis

<i>Start a grassroots RFID initiative! The relevance of communication and showcases on the success of RFID</i> Florian Resatsch, Joerg Aßmann, Thomas Schildhauer, Daniel Michelis	37
<i>Perceived Usefulness of RFID-enabled Information Services – A Systematic Approach</i> Hanna Krasnova, Matthias Rothensee, Sarah Spiekermann	55

Zum Einsatz von RFID in der Filiallogistik eines Einzelhändlers: Ergebnisse einer Simulationsstudie 71

Frédéric Thiesse, Elgar Fleisch

Wahrgenommene strategische Bedeutung von RFID aus Sicht von IT-Entscheidern in Deutschland - Eine empirische Analyse 89

Uta Franziska Knebel, Jan Marco Leimeister, Helmut Krcmar

eGovernment - effektiv, wirtschaftlich, transparent

Track-Chairs: Jörg Becker, Maria A. Wimmer, Willi Kaczorowski, Helmut Krcmar

Model-Driven Integration Engineering in der E-Government-Domäne Meldewesen 109

Stefan Kühne, Maik Thränert, Werner Rotzoll, Jan Lehmann

Support of Harmonisation of Public Processes - Through Modelling of Legal Constraints 127

Sebastian Olbrich, Paul Alpar

Prozessorientierte Gestaltung von Behördenkontakten - Untersuchung zu eGovernment-Anforderungen aus Unternehmenssicht 145

Petra Wolf, Helmut Krcmar

Zahlungsbereitschaft für elektronische Signaturen 163

Heiko Roßnagel, Oliver Hinz

Cross-organizational processes in Public Administrations: Conceptual modeling and implementation with Web Service Protocols 181

Jörg Ziemann, Timo Kahl, Thomas Matheis

Information Services in der Logistik

Track-Chairs: Christian Bierwirth, Kai Furmans, Herbert Kopfer, Dieter Pütz

ComEx: Kombinatorische Auktionen zum innerbetrieblichen Austausch von Logistikdienstleistungen 201

Oleg Gujo, Michael Schwind, Jens Vykoukal, Kilian Weiß, Tim Stockheim, Oliver Wendt

Prozessverantwortung und -dokumentation als Determinanten der Effizienz und Qualität der Transportplanung - Ein Modell auf Basis einer empirischen Untersuchung unter 1000 deutschen Unternehmen 219
Volker Lanninger, Tim Stockheim, Oliver Wendt

Mobile Business and Communications

Track-Chairs: J. Felix Hampe, Kai Rannenberg, Klaus Turowski

Privatsphärenfreundliche topozentrische Dienste unter Berücksichtigung rechtlicher, technischer und wirtschaftlicher Restriktionen 237
Mike Radmacher, Jan Zibuschka, Tobias Scherner, Lothar Fritsch, Kai Rannenberg

Allokation und Bepreisung kontextbezogener Kundenkontakte im Mobile Marketing - Rahmenbedingungen für Einsatz und Gestaltung von multi-attributiven Auktionen 255
Andreas Albers

Adoption and Impact of Mobile-Integrated Business Processes – Comparison of Existing Frameworks and Analysis of their Generalization Potential 273
Key Pousttchi, Bettina Thurnher

Critical Success Factors for Mobile Field Service Applications: A Case Research 291
Andreas Birkhofer, Sina Deibert, Franz Rothlauf

Migration als Ansatz zur Gestaltung mobiler Services 309
Oliver Bohl, Shakib Manouchehri, Udo Winand

Automotive Service Engineering - Systematische Entwicklung personalisierbarer und interaktiver mobiler Dienste für den Automobilsektor 327
Holger Hoffmann, Jan Marco Leimeister, Helmut Krcmar

Outsourcing und IT-Governance

Track-Chairs: Armin Heinzl, Michael Heym, Gerhard F. Knolmayer

Eine Typologie von Beziehungen im IT-Outsourcing: Ein konzeptioneller Ansatz 347
Stefanie Jahner, Tilo Böhmann, Helmut Krcmar

*Qualität von IT-Leistungen aus den Perspektiven von Anbietern und Nachfragern -
Ergebnisse einer Umfrage in der Schweiz* 365
Roger Gruetter, Gerhard Schwabe, Felix-Robinson Aschoff

*The Interplay of Outsourcing Risks and Benefits - A Study of Business Process
Outsourcing in the German Banking Industry* 383
Kim Wuellenweber

Product Life Cycle Management in Unternehmenssoftwaresystemen

Track-Chairs: Stefan Kirn, Peter Loos, C. Pohl

*Hybride Wertschöpfung im Maschinen- und Anlagenbau - Prozessorientierte
Integration von Produktentwicklung und Servicedokumentation zur Unterstützung
des technischen Kundendienstes* 403
**Oliver Thomas, Philipp Walter, Peter Loos, Michael Schlicker,
Markus Nüttgens**

Semantic Web

Track-Chairs: Jürgen Angele, Dieter Fensel, Rudi Studer

An Ontology Framework for Semantic Business Process Management 423
Martin Hepp, Dumitru Roman

Automated Selection of Configurable Web Services 441
Steffen Lamparter, Anupriya Ankolekar

Service Oriented Computing

Track-Chairs: Torsten Eymann, Alfred Geiger, Manfred Grauer

Negotiations in Service-Oriented Architectures 461
Falk Kretzschmar

*Evaluation of Economic Resource Allocation in Application Layer Networks –
A Metrics Framework* 477
**Werner Streitberger, Torsten Eymann, Daniel Veit, Michele Catalano,
Gianfranco Giuliani, Liviu Joita, Omer F. Rana**

<i>Service-Oriented Computing for Risk/Return Management</i>	495
Wolfgang Hackenbroch, Matthias Henneberger	
<i>Native Code Security for Grid Services</i>	513
Thomas Friese, Matthew Smith, Bernd Freisleben	
<i>A Service-Oriented Grid-Based Infrastructure for Supporting Virtual Prototyping in Manufacturing</i>	531
Manfred Grauer, Julian Reichwald, Thomas Barth	
<i>Entwurf eines Enterprise Architecture Frameworks für Serviceorientierte Architekturen - Betrachtung am Beispiel einer erweiterten UN/CEFACT Modeling Methodology</i>	549
Philipp Offermann, Christian Schröpfer, Marten Schönherr, Maximilian Ahrens	

Prozess-Engineering

Business Intelligence

Track-Chairs: Peter Chamoni, Bodo Rieger

<i>Kollaboratives Data Warehousing - Konzeption und prototypische Realisierung flexibler Schema- und Datenintegration</i>	569
Thomas Matheis, Dirk Werth, Peter Loos	
<i>Initialisierung des Dispositionssystems neu eröffneter Einzelhandelsmärkte mit Methoden des Data-Mining</i>	587
Jochen Frank, Michael Kohl, Otto K. Ferstl, Thorsten Ruffer	
<i>Konzeption des Unternehmensreportings - Ein modellgestütztes Vorgehensmodell zur fachkonzeptionellen Spezifikation</i>	605
Jörg Becker, Stefan Seidel, Christian Janiesch	
<i>Designing Open Source Business Intelligence for Public Administrations</i>	623
Björn Niehaves, Felix Müller-Wienbergen	
<i>Move-to-the-User? Eine Analyse der verlagernden Wirkung von Business Intelligence im Controlling</i>	641
Michael Samtleben, Thomas Hess	

Business Process Engineering

Track-Chairs: Dimitris Karagiannis, Peter Küng, Markus Nüttgens

Supporting Inter-Business Collaboration via Contract Negotiation and Enactment 661
Peter Rittgen

Integration of Conceptual Process Models by the Example of Event-driven Process Chains 677
Carlo Simon, Jan Mendling

Realisierungsformen des Geschäftsprozessmanagements - Eine explorative Klassifikationsanalyse 695
Tobias Bucher, Robert Winter

Individualisierung von Prozessen und E-Services mithilfe von Case Based Reasoning 713
Günter Schicker, Carolin Kaiser, Freimut Bodendorf

Plädoyer für die Entwicklung perspektivenspezifischer Problemlösungskomponenten zur Unterstützung der Prozessverbesserung 731
Ralf Knackstedt, Martin Pellengahr

Collaborative Business

Track-Chairs: Oliver Frick, Michael Koch, Gerhard Schwabe, Volker Wulf

Präsenzbasierte Echtzeitkommunikation - Eine prototypbasierte Untersuchung der Nutzbarkeit im Unternehmensberatungskontext 751
Kai Riemer

Audio vs. Chat bei Aufgaben mit Unsicherheit: Die Produktivität folgt anderen Regeln als bei mehrdeutigen Aufgaben 769
Andreas Löber, Gerhard Schwabe

Diskontinuierliche Innovationen fördern - Die Rolle von Idea Mirrors zur Unterstützung von Innovation und Kooperation im Unternehmen 787
Michael Koch, Kathrin Möslein

Conceiving an environment for managing the lifecycle of collaborative business processes 805
Dirk Werth, Philipp Walter, Peter Loos

Infrastrukturen zur Aneignungsunterstützung - Ein Konzept zur Integration von produkt- und prozessorientierter Flexibilisierung 823
Gunnar Stevens, Volker Wulf, Volkmar Pipek

Individual and Human-Assisted Computer Self Efficacy: An Empirical Examination 841
Jason Bennett Thatcher, Michael J. Gundlach, D. Harrison McKnight, Mark Srite

Customer und Supplier Relationship Management

Track-Chairs: Freimut Bodendorf, Andreas Geyer-Schulz, Roland Holten

Optimierung des Customer Lifetime Value auf Basis einzelvertraglicher Entscheidungen - Ein dynamisches Optimierungsmodell 861
Ulrich Faisst, Martin Gneiser, Nina Kreyer, Nina Schroeder

Spezifizierung des Kano-Modells zur Messung von Kundenzufriedenheit 879
Hans Ulrich Buhl, Dennis Kundisch, Annette Renz, Nicola Schackmann

eHealth / Informationsmanagement im Gesundheitswesen

Track-Chairs: Hans Czap, Jörg Haas, Detlef Ruland

Elektronische Intermediation im Gesundheitswesen - Das Beispiel Schweiz 899
Konrad Walser, Thomas Myrach

Process-based Performance Measurement in Healthcare Networks 917
Günter Schicker, Jörg Purucker, Freimut Bodendorf

Zielgruppenspezifische Dienste für Virtuelle Patientengemeinschaften 935
Achim Dannecker, Ulrike Lechner

Formalisierung und Automatisierung von SOPs in der Intensivmedizin 953
Martin Sedlmayr, Thomas Rose, Rainer Röhrig, Markus Meister, Achim Michel-Backofen

Hauptvorträge

Keeping CEOs Awake At Night

Innovation auf dem globalen Markt

Martin Jetter

Vorsitzender der Geschäftsführung
IBM Deutschland GmbH
Stuttgart
mjetter@de.ibm.com

Abstract

Die wirtschaftliche Makrodynamik der Globalisierung zwingt CEOs zur radikalen Änderung, wenn nicht gar Neuerfindung ihrer Unternehmen. Zunehmender Wettbewerbsdruck im Inland einerseits und enorme Chancen in den Wachstumsmärkten andererseits bestimmen dabei den Handlungsrahmen. Dank zunehmend vernetzter und leistungsfähiger IT-Infrastruktur migriert Arbeit an den Ort ihrer optimalen Erledigung. Wohin, das regeln die Marktfaktoren Wirtschaftlichkeit, verfügbarer Erfahrungsschatz und Offenheit.

Der Schlüssel für die Nutzung der aus der Globalisierung resultierenden Chancen ist die konsequente Ausrichtung auf Innovation. Echte Innovation geschieht am Schnittpunkt von Erfindung und der Erfahrung, wie und wo sie in Differenzierung im Markt und wirtschaftlichen Erfolg umgesetzt werden kann. Dabei kooperieren mehr und mehr CEOs mit Dritten, mit Partnern, Zulieferern und Kunden. Nur in dieser Besinnung auf gemeinsame Stärken liegt eine Chance für den Wirtschaftsstandort Deutschland.

1 The World Is Flat

1.1 Wirtschaftliche Makrodynamik

Verschaffen wir uns zu Beginn einen Überblick aus größerer Höhe über die wirtschaftliche Makrodynamik auf unserem Planeten. Kaum ein wirtschaftlicher Vortrag, kaum ein Wirtschaftsbuch, kaum ein Zeitungsbeitrag kommt ohne diese Betrachtung aus und am Stichwort Globalisierung vorbei. Und das aus gutem Grund: Kein Phänomen greift so tiefgehend in unser Leben und Wirtschaften ein, fordert uns so sehr heraus und verlangt so viel Anpassung wie die Globalisierung unserer Ökonomie.

Zwei Drittel der von IBM befragten 765 CEOs weltweit planen, ihr Unternehmen in den kommenden zwei Jahren radikal umzubauen, es neu auszurichten, seine Strategie zu justieren oder gänzlich neu zu erfinden. Solche massiven Transformationen verlangen Unternehmen und ihren Chefs einiges ab.

Vorbei sind die Zeiten reinen Kostensparens, das in den letzten Jahren die CIOs und CTOs in den Unternehmen sehr – zu sehr – in die Kostenspar-Ecke drängte. Das zeigte sich beispielhaft an der Bedeutung, die der Informationstechnologie im Unternehmen zugemessen wurde: Ihr strategischer Wertbeitrag zum Erfolg des Unternehmens rückte zusehends in den Hintergrund.

IT als Innovationsmotor und Werttreiber genoss oftmals nicht mehr den Stellenwert, der ihr gebührte. Die Stimmen der Zweifler waren laut und gipfelten im vieldiskutierten und zitierten Beitrag von Nick Carr in Harvard Business Review: ‚IT doesn’t matter’ – er behauptete darin, dass IT zur Massenware, zur Commodity geworden sei, aus der sich mithin kein Wettbewerbsvorteil mehr ableiten lasse. Sorry, Nick, you’re wrong. IT does matter!

Denn: Niemals war oder ist Informationstechnologie in ihrer Gesamtheit Massenware. Stets gibt es bessere und schlechtere Lösungen; attraktive, außergewöhnliche, innovative – und solche, die zum bloßen Mindeststandard gehören. Entscheidend ist, heraus zu finden, was Unternehmen tatsächlich weiterbringt und welche neuen, kreativen Ansätze sie verfolgen können.

Kreative Ansätze gibt es auf allen Ebenen, nicht nur im engeren Sinne bei der Technologie. Denken wir über Produkte und Technologien hinaus an Geschäftsprozesse, an neue Formen der Kooperation entlang der Wertschöpfungskette und an neue Nutzenerwartungen der Kunden.

IT ist *der* Innovationstreiber im Unternehmen und wächst gerade jetzt, im Zuge immer rasanter fortschreitender Globalisierung, zu beträchtlicher Bedeutung heran. Innovation ist nämlich – auch das belegt die IBM CEO Studie eindrucksvoll – *die* Antwort auf die Herausforderungen einer globalisierten Welt.

1.2 Herausforderungen in der Flat World

Diese Herausforderungen erläutern nur wenige so eindrucksvoll und augenöffnend wie Tom Friedman, New York Times Kolumnist und Autor des sehr lesenswerten Buches ‚The World Is Flat‘. Friedman beschreibt darin die evolutionären und revolutionären Änderungen unserer Welt und Wirtschaft in den vergangenen beiden Dekaden, die zumindest ökonomisch die Erde zur Scheibe werden ließen.

Manche dieser „Flattener“, wie Friedman sie nennt, sind vorwiegend politischer Natur, wie der Fall der Mauer, der die Machtbalance auf unserem Planeten zugunsten von Demokratie und Freiheit verschob und damit mehr Menschen Zugang zu Information, Waren und Dienstleistungen gab. Andere sind eher wirtschaftliche Ereignisse, wie der Netscape Börsengang, der massive Investitionen in die Breitbandverkabelung der Erde motivierte – oder das Offshoring von industrieller Fertigung, das China auf der wirtschaftlichen Weltbühne dramatisch wichtiger werden ließ.

Wie dramatisch, zeigen einige Zahlen: Zwischen 1970 und 2005 hat sich das Volumen der Weltwirtschaft etwa vervierfacht, von rund 12 auf 45 Billionen Dollar. Gleiches gilt für die Zahl der Arbeitskräfte in Indien (von 62 auf 220 Millionen) und China (von 86 auf 412 Millionen). In allen übrigen Industrieländern zusammen hat sich die Zahl der Arbeitskräfte im gleichen Zeitraum nur um 40 Prozent erhöht und liegt heute mit 416 Millionen etwa gleichauf mit den chinesischen Kollegen. Entscheidend ist: die Arbeitsmärkte waren 1970 noch weitgehend isoliert. Heute stehen sie in direkter Konkurrenz zueinander.

Denn: Die meisten Flattener sind im Kern und in ihren Ausprägungen informationstechnologisch getrieben: Workflow Software, Open Source, Wireless, nicht zuletzt das Internet – diese Technologien ermöglichen die vernetzte Wirtschaft, in der wir heute auf Augenhöhe mit Volkswirtschaften konkurrieren, die wir bisher entweder gar nicht oder bestenfalls als Güterexport-Adresse wahrgenommen haben.

Längst hat diese Vernetzung und Kooperation alle Bereiche des Wirtschaftens erreicht und ergriffen. Indische Service Companies haben nicht mehr nur „Brückenköpfe“ hier, wie wir uns in vergeblicher Selbstberuhigung in den letzten Jahren immer wieder suggeriert haben.

Sie treten massiv in unseren Markt ein und sie nutzen längst ihrerseits attraktive Economies, indem sie Teile ihrer Leistung zum Beispiel in Osteuropa sourcen. Das müssen sie übrigens tun. Mitte 2006 hat die Economist Intelligence Unit eine Studie zum Offshore Outsourcing veröffentlicht, die konstatiert: Zentral- und Osteuropa sowie Nordafrika sind insbesondere für Unternehmen in Europa die kommenden Lokationen für Outsourcing Center. Herausgehoben werden Länder wie Rumänien, Ägypten und Jordanien.

Indien dagegen wird nämlich allmählich teurer – selbst in Millionenstädten wie Bangalore und Chennai zeigt sich inzwischen ein Mangel an gut ausgebildetem Personal.

1.3 Partizipation an Wachstumschancen

IBM hat sich frühzeitig – schon vor mehr als fünfzig Jahren – hier engagiert und ist seit über fünfzehn Jahren in China präsent. Jüngst erfolgte ein rasanter Ausbau. So ist Indien mit rund 48.000 Mitarbeitern inzwischen die zweitgrößte IBM Landesgesellschaft außerhalb der USA. Wir investieren in den nächsten drei Jahren weitere 6 Milliarden Dollar in den Ausbau.

Das sind übrigens keine schlicht opportunistischen oder singulären Engagements, sondern Teil planvollen weltweiten Vorgehens: Die IBM integriert sich selbst global. Das wurde in der vergangenen Dekade zunächst mit der weltweiten Supply Chain umgesetzt – heute ist sie durchgängig – und setzt sich zur Zeit mit dem Servicegeschäft und auch den internen Supportfunktionen fort.

Langfristig betrachtet ist das nur konsequent: Waren wir in den 60er und 70er Jahren noch ein typisches „internationales“ Unternehmen mit Kernaufgaben wie Entwicklung und Produktion im Heimatland und Vertriebseinheiten im Ausland, so haben wir uns in den 80er und 90er Jahren zu einem „Multinational“ gewandelt. In den meisten unserer weltweit 170 Märkte haben wir Mini-IBMs aufgebaut – voll funktionsfähige Unternehmen, die sich hervorragend in die jeweilige lokale Wirtschaft einbrachten, in aller Regel von Managern aus dem jeweiligen Land geführt wurden und prächtig gediehen.

Was damals richtig war, entpuppte sich um die Jahrtausendwende zunehmend als redundant und inflexibel. Schnellen Reaktionen auf Marktveränderungen und Innovationen steht eine solche Struktur eher im Wege. Deshalb entwickeln wir uns heute zum „global integrierten Unternehmen“ weiter, das seine Ressourcen und Assets weltweit optimiert und managed.

So wie der Grad der Vernetzung wächst, alles mit allem verbunden ist, so wandert nun Arbeit an den Ort ihrer optimalen Erledigung – wie Wasser, das sich seinen Weg sucht. Was bestimmt diesen Weg des Wassers? Wohin fließt Arbeit und warum?

1.4 Wirtschaftlich, erfahren, offen

Wir beobachten drei Gesetze: Erstens ein ökonomisches. Den Regeln von Kosten und Profit kann sich niemand widersetzen. Wenn in Indien und China die Arbeitskosten um den Faktor zehn unter unseren liegen, haben wir einen Wettbewerbsnachteil. Den können wir nur mit den beiden anderen Gesetzen ausgleichen:

Zweitens gilt das Gesetz der Expertise: Die besten Köpfe veranlassen Arbeit zu migrieren – das heißt, die Arbeit folgt den besten Köpfen und nicht anders herum, wie oft angenommen wird. Die Qualität muss stimmen. Oder, um es wie unser Bundespräsident vor dem Europäischen Parlament zu sagen: „Wir müssen mindestens so viel besser sein, wie wir teurer sind.“

Drittens gilt das Gesetz der Offenheit: Je bereitwilliger Wissen und Information mit anderen geteilt werden, desto besser ist das Gesamtergebnis der Arbeit. Clustertheorie und Clusterpraxis liefern dafür zahlreiche Belege.

Verweilen wir noch einen Augenblick bei den Markttrends der Globalisierung und zeigen diese nochmals exemplarisch an der Informationstechnologie:

Global betrachtet wächst der Trend zum IT Offshore Outsourcing ungebrochen weiter. Der Weltmarkt in diesem Geschäft wird vom Economist auf 40 bis 50 Milliarden Dollar geschätzt. Jährliche Steigerungsraten von 30% werden ihn bis 2008 auf rund 100 Milliarden Dollar wachsen lassen.

Allein dieses Geschäft wächst damit zu einer Größe heran, die unseren gesamten deutschen IT-Markt deutlich übertrifft. Der lag 2005 laut Branchenverband BITKOM bei gut 68 Milliarden Euro Umsatz, verzeichnete in den vergangenen Jahren Wachstumsraten unter drei Prozent und bleibt auch 2006 und 2007 hinter den Erwartungen zurück.

1.5 Der Faktor Arbeit

Vorwiegend Seitwärtsbewegungen kennzeichnen die Beschäftigungssituation der rund eine Million Arbeitsplätze umfassenden IT-Branche. Deren Fortbestand und Ausbau ist keine simple Preisfrage. Ähnlich wie im globalen Wettbewerb bei den Arbeitskosten steigt auch in der Frage der Leistungsqualität der Wettbewerbsdruck stetig. Umso wichtiger sind die Aktualität und das Qualitätsniveau unserer verfügbaren Skills. Hier dürfen wir keine Kompromisse machen. Kontinuierliche Erneuerung und Erweiterung unserer Skills in einer offenen und kooperativen Umgebung ist die einzige Chance, die wir zur Verbesserung unserer ökonomischen Aussichten haben.

Tom Friedman hat die Folgen für den Faktor Arbeit in den reiferen Märkten in einem Interview so verdeutlicht: „When I was growing up, my parents told me, "Finish your dinner. People in China and India are starving." I tell my daughters, "Finish your homework. People in India and China are starving for your job."

Bis 2004 hatten die USA bereits rund 700.000 sogenannte White-Collar-Jobs an Indien verloren. Allein 2005 sind nochmals 250.000 Jobs dazugekommen. Ich rede hier gleichwohl nicht schlichtem Arbeitsplatz-Pessimismus das Wort.

Betrachtet man Zahlen des U.S. Bureau of Labor Statistics und des Weltwährungsfonds, dann stellt man fest: Viele reife Märkte importieren mehr Arbeit – gemessen an ihrem finanziellen Gegenwert – als sie exportieren. Sie sind Netto-Profiteure des Global Sourcing.

Großbritannien zum Beispiel hat 2004 Arbeit im Gegenwert von 19 Milliarden US-Dollar exportiert, im selben Jahr aber für ausländische Unternehmen Arbeit im Wert von 41 Milliarden US-Dollar - rund doppelt so viel - innerhalb des eigenen Territoriums erledigt. Den selben Trend zeigen die USA. Job-Exporten von gut 43 Milliarden Dollar stehen Importe von 61 Milliarden Dollar gegenüber. Weniger rosig sieht es allerdings bisher für Deutschland aus: 41 Milliarden Dollar-Job-Export, 26 Milliarden Dollar Job-Import. Andere Netto-Verlierer sind Österreich und Japan.

Welche Schlüsse lassen sich daraus ziehen? Über die Gründe für unterschiedlichen volkswirtschaftlichen Insourcing-Erfolg kann man trefflich spekulieren und gar streiten. Standortfaktoren aller Art spielen dabei eine Rolle: Der Fokus auf Bildung, Innovationsfähigkeit, Investitionsverhalten, Flexibilität des Arbeitsmarktes sowie politische und wirtschaftliche Stabilität ist sicher konsensfähig. Und nicht in allen Bereichen gehört Deutschland auf Anhieb zu den Top-Scorern.

2 Globale Integration und Innovation

2.1 Chancen der globalen Integration

Fest steht: Globale Integration veranlasst Unternehmen, sich neu zu erfinden. Sie integrieren spezialisierte Funktionen – Entwicklung, Produktion, Einkauf, Vertrieb, Personal – und treffen dabei flexible Make-or-Buy-Entscheidungen: Einkaufsprozesse globaler Unternehmen werden in Manila abgewickelt, Investment-Bank-Transaktionen in Dublin. Britische Radiologen senden Röntgenbilder zur Interpretation nach Australien und Customer Service Center in Neuschottland bearbeiten Garantieforderungen von Verbrauchern aus der ganzen Welt.

In der globalen Integration und der Zusammenarbeit mit Einheiten innerhalb und außerhalb des eigenen Unternehmens liegen mehr Chancen als Risiken. Es ist allerdings ein Geschäftsmodell,

dem man sich konsequent und rechtzeitig verschreiben muss. Dann muss man nicht mehr befürchten, links und rechts überholt zu werden.

Auch Tom Friedman sieht zahlreiche Chancen, die er ebenso wie die „Flattener“ ganz wesentlich der Informationstechnologie zuschreibt. In seiner New York Times Kolumne schrieb er unlängst: „We are seeing the emergence of collaborative business models that were simply unimaginable a decade ago. Today, there are so many more tools, so many more ideas, so many more people able to put these ideas and tools together to discover new things, and so much better communications to disseminate these new ideas across the globe.”

Wir stehen gemeinsam in der Verantwortung, die Entwicklung des global integrierten Wirtschaftens voran zu treiben: Wir brauchen den Brückenschlag zwischen Wissenschaft und Praxis. Die flexible Kombination, Kooperation und Integration von Funktionen und Skills verlangt ebenso flexible und neuartige IT-Systeme, die den Dialog der Unternehmenseinheiten im weltweiten Verbund stützen und steuern. Die Rolle der IT-Verantwortlichen ist deshalb wieder stärker investiv ausgerichtet, mehr darauf, Innovation voranzutreiben und das Unternehmen weiterzuentwickeln.

2.2 Innovation richtig verstanden

Der Begriff Innovation ist in aller Munde und das aus gutem Grund. Mancher mag ihn schon nicht mehr hören, so inflationär wird er benutzt und instrumentalisiert. Tatsächlich gilt: Innovation ist die Antwort auf globale Herausforderungen.

Die Frage ist: Wie versteht man Innovation? Innovation entsteht am Kreuzungspunkt von Erfindergeist und dem Wissen darum, wie und wo kluge Erfindungen angewendet werden. Erfindungen sind gut und notwendig, aber wertvoll werden sie erst dann, wenn sie sich kapitalisieren und in handfeste Wettbewerbsvorteile ummünzen lassen. Diese entstehen, wenn Technologien auf einzigartige Weise für den Kunden oder die Allgemeinheit einen Nutzen erbringen. Das muss sich ausdrücken in Umsatzwachstum, höheren Gewinnen, mehr Produktivität, Differenzierungsmerkmalen, Erschließung neuer Märkte, besseren Sicherheitsstandards, einer verbesserten Volksgesundheit und so weiter.

Im Zeitalter der Globalisierung kann Technologie nicht gehortet werden. Sie steht der ganzen Welt zur Verfügung oder - für eine begrenzte Zeit - zumindest demjenigen, der ein Patent darauf angemeldet hat. Wenn nur allein die Anzahl der Patentanmeldungen automatisch auf die Innovationsfähigkeit eines Landes schließen ließe, hätte der Standort Deutschland noch relativ gute Karten. Wurden doch im letzten Jahr insgesamt fast 50.000 deutsche Patente beim Deutschen und Europäischen Patentamt registriert. Und im EU-Innovationsindex des Statistischen Landesamts Baden-Württemberg führt das südwestdeutsche Bundesland vor Berlin, der Ile de France, Schweden und Bayern. Sechs der zehn ersten Positionen in diesem Ranking werden von deutschen Bundesländern belegt.

Mit Nordamerikanern und Japanern können hiesige Entwickler gleichwohl nicht mithalten. Vor allem im Bereich Mikroelektronik, Großcomputer und Biotechnologie ist die Differenz groß. Stark sind die deutschen Erfinder dagegen bei der Kraftfahrzeugtechnik und in manchen Sparten des Maschinenbaus.

Aber das Patentieren hervorragender Technologie allein reicht nicht aus. Es kommt darauf an, sie zu kapitalisieren. In Cent und Euro, in Produktivität und Arbeitsplätze. Und da haben wir in Deutschland deutlichen Nachholbedarf. Wenn wir im globalen Wettbewerb bestehen und eine führende Rolle einnehmen wollen, müssen wir mehr aus unseren Technologie-Ressourcen machen.

2.3 Worüber spricht die Welt?

Technologie revolutioniert Industrien oder degradiert sie zur Bedeutungslosigkeit. Was heute „in“ ist, muss es morgen schon lange nicht mehr sein. Der Walkman oder Bestellungen per Post? Niemand spricht mehr davon. Dafür um so mehr von Apples iPod oder Amazons Geschäftsmodell. Oder auch darüber schon nicht mehr? Statt dessen lieber von Web 2.0, Web 3.0, virtuellen Welten und realen Geschäften darin?

Warum besucht der IBM CEO, Sam Palmisano, in Gestalt seines Avatars die Onlinewelt SecondLife? Eine nette Spielerei? Wohl kaum – und es macht deshalb zu recht Schlagzeilen! Wie bereits Medienunternehmen, Konsumgüterhersteller, Eventveranstalter, Klein- und

Kleinstunternehmer experimentieren und investiert IBM in 3D-Welten und entwickelt innovative Zugänge zu neuen Märkten.

Die Zeit ist günstig für Innovatoren. Viele Technologien sind ausgereift. Datenverarbeitungskomponenten werden stetig kleiner, leistungsfähiger und kostengünstiger. Im letzten Jahr produzierte die Welt zum Beispiel mehr Transistoren - und zu geringeren Kosten - als Reiskörner. Alles ist mit allem – Menschen mit Menschen, Geräte mit Geräten - immer und überall miteinander vernetzt. Die Anzahl der weltweiten Breitbandleitungen lag 2005 bei 215 Millionen und wird im Jahr 2010 500 Millionen erreichen. Supercomputing gibt's für jedermann.

Es stellt sich die Frage: Was mache ich damit? Antwort: Etwas Sinnvolles. Das gibt genau den Kick, der Reserven für Wachstum freisetzt, Kosten senkt, Effizienz und Flexibilität steigert und auch dafür sorgt, dass Unternehmen schneller und beweglicher im Hinblick auf neue Wachstumschancen werden. Gerade CEOs sehen im Thema Innovation die größten Potenziale, sich im Markt zu differenzieren und neue Wachstumschancen zu erschließen. Das haben sie in der eingangs erwähnten Studie eindeutig formuliert. Damit liegen sie genau richtig, denn: Rauchende Köpfe schaffen mehr Wohlstand als rauchende Schloten.

2.4 Innovation dank Kooperation

Und die CEOs heben noch etwas hervor: Zusammenarbeit mit Dritten und gemeinsame schöpferische Kraft. Kunden, Partner, Wissenschaft – sie schließen in immer mehr Unternehmen zur Entwicklungsabteilung als Innovationsmotor auf oder laufen ihr sogar den Rang ab.

Auch die IBM probiert neue Wege aus: Zwar bleiben die Forschungsinvestitionen auf konstant hohem Niveau von fünf bis sechs Milliarden Dollar jährlich, aber wir bauen auch auf die kollektive Intelligenz unserer gesamten Belegschaft – und darüber hinaus auch auf Kunden und Partner. So haben wir in diesem Jahr mit unserem InnovationJam eine massives weltweites Online-Brainstorming veranstaltet, zu dem zehntausende IBMer und zahlreiche Dritte tausende von Ideen beigetragen haben. Diese wurden verdichtet und dann in einem zweiten Online-Jam die besten 34 Ideen in Richtung Business Plan weiterentwickelt. Auf eine weitere Verdichtung

folgte die Ankündigung, in die zehn erfolgversprechendsten Ideen 100 Millionen Dollar zu investieren. Eines dieser Projekte ist das schon erwähnte 3D-Internet und die Entwicklung entsprechender Geschäftsmodelle.

Auf solche kreativen und innovativen Prozesse muss sich auch Deutschland besinnen und aufbauen. Wir dürfen uns nicht durch die Globalisierung einschüchtern lassen! Wir können uns nicht abschotten gegen die neuen Rahmenbedingungen. Natürlich fordert und verändert uns die Globalisierung. Die Dynamik der Flat World macht uns in Deutschland zu schaffen.

Aber: Brain Power, Qualifikation, Kreativität, Vernetzung – das sind unsere Stärken. Wir brauchen die besten Universitäten und das beste Hochschulsystem. Wir brauchen Vernetzung zwischen Wissenschaft und Wirtschaft. Nicht nur punktuell, sondern systematisch und institutionalisiert. Wir brauchen Internationalität und Bedingungen, die es den Besten ermöglicht, zu forschen, zu kooperieren und zu bleiben!

Unser Unternehmen steht in einer langen Tradition enger Kooperation mit führenden technischen und ingenieurwissenschaftlichen Hochschulen – hier in Deutschland beispielsweise zwischen der Universität Bonn und dem IBM Entwicklungslabor in Böblingen im Bereich Chipentwicklung und -design. Mit der Veränderung des Marktes und unseres Geschäftsmodells, mit der zunehmenden Verschmelzung von Informationstechnologie und Business nämlich, erweitern wir diese Kooperationen und Engagements. Beispielhaft erwähnt sei die Zusammenarbeit mit dem Elitestudiengang „Finance and Information Management“ an der TU München und der Universität Augsburg. Hier entwickelt sich ein Kompetenzzentrum für Services Science, einem der wichtigsten akademischen Wachstumsthemen für unser Land.

Deutschlands Ressourcen sind seine Menschen. In sie muss investiert werden – in ihre Ausbildung, ihre Kreativität und ihren Willen sich Veränderungen zu stellen und sie als Chance zu nutzen. Die Exzellenzinitiative der Bundesregierung ist ein Schritt in die richtige Richtung. Den Kostenwettbewerb mit den prosperierenden Märkten im nahen und fernen Osten können wir nicht gewinnen. Wollen wir unseren erarbeiteten Lebensstandard halten, müssen wir uns auf unsere geistigen und kreativen Stärken konzentrieren und darüber hinaus den Willen haben, besonders und einmalig zu sein. Kurz innovativ zu sein!

Serious Games: The Impact of Pervasive Gaming on Business

Matthias Jarke

Präsident der Gesellschaft für Informatik e.V.,
Leiter des Fraunhofer-Institut für Angewandte Informationstechnik FIT
Lehrstuhl für Informationssysteme RWTH Aachen

jarke@informatik.rwth-aachen.de

In the public debate, computer games tend to have a bad name due to their perceived negative influence on young people. Worsening school performance of boys relative to girls, even massacres in schools, are attributed to this influence.

If nothing else, this debate points out that computer games cause lots of emotions in people, get them involved, and facilitate learning. The short-term and long-term impact has even been confirmed by recent neuro-imaging studies of player brains. This impact grows further with media richness and mobility in what is called pervasive gaming.

Business has begun to take advantage of these effects in many different ways. The obvious ones include new game products as well as gaming services related to existing products, e.g. for advertising or explanatory purposes. An interesting family of such products include those where mobile multimedia technologies are being combined with geographic information services, e.g. history games in city tourism, or city or company site planning.

But there are also other important application domains. The future of IT lies in networked embedded systems whose requirements are extremely difficult to analyse before these systems have been built. Game-based design is being pursued as a novel analysis strategy in applications as diverse as production process optimization, or wearable computing in firefighter guidance and protection.

Last not least, pervasive gaming can be used to alleviate the increasing shortage of well-trained personnel, or to retrain people for new business tasks. The effective design of such learning systems, embedded in the normal work process as opposed to classroom settings, is still quite a challenge from a technology, usability, and business perspective. We have found it surprisingly useful to gain initial experiences with learning environments for users with special needs, which then lead to general innovations.

The talk will present an overview of modern pervasive gaming technologies, their business applications, and remain challenges for Business Informatics research, illustrated by a number of projects we have conducted in this domain over the past few years.

Dezentrales Wissensmanagement bei der EnBW

Energie Baden-Württemberg AG

Dr. Bernhard Beck LL.M

Chief HR and Information Officer
Mitglied des Vorstands und Arbeitsdirektor der EnBW

Ist eOrganisation überhaupt ein Thema in der Energiewirtschaft? Als verantwortlicher Vorstand für Personal, IT und Recht zeigt Dr. Beck am Beispiel der EnBW AG auf, wo die Herausforderungen bei der „Vernetzung“ von Menschen, Informationen und Prozessen in modernen Energieversorgungsunternehmen liegen.

Vor dem Hintergrund der Liberalisierung der Energiemärkte agiert die EnBW AG als föderalistisch aufgestelltes Unternehmen mit rechtlich eigenständigen Gesellschaften in einem sich zunehmend verschärfenden Wettbewerbsumfeld. In diesem Kontext gilt es im Sinne von Effizienz und Effektivität, das Gesamtunternehmen konsequent entlang der Wertschöpfungskette zu organisieren. Zwei Aspekte gewinnen dabei immer mehr an Bedeutung: Wissen und Zusammenarbeit.

In beiden Feldern kann der zielgerichtete Einsatz von Informationstechnologien wesentlich zum unternehmerischen Erfolg beitragen: Dabei wird auch dezentral vorhandenes Wissen mit Hilfe neuer IT-Technologien immer schneller und besser nutzbar.

Der eOrganisation muss es gelingen, die Akzeptanz zur konsequenten Nutzung dieser innovativen Informationstechnologien sicherzustellen. Erfolgskritisch ist dabei, dass die Technologien am operativen Bedarf und an den kulturellen Gegebenheiten orientiert sind und so die IT eine besondere Rolle bei der zielgerichteten Vernetzung des Wissens der Mitarbeiter spielt.

Flexible Prozessgestaltung als Basis innovativer Geschäftsmodelle

Von der Service-Orientierten Architektur zur Vision des Business Webs

Lutz Heuser

Leiter SAP Research, SAP AG
69190 Walldorf
lutz.heuser@sap.com

Stefan Lacher

SAP Research, SAP AG
69190 Walldorf
stefan.lacher@sap.com

Simone Perlmann

SAP Research, SAP AG
69190 Walldorf
simone.perlmann@sap.com

Kurzfassung

Die steigenden Anforderungen an die Flexibilität der Unternehmen hinsichtlich ihrer Geschäftsprozesse und -modelle führt zu neuen Ansprüchen an die IT-Systeme. Web Services bzw. das Konzept einer Service-Orientierten Architektur (SOA) versprechen, einem Großteil dieser Bedürfnisse gerecht werden zu können. Für die Forschung ergeben sich aus dem Übergang zu SOA zahlreiche Forschungsfelder. Einer der prominentesten Leitgedanken für SAP Research ist hierbei die Vision eines „Business Web“, einer umfassenden Plattform zur dynamischen Komposition betriebswirtschaftlicher Funktionen und Dienste.

1 Einleitung

Vor dem Hintergrund einer zunehmend globalisierten Welt und steigenden Kundenansprüchen sind Unternehmen jeglicher Größe heute dringender denn je auf Innovation angewiesen. Ihre Fähigkeit, Strategien und Geschäftsmodelle rasch anzupassen, gewinnt dabei als

Wettbewerbsfaktor zunehmend an Bedeutung. So weist beispielsweise die Studie des Marktforschungsinstituts "Economist Intelligence Unit" [Econ05] darauf hin, dass sich ‚Flexibilität‘ als eine der größten Herausforderungen erweist, der sich Unternehmen in den nächsten fünf Jahren gegenübersehen. Die zunehmenden Anforderungen an Flexibilität, aber auch an die Integration unternehmensübergreifender Geschäftsprozesse ist insbesondere auf die Konsolidierungs- und Spezialisierungstendenzen in den meisten Branchen zurückzuführen, infolge deren heute nicht mehr einzelne Unternehmen miteinander konkurrieren, sondern zunehmend ganze Wertschöpfungsketten untereinander in Wettbewerb treten. So vollzieht sich derzeit ein Wandel von linearen, statischen Lieferketten zu flexiblen Wertschöpfungsnetzwerken, deren Wert durch die Vielzahl und die Dynamik der Geschäftsverbindungen entsteht (Abbildung 1). Die heutigen Unternehmen müssen sich dieser Dynamik stellen, um auf die Herausforderungen und die Möglichkeiten der Globalisierung reagieren zu können. Innovative Internet-basierte Geschäftsmodelle werden es spezialisierten Unternehmen ermöglichen, Dienste anzubieten, zu vermitteln und zur Verfügung zu stellen. Dies erfordert jedoch inhaltlich neu definierte Prozesse, sowie eine neue Form der unternehmensübergreifenden Prozessmodellierung innerhalb der Unternehmensnetzwerke [KaOe06a].

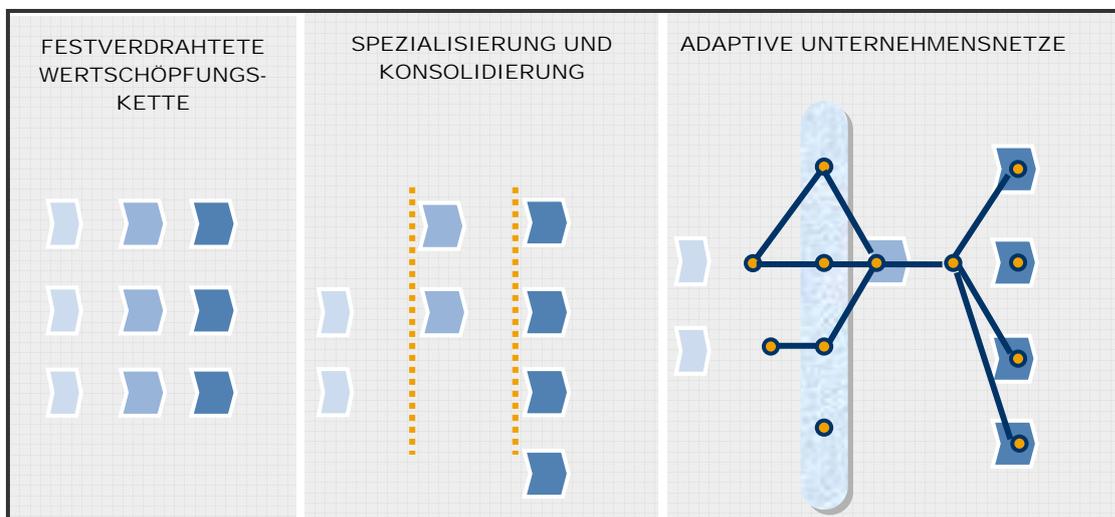


Abbildung 1: Adaptive Unternehmensnetzwerke (Quelle: SAP AG)

Die Voraussetzung für eine schnelle Reaktionsfähigkeit auf die Anforderungen des Marktes liegt in einer flexiblen IT, die offen, modular, integrationsfähig und kosteneffizient ist. Jedoch haben viele Unternehmen mittlerweile erkannt, dass ihre IT-Architektur nicht die nötige Flexibilität aufweist, um dem heutigen Wettbewerb standhalten zu können. Sie suchen daher

nach Mitteln und Wegen, ihre IT-Architektur zu optimieren und anpassungsfähiger zu gestalten. Vor diesem Hintergrund zeichnet sich ein Trend zu serviceorientierten Architekturen ab, die von vielen IT-Experten und -forschern als Schlüssel für kostengünstige und flexible IT-Systeme betrachtet werden [KaOe06b].

2 Flexible IT-Architektur durch modulare Services

Da die Geschäftsprozesse in IT-Architekturen abgebildet werden, müssen sich folglich mit den veränderten Abläufen auch die gewachsenen IT-Strukturen ändern. Hinzu kommt, dass in den herkömmlichen IT-Architekturen oftmals ein Bruch zwischen der Geschäftslogik und der Ausführung der Prozesse zu verzeichnen ist. In Folge dessen müssen nun komplexe Prozesse in Einzelschritte zerlegt und Softwarebausteinen zugeordnet werden, um die Logik der Geschäftsabläufe von der Komplexität der existierenden heterogenen Applikationslandschaften und Inselanwendungen zu entkoppeln. Der Vorteil: Die ‚Softwarebausteine‘ müssen fortan nicht mehr im Unternehmen selbst vorgehalten werden, sondern können beispielsweise über das Internet abgerufen werden. Hierbei handelt es sich um so genannte Web-Services, die über offene, standardisierte Schnittstellen wie XML (eXtensible Markup Language) eingebunden werden [Alon04]. Werden in einem nächsten Schritt die technischen Bausteine mit betriebswirtschaftlichen Funktionen verknüpft, so erhält man Enterprise-Services. Enterprise Services stellen die Geschäftslogik bereit, die sich für kommende Anforderungen rasch und flexibel kombinieren lassen. Das schafft die notwendige Flexibilität und Offenheit, um neue Geschäftsmodelle schneller umzusetzen und Prozesse zu automatisieren. Diese Offenheit lässt zudem die Integration bestehender Applikationen ebenso zu wie die Neuentwicklung von Programmen und die Einbindung von Software beliebiger Hersteller. Damit ist der Weg frei für eine IT-Architektur, die dem Anwender baukastenartig die Möglichkeit eröffnet, seine Prozesse zu modellieren und dabei automatisch seine IT-Architektur anzupassen. Das Konzept der SAP hierfür heißt Enterprise Service-Oriented Architecture (Enterprise SOA) [WoMa06].

2.1 Enterprise Service-Oriented Architecture

Die Enterprise Service-Oriented Architecture (Enterprise SOA) erweitert das Konzept der Web-Services zu einer Architektur für Geschäftsanwendungen: Während Web-Services zunächst nur

ein technisches Konzept darstellen (siehe Abbildung 2), ist Enterprise SOA der Entwurf für umfassende und Service-basierte Geschäftsanwendungen [Zenc06].

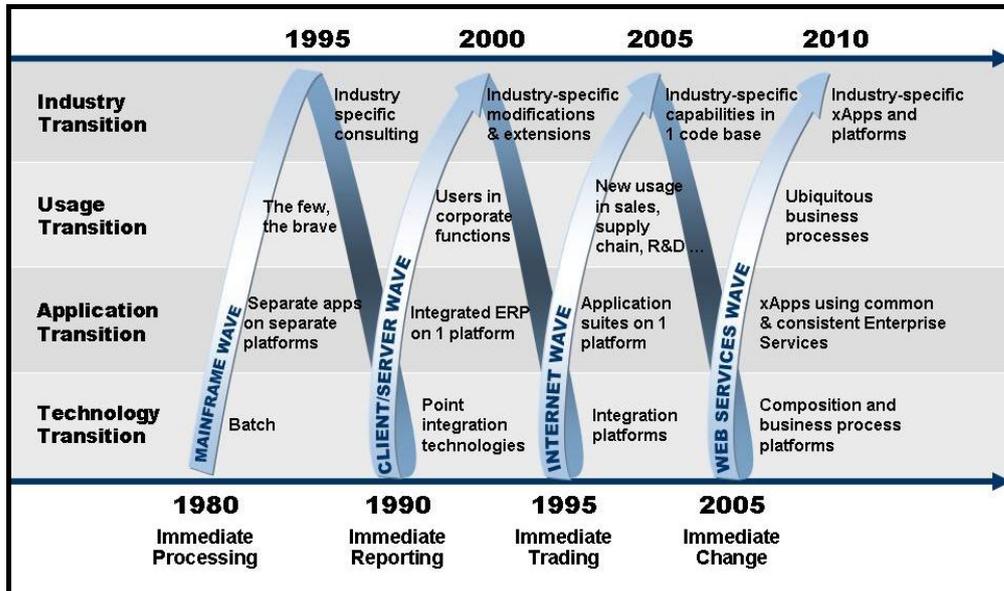


Abbildung 2: Evolutionsschritte in der Geschichte der Unternehmenssoftware (Quelle: SAP AG)

Enterprise SOA ermöglicht das Design der kompletten Geschäftsprozess-Lösung. Hierbei werden sowohl existierende Systeme und Anwendungen einbezogen als auch der Einsatz neuer Funktionalitäten beschleunigt. Die zentrale Komponente ist dabei das Composite Application Framework, eine Modellierungs- und Entwicklungsumgebung für die Komposition von Web Services (siehe Abbildung 3). SAP NetWeaver dient dazu, eine Enterprise SOA umzusetzen [SAP06] und stellt entsprechende Entwicklungs-, Integrations- und Laufzeitumgebungen zur Verfügung.

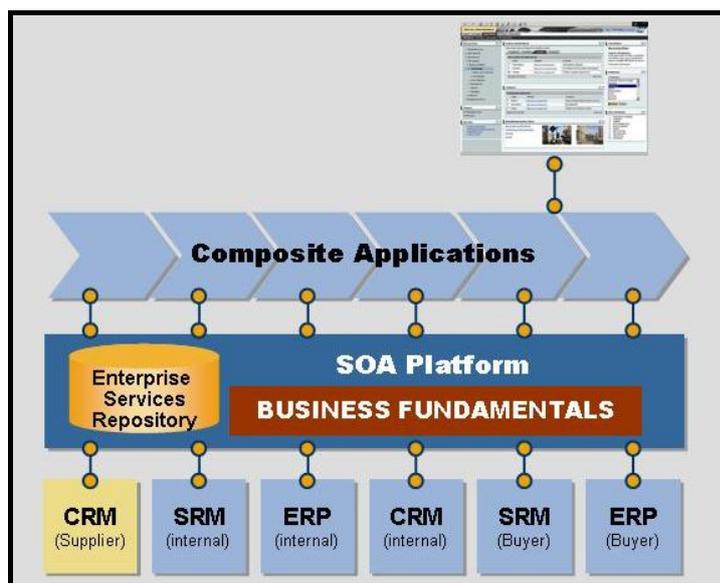


Abbildung 3: Enterprise SOA (Quelle: SAP AG)

2.2 Die Geschäftsprozessplattform

Offen bleibt nun die Frage, wie die technische Umsetzung einer Enterprise SOA-Lösung erfolgt: Vorrangiges Ziel der Enterprise SOA ist eine service-basierte Geschäftsprozessplattform auf Basis von SAP NetWeaver, die eine anwenderfreundliche ‚Komposition‘ von Anwendungen ermöglicht [Kage05]. Diese Plattform umfasst Tools zur Planung, Entwicklung und Koordination komplexer Services und Geschäftsprozesse innerhalb der serviceorientierten Landschaft und schließt Sicherheitsmanagement und Stammdatenverwaltung sowie die Datenkonsolidierung aus verschiedenen Quellen ein [Heus06b]. Diese Entwicklung steht im Einklang mit der zunehmenden Industrialisierung der Softwareentwicklung, da Plattformansätze beispielsweise in der Automobilindustrie schon seit vielen Jahren einen wichtigen Erfolgsfaktor darstellen [GaCu02]. Neben der Nutzung von Skaleneffekten sowie einem hohen Maß an Anpassungsfähigkeit hat der Plattformansatz in der Softwareindustrie noch weitere Vorteile (siehe Abbildung 4): So ermöglicht eine solche Plattform in neuem Masse eine Öffnung gegenüber Drittanbietern, die mit eigenen Anwendungen die Plattform nützen können bzw. denen auch die Möglichkeit für Verbundapplikationen mit anderen Anbietern gegeben wird [AgZe05].

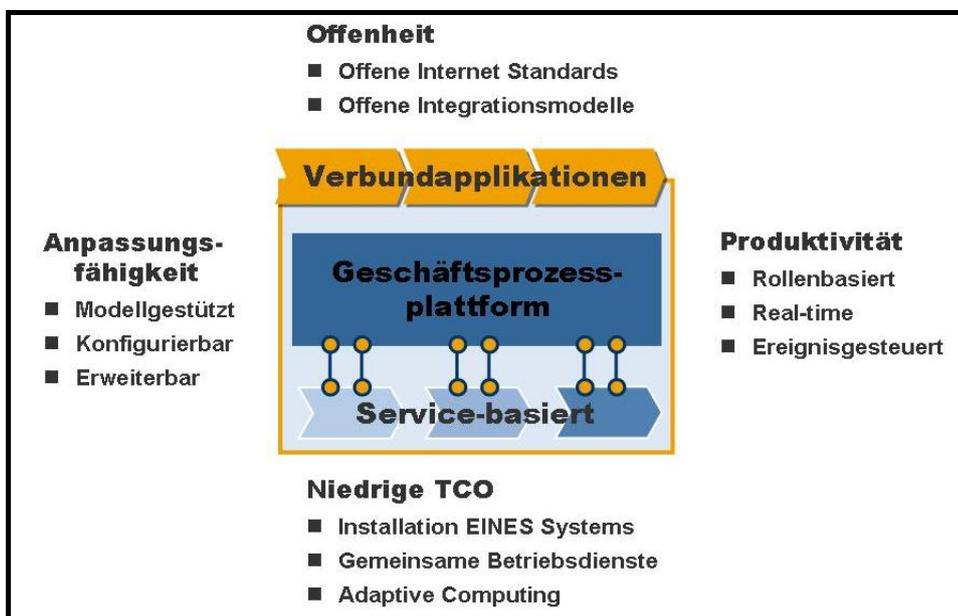


Abbildung 4: Der Nutzen der Prozessplattform (Quelle: SAP AG)

Dies jedoch führt zu erheblichen Herausforderungen an die Interoperabilität der Anwendungen. Teilweise können diese über eine einheitliche geschäftliche Ontologie geregelt werden, sie erfordern darüber hinaus jedoch eine zentrale Koordination. Hierzu sieht SAP eine Abstimmung der Services mit dem ‚Enterprise Service Repository‘, dem zentralen Verzeichnis aller Dienste

der Geschäftsprozessplattform vor. Dieser Kern der Architektur wird von SAP verwaltet, so dass trotz Öffnung der Lösung die Integrität und Einhaltung regulatorischer Richtlinien stets sichergestellt bleibt [Kage06]. Somit bietet die Geschäftsprozessplattform dem Kunden sowohl die Vorzüge einer Standardsoftware mit „Best Practice“ Prozessen und die Sicherstellung der Prozessintegrität als auch die Flexibilität einer individuell erweiterbaren Lösung [AgZe05]. Unterstützt wird dieses Konzept durch das Partner Ecosystem der SAP, da die SAP natürlich nicht jede einzelne Funktion für Unternehmen jeder Größenordnung in jeder Branche entwickeln kann. Grundsätzlich kann im Rahmen des zuvor vorgestellten Konzepts jede Funktion, die als Service bereitgestellt wird, wieder verwendet werden, wenn neue ‚Schichten‘ auf die Anwendung gesetzt werden. Solche Schichten benötigen Services und fügen der Anwendung Funktionen hinzu, die ursprünglich nicht für diese Anwendung konzipiert wurden. Somit eröffnet SAP mit Hilfe der Services über ihre Anwendungen hinaus neue Funktionen und ermöglicht den Systemhäusern (Independent Software Vendors ISVs) des Partnernetzes ein flexibles Wachstum [Kage06].

3 Zukünftige Forschungsfelder

Vor diesem Hintergrund werden einige der zentralen Themen von SAP Research deutlich: So ist absehbar, dass das Enterprise SOA-Konzept sowie eine entsprechende Infrastruktur in naher Zukunft eine passende Möglichkeit für Unternehmen bietet, sich auf ihr dynamisches Umfeld einzustellen und ihre Geschäftsmodelle entsprechend flexibel an die veränderten Gegebenheiten einer globalen Internet-basierten Wirtschaft anzupassen. Zumeist konzentriert sich die Wertschöpfung jedoch nicht auf einen einzelnen Teilnehmer, sondern setzt sich aus einer Vielzahl von Partnern zusammen, die einen gemeinsamen Mehrwert für den Kunden anstreben. Folglich erscheint es notwendig, dass ‚Service Communities‘ schnell und einfach Kollaborationen mit anderen Geschäftspartnern aufbauen, ändern und ggf. auch unproblematisch beenden können. In diesem Zusammenhang entwickelte SAP Research die Vision der Business-Webs.

Business-Webs verfolgen hierbei die Grundidee zukünftiger Service-Marktplätze, die eine einfache und flexible Integration spezialisierter Dienste zu einem Gesamtangebot für den Dienstanutzer entstehen lassen. Um die Kosten eines solchen Ansatzes möglichst gering zu halten und die Integration der Services zu einem größtmöglichen Teil automatisiert

vorzunehmen, forscht SAP Research einerseits im Bereich semantischer Technologien und andererseits an passenden Werkzeugen und Entwicklungsprozessen zur vereinfachten Komposition derartiger Services sowie an den hierfür notwendigen Sicherheitstechnologien. Als erste Ansätze weiterentwickelter Web-Anwendungen können beispielsweise „Mash-ups“ gesehen werden, wie sie in letzter Zeit im Rahmen der Web 2.0-Aktivitäten wiederholt im Gespräch waren.

Um die Vision von den Business-Webs zu verwirklichen, arbeitet SAP Research zudem an einer Reihe von weiteren Problemstellungen, die von ökonomischen und rechtlichen Gesichtspunkten (z.B.: Service Innovationen oder Haftungsregelungen in Business-Webs) über die Koordination bzw. Bereitstellung verteilter Dienste bis hin zu einer semantischen Infrastruktur und der Interoperabilität zwischen Business-Webs reichen. Diese Konzepte werden den Unternehmen von morgen weitere Möglichkeiten eröffnen, die Chancen und Herausforderungen der Internet-basierten Globalisierung flexibel zu gestalten.

4 Conclusio

Unternehmen werden sich in den kommenden Jahren weiterhin verstärkt durch Innovationen im Bereich von Diensten, Prozessen und speziell auch Geschäftsmodellen vom Wettbewerb differenzieren. Der bereits heute wichtige Faktor der IT wird dabei zu einem strategischen Differenzierungskriterium, das hinsichtlich der Anpassung der Geschäftsmodelle an die jeweiligen Marktgegebenheiten maßgeblich beteiligt sein wird. Die SAP bietet ihren Kunden auf Basis ihrer Geschäftsprozessplattform sowie der Enterprise Services die Möglichkeit einer flexiblen Anpassung der Geschäftsanwendung an die eigenen Bedürfnisse. Unabhängigen Softwareanbietern wird die Chance eröffnet, im Rahmen eines Partnernetzwerks mit anderen Anbietern zu kooperieren und über dieses Modell unternehmensübergreifende Innovationen voranzutreiben. Für die Kunden entstehen somit „Best Practice“ Lösungen der jeweiligen Spezialanbieter für Software, wobei gleichzeitig die Integrität und die Gesetzeskonformität der Daten über die zentrale Plattform sichergestellt wird [Heus06a].

Die Enterprise Services ermöglichen den Unternehmen darüber hinaus ein hohes Maß an Flexibilität für die eigenen Kundenangebote, aber auch für die zunehmend wichtiger werdende dynamische Gestaltung der Wertschöpfungskette. Hierfür soll eine Umsetzung der Vision des

Business-Webs in den nächsten Jahren den entsprechenden ökonomischen und technischen Rahmen bieten [Heus06b].

Literaturverzeichnis

- [AgZe05] Agassi, S.; Zencke, P.: Best-of-Both-Worlds Benefit, Interview in SAP INFO, 27.03.2005. Verfügbar unter: <http://www.sap.info>.
- [Alon04] Alonso, G. et al.: Web Services – Concepts, Architectures and Applications, Springer, Berlin, Heidelberg, 2004.
- [Econ05] The Economist Intelligence Unit: Business 2010 - Embracing the challenge of change, London, New York, Hong Kong, 2005. Verfügbar unter: http://www.eiu.com/site_info.asp?info_name=eiu_SAP_business2010.
- [GaCu02] Gawer, A.; Cusumano, M. A.: Platform Leadership. Harvard Business School Press, Boston, 2002.
- [Heus06a] Heuser, L.: Das Business Web – eine zentrale Vision von SAP Research, Pressekonferenz zur 36. Jahrestagung der Gesellschaft für Informatik, 04.10.2006.
- [Heus06b] Heuser, L.: Unsere Herausforderung liegt in der Servicekomposition, Interview in Computer Zeitung, 13.11.2006.
- [Kage05] Kagermann, H.: SAP plans new platform as competitive weapon, Interview in InfoWorld, 31.01.2005.
- [Kage06] Kagermann, H.: Balancing Change and Stability in the Evolution of SAP's Enterprise Software Platform, Interview in Knowledge@Wharton, 04.10.2006.
- [KaOe06a] Kagermann, H.; Oesterle, H.: Geschäftsmodelle 2010 - Wie CEOs Unternehmen transformieren, Frankfurter Allgemeine Buch, 2006.

- [KaOe06b] Kagermann, H; Oesterle, H.: Nur mit SOAs bleiben Firmen wettbewerbsfähig, Interview in CIO Online, 30.08.2006. Verfügbar unter: <http://www.cio.de/index.cfm?pid=287&pk=826419>.
- [SAP06] SAP AG: SAP INFO Glossar, 2006. Verfügbar unter: <http://www.sap.info/public/DE/de/glossary/de>.
- [WoMA06] Woods, D.; Mattern, T.: Enterprise SOA – Designing IT for Business Innovation, O’Reilly Media, Sebastopol, 2006.
- [Zenc06] Zencke, P.: Geschäftsmodell 2010 - Paradigmenwechsel durch SOA, Interview in CIO Online, 27.09.2006. Verfügbar unter: <http://www.cio.de/index.cfm?pid=324&pk=827738>.

Informationen zu SAP

Die SAP AG, mit Hauptsitz in Walldorf, Bundesrepublik Deutschland, ist der weltweit führende Anbieter von Unternehmenssoftware. Das Portfolio der SAP umfasst Geschäftsanwendungen für große Unternehmen und den Mittelstand, die auf der SAP NetWeaver-Plattform aufbauen, sowie leistungsfähige Standardlösungen für kleine und mittelgroße Firmen. Darüber hinaus unterstützt SAP mit mehr als 25 branchenspezifischen Lösungsportfolios Kernprozesse in Industrien wie Handel, Finanzen, High-Tech, im Gesundheitswesen und öffentliche Verwaltungen. Damit sind Organisationen in der Lage, ihre Geschäftsprozesse intern sowie mit Kunden, Partnern und Lieferanten erfolgreich zu organisieren und die betriebliche Wertschöpfung maßgeblich zu verbessern. SAP-Lösungen sind bei über 32.000 Kunden in mehr als 120 Ländern im Einsatz. SAP wurde 1972 gegründet und ist heute der weltweit drittgrößte unabhängige Softwareanbieter, mit Niederlassungen in über 50 Ländern. Im Geschäftsjahr 2005 erzielte das Unternehmen einen Umsatz von rund 8,5 Mrd. Euro. Derzeit beschäftigt SAP über 35.000 Mitarbeiter. Weitere Informationen unter: www.sap.com.

Informationen zu SAP Research

SAP Research ist die Forschungsabteilung der SAP AG und als solche verantwortlich für die Bereiche der Angewandten Forschung sowie der Vorentwicklung. SAP Research kooperiert eng mit führenden Universitäten und Forschungseinrichtungen. SAP Research verfügt über zahlreiche Forschungsstandorte weltweit. Die wesentlichen Teilaufgaben liegen in der:

- Technologiefrühaufklärung und Erstellung des Forschungsportfolios;
- Identifizierung von Zukunftstechnologien sowie deren Anwendungspotentiale im Markt für Unternehmenssoftware;
- Sicherung der Technologiepotentiale mittels Patentierung; sowie der
- Bereitstellung, Speicherung und Verwertung des Wissens nach dem Grundprinzip der kollaborativen Forschung.

Search Engine Economics

Hal R. Varian

School of Information

University of California at Berkeley

102 South Hall, Berkeley, CA 94720-4600, USA

hal@sims.berkeley.edu

Bio: <http://www.ischool.berkeley.edu/~hal/>

Search engine advertising has become a big business, with the combined revenue of industry leaders Yahoo and Google exceeding \$11 billion in 2005. Most all of these ads are sold via auction, with literally billions of these auctions being held per week.

This talk is concerned the economics of search engine ad auctions: how they evolved, how they work, and where they are going. I also speculate on how search engine advertising will affect other marketing efforts.

Decentralised Control of Complex Systems

Nicholas R. Jennings

School of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ
UK
nrj@ecs.soton.ac.uk

Bio: <http://www.ecs.soton.ac.uk/~nrj/>

Many modern computing systems have to operate in environments that are highly interconnected, highly unpredictable, are in a constant state of flux, in which there is no centralized point of control, and in which the constituent components are owned by a variety of stakeholders that each have their own aims and objectives. Relevant exemplars include the Web, Grid Computing, Peer-to-Peer systems, Pervasive Computing and many eCommerce applications. Now, I believe that all of these systems operate under the same conceptual model: (i) entities offer a variety of services in some form of institutional setting; (ii) other entities connect to these services (covering issues such as service discovery, service composition and service procurement); and (iii) entities enact services in a flexible and context sensitive manner. Moreover, I believe agent-based computing is an appropriate model computational for such systems (Jennings, 2000; Jennings 2001). In particular, autonomous agents are a natural way of viewing flexible service providers and consumers and the interactions between these autonomous components are naturally modeled as some form of economic trading process that, if successful, results in a service contract (or service level agreement) between the agents involved.

In this talk, the focus will be on the design of the agents and their interactions. Specifically, I will consider the design and implementation of various computational service economies for a number of real-world applications including: virtual organizations (Norman et al., 2004), sensor

networks (Padhy et al., 2006; Rogers et al., 2005; Rogers et al., 2006) and personalized recommendations (Wei et al., 2005; Payne et al., 2006). In so doing, I will touch upon some of the techniques and advances we have made in the areas of game theory (Gerding et al., 2006), auctions (Dash et al., 2007, David et al., 2005), coalition formation (Dang and Jennings, 2006; Dang et al., 2006; Rahwan et al., 2005; Rahwan and Jennings, 2005), automated negotiation (Fatima et al., 2004; Fatima et al., 2006) and computational mechanism design (Dash et al., 2003).

References

V. D. Dang and N. R. Jennings (2006) "Coalition structure generation in task-based settings" *Proc. 17th European Conference on AI*, Trento, Italy, 210-214.

V. D. Dang, R. K. Dash, A. Rogers and N. R. Jennings (2006) "Overlapping coalition formation for efficient data fusion in multi-sensor networks" *Proc. 21st National Conference on AI (AAAI)*, Boston, USA, 635-640.

R. K. Dash, P. Vytelingum, A. Rogers, E. David and N. R. Jennings (2007) "Market-based task allocation mechanisms for limited capacity suppliers" *IEEE Trans on Systems, Man and Cybernetics (Part A)*.

R. K. Dash, D. C. Parkes and N. R. Jennings (2003) "Computational Mechanism Design: A Call to Arms" *IEEE Intelligent Systems* **18** (6) 40-47.

E. David, A. Rogers, J. Schiff, S. Kraus, and N. R. Jennings (2005) "Optimal design of English auctions with discrete bid levels" *Proc. of 6th ACM Conference on Electronic Commerce (EC'05)* Vancouver, Canada, 98-107.

S. S. Fatima, M. Wooldridge and N. R. Jennings (2006) "Multi-issue negotiation with deadlines" *Journal of AI Research* **27** 381-417.

S. Fatima, M. Wooldridge and N. R. Jennings (2004) "An agenda based framework for multi-issues negotiation" *Artificial Intelligence Journal* **152** (1) 1-45.

E. H. Gerding, R. K. Dash, D. C. K. Yuen and N. R. Jennings (2006) "Optimal bidding strategies for simultaneous Vickrey auctions with perfect substitutes" *Proc. 8th Int Workshop on Game Theoretic and Decision Theoretic Agents*, Hakodate, Japan, 10-17.

N. R. Jennings (2000) "On Agent-Based Software Engineering" *Artificial Intelligence Journal*, **117** (2) 277-296.

N. R. Jennings (2001) "An agent-based approach for building complex software systems" *Comms. of the ACM*, **44** (4) 35-41.

- T. J. Norman, A. Preece, S. Chalmers, N. R. Jennings, M. Luck, V. D. Dang, T. D. Nguyen, V. Deora, J. Shao, A. Gray and N. Fiddian (2004) "Agent-based formation of virtual organizations" *Int. J. Knowledge Based Systems* **17** (2-4) 103-111.
- P. Padhy, R. K. Dash, K. Martinez and N. R. Jennings (2006) "A utility-based sensing and communication model for a glacial sensor network" *Proc. 5th Int. Conf. on Autonomous Agents and Multi-Agent Systems*, Hakodate, Japan, 1353-1360.
- T. R. Payne, E. David, N. R. Jennings and M. Sharifi (2006) "Auction mechanisms for efficient advertisement selection on public displays" *Proc. 17th European Conference on AI*, Trento, Italy, 285-289.
- T. Rahwan, S. D. Ramchurn, V. D. Dang and N. R. Jennings (2007) "Near-optimal anytime coalition structure generation" *Proc 20th Int. Joint Conf. on AI (IJCAI)*, Hyderabad, India.
- T. Rahwan and N. R. Jennings (2005) "Distributing coalitional value calculations among cooperating agents" *Proc 25th National Conf on AI (AAAI)*, Pittsburgh, USA, 152-157.
- A. Rogers, E. David and N. R. Jennings (2005) "Self organized routing for wireless micro-sensor networks" *IEEE Trans. on Systems, Man and Cybernetics (Part A)* **35** (3) 349-359.
- A. Rogers, R. K. Dash, N. R. Jennings, S. Reece and S. Roberts (2006) "Computational mechanism design for information fusion within sensor networks" *Proc. 9th Int. Conf. on Information Fusion*, Florence, Italy.
- Y. Z. Wei, L. Moreau and N. R. Jennings (2005) "A market-based approach to recommender systems" *ACM Trans on Information Systems* **23** (3) 227-266.

Einführung in den Track

Das Internet der Dinge: Betriebliche Anwendungen des RFID und Ubiquitous Computing

Prof. Dr. Elgar Fleisch

Universität St. Gallen

Prof. Dr. Franz Lehner

Universität Passau

Prof. Dr. Karl Kurbel

Universität Frankfurt (Oder)

Dr. Orestis Terzidis

SAP Research

Technologien wie RFID (Radio Frequency Identification), Sensornetzwerke und Ubiquitous Computing machen das Internet der Dinge möglich. Im Unterschied zum Mobile Computing, bei dem die Geräte als eigenständige Entitäten sichtbar sind und als solche wahrgenommen werden, stehen bei Ubiquitous und Pervasive Computing integrierte und eingebettete Systeme im Mittelpunkt, die jederzeit und überall genutzt werden können.

Wenn nicht mehr nur die Abstraktionen der realen Dinge Gegenstand der Informationsverarbeitung sind, sondern die Dinge selbst Daten senden, empfangen und verarbeiten, dann eröffnen sich für die Wirtschaftsinformatik neue Chancen, aber auch Herausforderungen: Wie kann der riesige Datenanfall semantisch adäquat bewältigt werden, so dass relevante Managementinformationen gewonnen werden? Welche Sicherheits- und Privacy-Anforderungen sind zu beachten und welche rechtlichen Rahmenbedingungen sind zu schaffen? Welche Infrastrukturen werden künftig benötigt werden? Wie wird sich der Markt entwickeln, welche Rolle spielen Standards, und wer werden die Key Players sein? Was ist bei der Einführung zu beachten und welche Erfolgsfaktoren können identifiziert werden?

Programmkomitee:

Prof. Dr. Sahin Albayrak, Technische Universität Berlin

PD Dr. Christian Becker, Universität Stuttgart

Prof. Dr. Schahram Dustdar, Technische Universität Wien

Prof. Dr. Alois Ferscha, Universität Linz Prof.

Dr. Elgar Fleisch, Universität St. Gallen Konrad Klöckner, FIT, Sankt Augustin

Prof. Dr. Gabriele Kotsis, Universität Linz

Prof. Dr. Karl Kurbel, Universität Frankfurt (Oder)

Dr. Marc Langheinrich, ETH Zurich

Prof. Dr. Franz Lehner, Universität Passau

Ulrike Lucke, Universität Rostock

Prof. Dr. Friedemann Mattern, ETH Zurich

Dr. Kay Roemer, ETH Zurich

Dr. Albrecht Schmidt, Universität München

Dr. Orestis Terzidis, SAP Research

Jun.-Prof. Dr. Frank Teuteberg, Universität Osnabrück

Start a Grassroots RFID Initiative!

The Relevance of Communication and Showcases on the Success of RFID.

Florian Resatsch, Jörg Aßmann, Thomas Schildhauer

Institute of Electronic Business
University of the Arts Berlin
D-10639 Berlin
{resatsch,assmann,schildhauer}@ieb.net

Daniel Michelis

Institute of Media and Communication Management
University of St. Gallen
CH-9000 St. Gallen
daniel.michelis@unisg.ch

Abstract

Since about 1994 there are now more microprocessors than humans on this planet. With technology advancements and lowered prices the trend will go on. But in a world of ubiquitous connectivity, computing-not computers-characterizes the next era of the computer age. Enabling technologies that support wireless connections, such as Radio Frequency Identification, are being implemented in many businesses to optimize processes and to track and trace objects automatically. But researchers around the world are already working to develop further applications and ideas related to ubiquitous computing beyond tracking and tracing. To discover why most discussions outside the university and the consulting arena haven't yet reached the point where the consumer perceives RFID as a living technology, we conducted a study during a large trade fair in Germany. The results and implications are presented in this paper.

1 Introduction

The multifarious term "Ubiquitous Computing" (Ubiocom) is widely used throughout business and media today. The vision behind Ubiocom was described more than 10 years ago by Mark

Weiser as the “Enhancing of computer use by making many computers available throughout the physical environment, but making them effectively invisible to the user.” [Weis91, Weis93]. Weiser coined the term in reference to computers that assist people in their daily lives at home and work [BCLM05]. Favored by technological development processes, such as miniaturized chips, the vision comes closer these days [Mano01; GeKC04; FIMa05]. Despite the basic ideas of Ubicom, the potentially ubiquitous available technology, Radio Frequency Identification (RFID), is currently integrated mostly in tracking and tracing processes to improve and optimize logistics. In order to focus on the topic RFID in the consumer world, we conducted a study on RFID during an international consumer electronics trade fair in Germany. The results show that current discussions of RFID in the media are misleading. Most media discussions center around privacy issues, mostly because “Ubicom systems are relatively unusual” [Beck03] and designers of Ubicom systems lack a certain amount of knowledge about the ideas of potential users. There is also a gap between media communication and public perception, in that both Ubicom and RFID seem to be rather unknown entities. The question arises whether discussions about privacy hold back the full potential of RFID applications in both consumer and business arenas and thereby restrict future developments. The aim of this paper is to show why researchers as well as practitioners should focus on more showcases and prototypes for consumers that include reasoning about privacy, but also start discussions based on valid application development regarding usage and benefits to the consumer. The main question is how to show the benefits of a technology based principally on invisibility. What are consumers really thinking and what information could be made available to people in order to make the technology well-known so that the consumer is able to accurately evaluate it before using or purchasing it [DaKa73]? This paper describes the current state of interest in RFID technology in section two, followed by a description of the study design in section three. Section four points out the empirical findings of the study. In section five we propose a way to categorize discussions around RFID-related issues and finish with a conclusion and limitations in section six.

2 RFID in the Innovation Cycle

RFID technology has been broadly discussed in literature [AbMy00; Matt02, Matt03; Flei04; FIMa05; FKTT06; KnLK06; Want06]. However, new or somewhat new technologies take a

long time to impact the lives of ordinary people [Norm99; Kaas05]. The adoption of “disruptive technologies” is an even longer and more complicated process. “Disruptive technologies” are technologies that may cause revolutionary changes in people’s lives [Chri97], a fact that is being spread by many Ubicom visionaries [Matt02; Matt03]. Christensen uses the term “Disruptive *innovation*” in his later work to underline that strategy creates the disruptive impact, not the technology in itself – only a few technologies have the intrinsic potential to be disruptive [Chri03]. Real “disruptive technologies” provide a new use case category for a technology that was previous unavailable. The question is to what extent can RFID be considered a disruptive technology and what would be the implication of this? Although RFID technology has been available since the late 40’s [Stoc48], within the last years we see a new era in mass usage [KnLK06]. The supporters of the technology, such as huge retail companies, regard RFID as an incremental improvement that helps to optimize their processes. In North America, especially the efforts of EPCglobal have mainly dominated the development of RFID applications with focus on the use cases of the aforementioned companies and have generated a main interest in supply chain management [FKTT06]. Fine, Klym et al. examine four key trends. First: the issue of standardization that leads towards interoperable applications. Second: most RFID systems exploit the combination of short-range radio-based communication networks that have thus far evolved independently. Third: they suggest that newer optical tagging systems should compete with RFID despite the performance advantages [FKTT06]. All of these can be considered to be more incremental than disruptive innovations, however as a final point, Fine et al. see RFID and tagging technologies making their way into the hands of end-users – which is currently not yet the case [WiDe04]. This would significantly increase the potential for disruptive rather than incremental improvements. The development of more grassroots projects would provide the necessary counterbalance to the tracking and tracing efforts [FKTT06]. Such grassroots tagging projects can already be seen in some areas of interaction design and human-computer interaction. Especially since large mobile phone producers introduced Near Field Communication phones including RFID readers, new ideas on bridging the gap between the real and virtual world pop up. Users can create their own data related to an object code, or use their own tags and ideas to build new applications, such as the Address Book Desk by Timo Arnall [Arna05] with a real use to people. Within the supply chain and logistics, RFID is more likely to be discontinuous improvements contrary to optical systems, but in the area of consumer application, RFID could still be a disruptive innovation. If

that were the case, it is important to locate the technology in the innovation cycle to determine which steps are necessary to create applications that are widely accepted by the users. Broad consumer acceptance would then automatically lead to greater acceptance of important industry use-cases such as item-level tagging. According to Sheffi, RFID “Is still not out of the fog of innovation: the benefits of the technology are not entirely clear, especially its advantages over bar code technology” [Shef04]. Furthermore he states that the discussions about standards and privacy issues are indicative of a technology still in its infancy. One of the aspects of technology is that the move from invention to innovation, only takes place when consumers adopt and accept a technology in large numbers [Shef04]. Based on Davis’ early work regarding technology acceptance [Davi89], Davis and Venkatesh suggest that pre-prototypes can also be used to assess usefulness [DaVe04; Kaas05]. Most RFID applications are within the supply chain, and as a result there are fewer prototypes visible and testable to the consumer, thus limiting their possibilities to accept the new technology. The question is how far the focus on technology acceptance will help the dissemination of RFID. Consumers and users might prefer to experience it first. Rogers describes the innovation adoption process in five steps “Through which an individual [...] passes from first knowledge” [Roge95] to finalizing the adoption decision. The dissemination itself is regarded as the process by which an innovation is communicated through various channels among members of a social system [Roge95]. Rogers defines five adopter categories in his work: innovators, early adopters, early majority, late majority and laggards. While the innovators are risk-takers, the early adopters are opinion leaders who make the initial evaluation of an innovation and communicate the result to the other members of their group. If the early adopters are convinced by innovators that a technology is worth it, the technology can spread into the mass market after crossing the “chasm” - the different needs of early adopters and early majority [Moor02]. The decision to accept an innovation depends on the innovation-decision of other members, and here networking exerts a major influence. This also means that communication channels such as mass media have a powerful effect on spreading the message. But the most trustworthy channel for leaders of opinion is word-of-mouth - even more so than mass media. The adoption of innovation is therefore largely an information processing activity to which potential adopters have gathered enough information on the particular innovation [Hain05]. Rogers also states that the available information about an innovation may positively influence its adoption probability [Roge95]. In the case of RFID, this would mean that - if the invention RFID is an innovation and if we are

still “in the fog of innovation” more information (equals experience and visible benefits) on RFID is crucial for large-scale adoption. We used the free tool “Google Trends” to ascertain discussions about RFID, showing the interest of the Internet users in specific topics. Search queries reflect a late majority indicator since people start searching for a specific term when they have already heard about it via a different channel. For example, if someone had heard about RFID in the media they might be conducting a search about it.



Fig. 1: The search queries for RFID and supply chain (Source: Google Trends)

Figure 1 shows that search queries for RFID and supply chain show a similar course over time. The queries maintain a stable level; therefore people might search for supply chain and RFID in the same manner. Nevertheless, RFID still proves to be of value outside classical track and trace applications. Figure 2 compares a real consumer product like the Apple™ iPod® with RFID. It can be clearly seen that for a significant period the news reference volume was almost equal while the iPod® related search queries quickly exceeded the RFID search queries. The news reference volume maintained the same relation for a while until new iPods® were released. However, iPod® search queries rose sharply, consistently maintaining a higher amount than RFID-related queries. This indicates significant media-based coverage and shows the interests of the majority of internet users after the adoption of previous groups (See Figure 2).

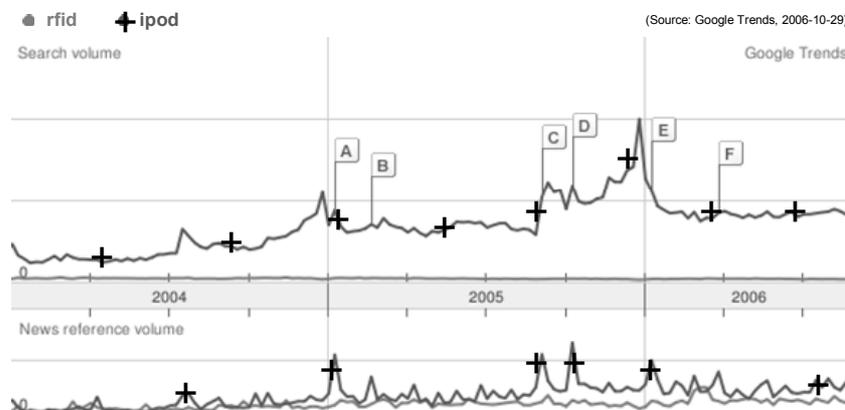


Fig. 2: Comparison of search volume vs. news reference of the consumer product iPod and the not-yet-consumer-product RFID (Source: Google Trends – Letters A to F mark relevant incidents)

Comparing the amount of search results for the term „spychips“ (German: Schnüffelchips) with “RFID” in Google shows 302.000 search results for spychips vs. 55.400.000 for RFID. However, more consumers might remember the discussion on “spychips” than any industry RFID proposal. The usage possibilities of RFID have led to many discussions within the media in terms of data security and privacy issues. The absolute number of Google search queries on the term “RFID” in Google trends worldwide is almost constant. Also, it can be stated that the type of news varied from highly positive to more negative in the last several years. It was initially euphoric (A) while in more recent times we see a focus on negative issues such as RFID viruses (E). The influence of negative news is shown in the chart below. Here Scott Silverman announced the idea of implanting RFID chips in immigrants and guest workers. This has led to a very high number of web-logs mentioning the quote (See Figure 3).

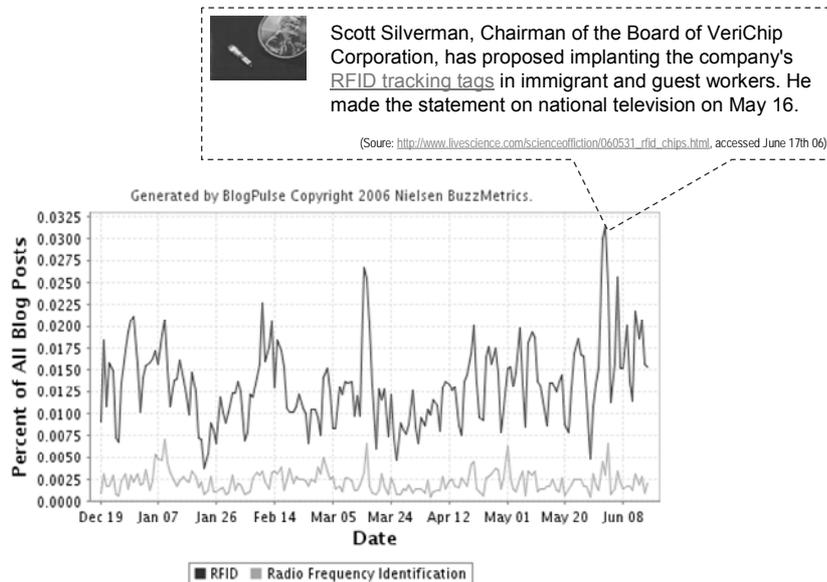


Fig. 3: The reaction in web-logs to bad news (Source: BlogPulse - We chose to search for both the acronym RFID and the term Radio Frequency Identification, because some web-logs use the acronym while others the whole set)

Clearly, privacy is an issue for RFID applications, but overall negative media coverage leads to a significant increase of reporting in many communication channels on the Internet, such as Weblogs, Communities, etc.. Since we are not able to determine the difference of opinion leaders’ opinions and other consumers with the tools used above, it seems that the search queries show an interest in the technology that is not at the same intensity than it could be for being consumer good (A test with myspace.com instead of iPod® shows an even higher difference). Although researchers and experts are very aware of the potential opportunities and

threats of RFID, the consumers might not be able to judge them appropriately. The aforementioned statements show that RFID is a potentially interesting technology for mass markets, one that *might* lead to a disruptive innovation – an effect that could cycle back into business – where the value of RFID is already proven in many cases [FICD05]. Our research focus was centered around seeing how people currently value RFID and what their opinion is of the technology. Therefore we decided to conduct a study during a world trade fair that deals with information technology and RFID. Without pre-qualifications, we asked people at the fair about different aspects of RFID. The next section describes the study design.

3 Study Design

For the study we selected a large trade fair (CeBit 2006) with presumably technologically savvy visitors. According to the organizers, the percentage of professionals in attendance was extremely high in 2006 (85% of all participants). Since the trade fair had more than 450,000 visitors and 6,262 exhibitors, we concentrated on the main hall with RFID as a major topic (#6) as well as the two adjacent halls (#5, #7). The following table (Tab. 1) gives an overview of the research design.

Research framework	Quantitative analysis/survey
Method of data collection	Personal interview with questionnaire
Period	Thursday, March 9 th 2006 and Saturday, March 11 th 2006
Measuring method	Interval 7-item-scale
Universe	World's largest trade fair for digital IT and telecommunications solutions for home and work environments; halls 5-7; Visitors of the fair.
Sample type	Random
Sample number	n=336

Tab. 1: Study data

Within a quantitative analysis in a non-representative framework, the sample type was random since the interviewers randomly approached participants in the halls. They used industrial personal digital assistants (PDAs) with questionnaire software installed (Software: mQuest). The only pre-definition was the relation of the sexes, since typically 30% of the participants at this trade fair are female.

Characteristics of the Participants

Within the study 30% of the participants were female and 70% were male. The primary business of their companies was information technology (47%), followed by research (12%),

insurance (7%), government (7%), retail (6%) and others. Most of the companies employing the participants had 20 to 99 employees (33%), followed by smaller companies ranging from 1 to 19 employees (29%). Among the participants only 5% worked for large enterprises with more than 10.000 employees.

Research Questions

The research questions were based on the aforementioned theoretical findings. Based on the premise that positive information is the key issue behind the diffusion of innovation, we asked the people about their knowledge (subjective, objective), importance (company, personal) and benefits of RFID on a 1-to-5 scale with “1” having the highest rejection and “5” having the highest acceptance. Based on the theoretical overview in section 2, we tested two hypotheses:

H1: Participants who work in branches, in which RFID prototypes already exist and are exposed to public media coverage, view RFID more positively (communication).

And, H2: Personal attributes play no role in the evaluation of RFID (gender, position)

The following section describes the findings.

4 Empirical Findings

Results of Knowledge about RFID

To get an overview of the basic knowledge of the participants we first asked if they knew what the abbreviation RFID stands for. 90% correctly identified the acronym RFID as Radio Frequency Identification. Following that, we asked the participant to rate their knowledge of RFID, which we defined as subjective knowledge (Figure 4).

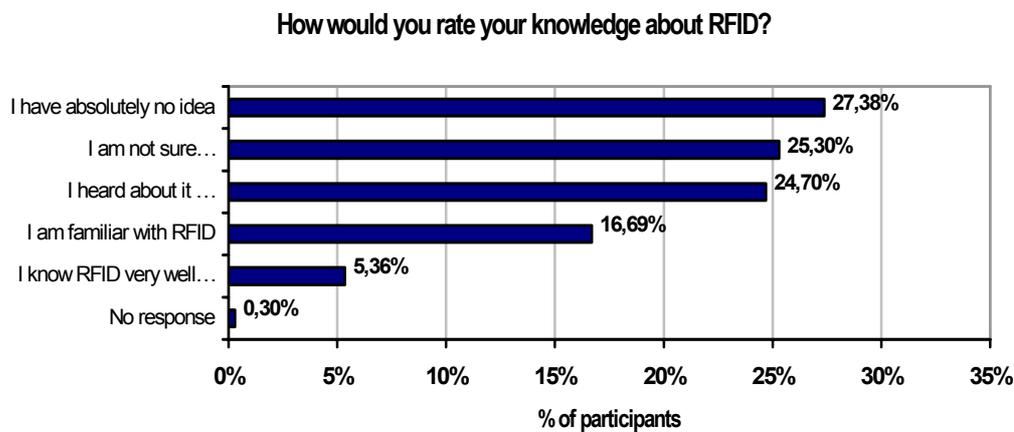


Fig. 4: Question on subjective knowledge

More than 75% of all participants rated their knowledge from “non-existent” to “I’ve heard about it”. Only a few people considered themselves to be significantly informed. Although the people did not think they know much about RFID, a clear connection emerged between answers to the question “In which areas do you see the most value of RFID?” and the branches that have been exposed to the most RFID media coverage: logistics and retail (Figure 5).

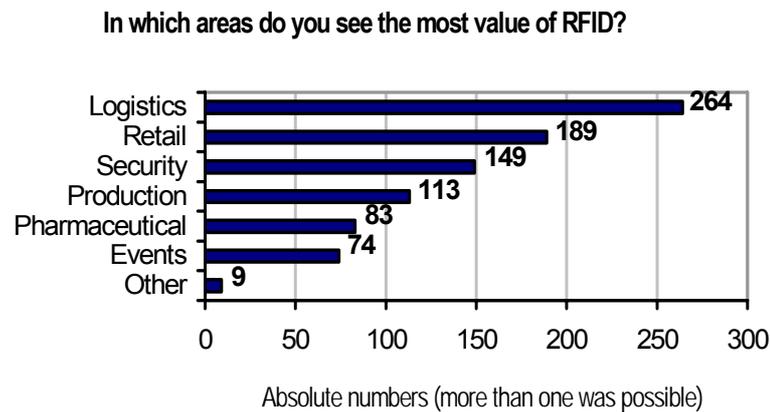


Fig. 5: Branches in which the participant sees the most value for RFID

Company Sphere Results

The second area of questions focused on company issues concerning RFID. The participants were asked to evaluate RFID for the company in which they are employed. People who specified no company were either students or involved in science or were not capable of evaluating RFID for their company. Only 24% considered RFID to be important for their company, while 44% saw no importance of RFID for their businesses. Bearing in mind that the survey was conducted *during a technology trade fair*, this is rather surprising considering the amount of RFID media coverage before the event. We also asked about the requirements for bringing RFID to mass markets: 82% of the participants said that the benefits of the technology needed to be communicated more clearly. Contrary to that finding however, 92% of all respondents answered that RFID is definitely coming. Since there are many studies on the market on the company perspective, we wanted to focus on researching the personal opinion of the people at the trade fair.

Personal Sphere Results

Evaluating something requires proper background information, or people will rely on other resources to determine personal opinions. The first question concentrated on the importance that people place on RFID in their daily lives. Only a few - 19% - rated RFID as being important for

their lives today. The rest responded either neutrally or did not see any relevance for RFID (Figure 6).

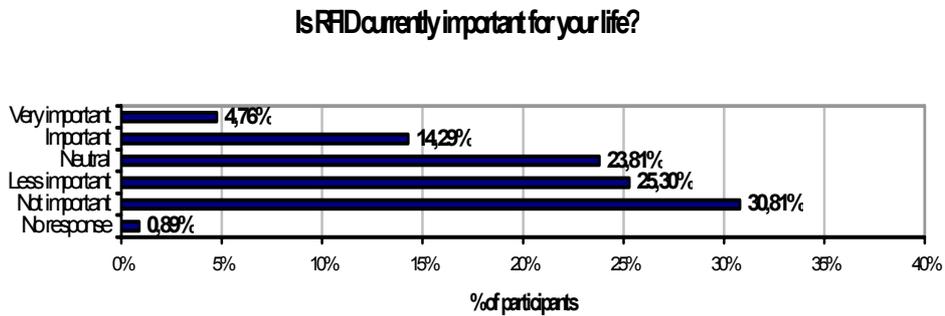


Fig 6: Importance of RFID in daily life

A possible explanation is that people do not see any benefits in RFID or they have not had any real contact with RFID, or finally there is currently no application that would add substance to people’s lives. RFID functions within the context of ubiquitous computing that is often times associated with control or perceived control. Research in this area on perceived control [GüSp05] exists in order to increase the acceptance of RFID applications. In our study, 47% strongly agreed that RFID would give others control over their lives, however, this also assumes that the people responding have enough information or experience with the technology to make an accurate assessment – which is not the case according to their own evaluations of how well they are informed (Figure 7).

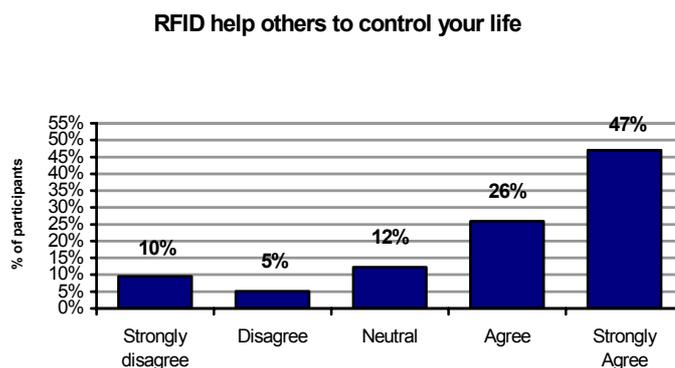


Fig. 7: RFID and control

To get a more detailed view of people’s opinion of the technology, we asked if RFID has already improved the life of the participant personally. While 63% did not see any personal

benefit yet, 10% were neutral. Overall, 19% saw a personal benefit or could at least imagine a use for RFID, 7% strongly agreed that RFID has benefits (Figure 8).

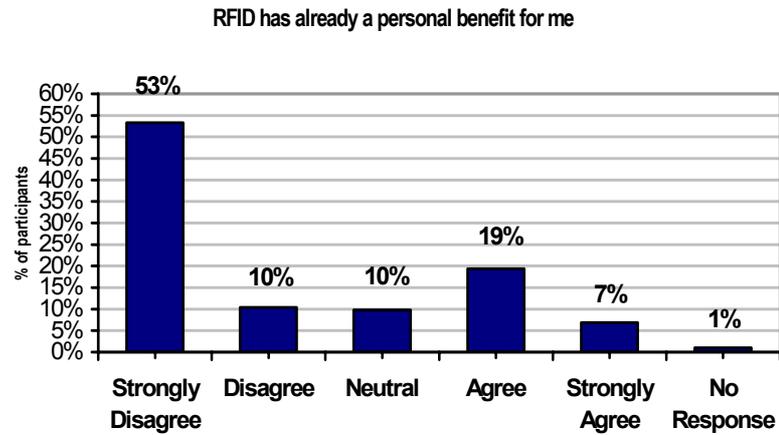
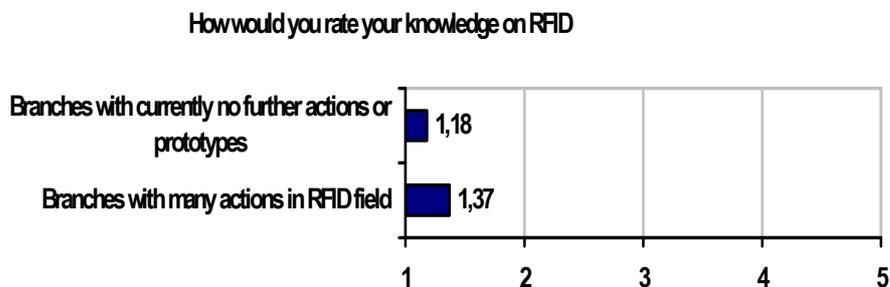


Fig. 8: Personal benefit of RFID

Another question concerned RFID and its influence on people lives (in whatever form). 52% of the answers were neutral, 15% of the participants assumed negative to very negative changes, and 20% rated the change rather positively. Based on these descriptive data, the picture of RFID that people had at the trade fair was rather negative. People are not necessarily informed about RFID, but in most cases rate RFID in exactly the same way that it is discussed in the media.

Hypothesis Results

The given results are possibly shaped by the position and or by the area in which the person works. Especially in terms of the personal sphere the answers from people in high or low positions were not significantly different. Contrary to that result however, people working in areas that have already seen prototypes and actions with RFID - such as information technology or retail - rate the degree to which they are informed higher than people working in other branches. (1= low level of knowledge in RFID; 5= high level of knowledge in RFID)



H1 assumed that people in different branches value the degree to which they are informed differently. This hypothesis proved correct as well as for H2, where we postulated that neither gender nor branch plays a role in the current evaluation of RFID.

Other Studies

Other studies in the area of RFID also provide rather similar findings. A recently conducted study by Knebel et. al. deals with the Chief Information Officer's (CIO) perspective on RFID [KnLK06]. Among all participating companies the dissemination of knowledge about RFID was very low. In addition, many information technology decision makers have heard about and are interested in the technology but are still far from actually implementing it. The study also points out the oppositional behavior of many CIOs. On the one hand, companies expect their RFID budgets to rise over the next several years. But the high-level concepts discussed in the media or in consulting have not yet found their way into day-to-day reality of companies. Participants rated their knowledge about RFID as good or very good (42%), whereas only a small number have really ever implemented a RFID system for their company (7%). Considering this fact, CIOs think they know RFID, but the practical experience is still rather low. Nevertheless, industrial applications can be more easily programmed compared to closed loop systems in machine-based scenarios. Here questions about privacy play a less important role than in consumer markets where open loop and people-based is the rule [KnLK06]. Another study was conducted in 2004 by CapGemini on the consumer's perception of RFID [Park04]. In contrast to what media commentators often quote, only 10% of the people who have heard about RFID have an unfavorable perception. The top concerns were: "Consumer data used by third party" (69%), "Targeted more with direct marketing" (67%) and "Tracking of consumers via purchase" (65%). Yet again, most respondents assumed that RFID tags will be placed on all products within the next 5 years, and only 3% said RFID will not become reality. The conclusion also clearly shows that while not many people know about RFID, they are still very much interested in the technology. A recommendation that came out of the findings centered around disconnecting RFID from the broader privacy debate [Park04]. Another study in 2005 [CapG05] shows similar results, e.g. "Helping consumers to form a positive impression of RFID by debunking myths" might be a way to educate the consumer [CapG05].

5 Discussion: Information as a Key to Successful RFID Implementations

Sources of information, such as Google Trends, show that RFID has the potential to be a “disruptive technology” but so far this is not a reality. Existing applications are limited to logistical tracking and tracing, signifying that only a few people have actually had first-hand experience with it. Our initial findings lead us to believe that the image of RFID is based more on misinterpreted facts or preconceptions than on the facts themselves. The benefits of the technology are not noticeably tangible. Even at large trade fairs where the participants include a large proportion of innovators and early adopters, one would expect the leaders at shaping opinion to give a more positive view. This might explain the reason behind the massive negative media hype and the rather low popular acceptance of RFID: if not enough positive information is available in mass markets, then the networking effects cannot be realized. The chain of communication is disconnected. Although many prototypes exist in industry segments as mentioned above, broad acceptance is not simply a question of handling privacy issues but rather one of communication and experience. Most of the participants in our study rated RFID as either dangerous (control) or irrelevant to their companies and lives. Nevertheless everyone saw a positive perspective or at least agreed on a broad future for RFID. In order for RFID to bring more benefits to companies and people, the industry needs to lay the foundations for more showcases and prototypes that can be tested and experienced by the consumer and establish more effective communication. In branches where prototypes have already received media coverage, people consider themselves to be aware of RFID and value the technology more positively. From the perspective of the individual, everyone rates RFID the same. This means that additional consumer prototypes would help people understand the technology better and lead therefore to more useful discussions. The successful marketing of RFID depends primarily on getting people involved in using the technology. To facilitate the communication categorization of potential RFID applications we propose the following simple matrix of RFID application types with a different emphasis on application communication:

Open loop vs. closed loop: It is necessary to define open and closed loop since the terms are used in several different disciplines with differing meanings (e.g. in electrical engineering or sports science). A closed loop RFID application is defined as an application in which the RFID tag is only used within a single system - if the system feeds information on the tag back into itself [WCPR06] and the tag is not leaving the process of the system in which it was initially used. By comparison, an open loop application is considered to be an application in which the

RFID tag and its information can be used outside of the initial system in one or more systems. According to Fleisch et al. most applications today begin with closed loop systems because of the clear cost and benefit calculations [FICD05]. The degree of integration varies of course between open and closed loop systems with open loop systems more likely to bring RFID in contact with users.

Machine-based or people-based applications: In the case of the proposed matrix we define machine-based applications as applications that do not include interactions with people on a user level. This counts for most of the business-to-business applications where the purpose of the application is to use the information on the tag for controlling either other machines or for using it as an information source for other applications. People-based applications are primarily designed for direct person-to-person interaction and built for end-users and consumers. Examples are the FIFA World-Cup 2006 tickets [FIFA04] or the aforementioned Address Book Desk by Timo Arnall [Arna05]. The reason for this distinction has to do with the reactions in the media regarding privacy and data security questions. As soon as people-based applications are considered, discussions arise. In the case of the World-Cup tickets the usage of RFID led from negatively critical articles [FBUD '06] to the public call for the Chaos Computer Club (CCC) to send in the World-cup tickets for CCC intense research [CCC06]. To sum up, potential applications can be classified in the table as shown below in Figure 9.

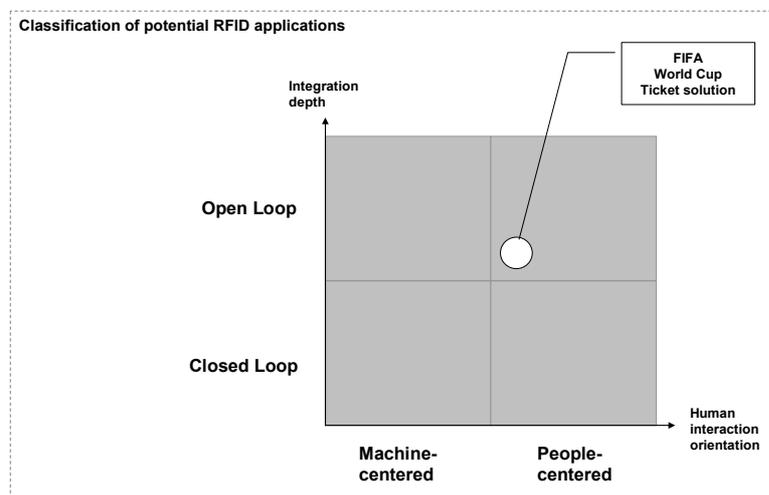


Fig. 9: Table for RFID applications (Integration depth, see [Fleisch, Christ et al. '05])

With most applications in operational business, identifiable in the first quadrant of the table, discussions of privacy issues are not in the fore. However, when it comes to open-loop, people-based systems, discussions will arise and emphasis should be on valid information or better experience with prototypes given to the users before introducing large scale applications.

6 Conclusion and Limitations

The theoretical framework shows that the adoption of innovations relies on positive discussions and useful information. The discussions concerning RFID center around tracking & tracing issues, supply chain support, and privacy and security issues. Privacy is definitely a very important topic, but should be discussed when people have had the chance to experience and learn about the technology with showcases and hands-on prototypes. If using the kind of model that Beckwith [Beck03] proposes according to Adams, information on the following topics should be given to the user: *information receiver* (who will use or have access to the data?), *information usage* (how will the information be used?) and *information sensitivity* (how sensitive is the data?). If this is done and shown, people will have the opportunity to use the application and decide freely if they like it, thereby giving RFID the chance to enter the consumer market place and cycle back on a major scale into industry and the world of small-and-medium enterprise. The acceptance of the technology and the potential benefits that come with it depend on information and perception. The study makes clear that people do not necessarily know enough about RFID to rate it properly. Using many more prototypes to show the real value of RFID in end-consumer areas would ultimately lead to a greater overall acceptance of the technology. This would also lead to more understanding and less fear of RFID-use within operational businesses. Even in his 1993 paper on Ubicom, Mark Weiser states that the establishment of “working prototypes of the necessary infrastructure in sufficient quantity to debug the viability of the systems in everyday use” is crucial to the research of Ubiquitous Computing. Even today’s surveillance technologies are more accepted because of their perceived benefit to people. People claim that they feel safer with surveillance cams - although statistics prove that surveillance cameras do not significantly prevent crimes in monitored areas [Wehr00]. In terms of World-Cup tickets for example, at the writing of this paper, no data was found on the chip that could have been used in a malicious way [CCC06]. If the benefits of RFID can be demonstrated to consumers with valid and fun prototypes, the overall perception might be better and the path will be clearer for its implementation in many areas. Many approaches allow the creation of fun prototypes, such as free RFID-kits given to schools and universities or application development competitions. The industry has the chance to and should sponsor and develop such events and actions. For researchers it is getting more important to conduct research in the context of real life of users with appropriate prototypes.

The nature of our research has had certain limitations. The research results are not necessarily representative since they are based on a random sampling of people without prior assessment of their qualifications. Compared to the total number of visitors to the trade fair, the questioned sample was too small to generate a representative number of interviews. Secondly, most questions were answered from a subjective point of view, e.g. for the question concerning one's knowledge of RFID, it isn't clear how to compare the individual assessment values since each participant may rate his or her knowledge with a different scale. In this way the exact dimensions of a particular aspect remain somewhat unclear - the use value of an application is a subjective determinant that can't be easily compared. Thus the study should be seen as an initial investigation that needs further evaluation at other fairs, public spaces or showcases. The usage of Google trends is not necessarily accurate and should be considered an approximation to the conclusions.

Moreover every RFID application must surely have an appropriate benefit to be for any use to the consumers. The usage of RFID for the sheer existence of the technology prevents an evolution of useful systems in the sense of the original vision. We emphasize a discussion on data security, but also see the chance to educate people on the benefits of this technology experiencing the technology. More grassroots projects and toolkits are a proposed solution.

References

- [AbMy00] *Abowd, G. D. and E. D. Mynatt.*: Charting Past, Present and Future Research in Ubiquitous Computing. In: ACM Transactions on Computer-Human Interaction 7 (2000) 1, P.30.
- [Arna05] *Arnall, T.*: The address book desk. <http://www.elasticspace.com/2005/12/address-book-desk>., Retrieved July 7, 2006
- [Beck03] *Beckwith, R.*: Designing for Ubiquity: The Perception of Privacy. In: Pervasive Computing (2003) 3, P. 40-46.
- [BCLM05] *Bohn, J., Coroama, V., Langheinrich, M., Mattern, F., Rohs, M.*: Social, Economic, and Ethical Implications of Ambient Intelligence and Ubiquitous Computing. Ambient Intelligence. In: W. Weber, J. Rabaey and E. Aarts, Springer-Verlag (2005), S.5-29.

- [CapG05] *CapGemini Consulting.*: RFID and Consumers.
<http://www.us.capgemini.com/DownloadLibrary/files/CapgeminiEuropeanRFIDreport.pdf>. Retrieved October 18, 2006
- [CCC06] *Chaos Computer Club.*: Scannt Eure Tickets, ehe es wer anders tut!
<http://www.ccc.de/updates/2006/rfid-tickets-fuer-berlin?language=de>. Retrieved July 7, 2006
- [Chri97] *Christensen, C. M.*: The Innovator's Dilemma. Harvard Business School Press 1997.
- [Chri03] *Christensen, C. M.*: The Innovater's Solution. Creating and Sustaining Successful Growth. Harvard Business School Press 2003
- [DaKa73] *Darby, M. R., Karni E.*: Free Competition And The Optimal Amount Of Fraud. In: The Journal of Law and Economics (1973) 16: P. 67-88.
- [Davi89] *Davis, F. D.*: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. In: MIS Quaterly (1989) 13. P. 319-339.
- [DaVe04] *Davis, F. D., Venkatesh, V.*: Toward preprototype user acceptance testing of new information systems: implications for software project management. In: IEEE Transactions on Engineering Management 51 (2004) 1.
- [FIFA04] *FIFA.*: 2006 FIFA World Cup OC agrees deal with CTS.
<http://fifaworldcup.yahoo.com/06/en/040816/1/203i.html>. Retrieved July 7, 2006
- [FKTT06] *Fine, C., Klym, N., Trossen, D., Tavshikar, M.*: The Evolution of RFID Networks. Value Chain Dynamics Working Group (VCDWG) Cambridge, MIT Sloan School of Management - MIT Communications Futures Program (CFP) Research Network Cambridge University Communications, Massachusetts 2006.
- [Flei04] *Fleisch, E.*: "RFID – Technologie: Neuer Innovationsmotor für Logistik und Industrie?", 2004.
- [FICD05] *Fleisch, E., Christ, O., Dierkes, M.*: Die betriebswirtschaftliche Vision des Internets der Dinge. In: Das Internet der Dinge. E. Fleisch and F. Mattern, Springer, 2005.
- [FIMa05] *Fleisch, E., Mattern, F. (Eds)*: Das Internet der Dinge - Ubiquitous Computing und RFID in der Praxis: Visionen, Technologien, Anwendungen, Handlungsanleitungen, Springer, 2005.
- [FBUD06] *FoeBUD.*: WM-Tickets absurd. <http://www.foebud.org/rfid/wm-tickets-absurd>. Retrieved July 7, 2006.
- [GeKC04] *Gershenfeld, N., Krikorian, R., Cohen, D.*: The Internet of Things. Scientific American (2004). P. 76-81.
- [GüSp05] *Günther, O., Spiekermann S.*: RFID And The Perception of Control: The Consumer's View. Communications of the ACM 48 (2005) 9. P. 73-76.

- [Hain05] *Hainbuchner, C.*: Technology Acceptance of Complex Products and Systems - The Case of Terrestrial Trunked Radio (TETRA). Dissertation. Vienna, 2005.
- [Kaas05] *Kaasinen, E.*: User acceptance of mobile services – value, ease of use, trust and ease of adoption. Dissertation. Espoo, 2005.
- [KnLK06] *Knebel, U., Leimeister, J. M., Krcmar, H.*: Strategic Importance of RFID – The CIO Perspective; An Empirical Analysis in Germany. Americas Conference on Information Systems, Acapulco, Mexico, 2006.
- [Mano01] *Manovich, L.*: The Language of New Media. MIT Press, Cambridge, Mass., 2001.
- [Matt02] *Mattern, F.*: "The Vision and Technical Foundations of Ubiquitous Computing." Retrieved May 15, 2005.
- [Matt03] *Mattern, F.*: Die technische Basis für das Internet der Dinge. Institut für Pervasive Computing. Zuerich, 2003..
- [Moor02] *Moore, G. A.*: Crossing the Chasm, HarperCollins Publishers, 2002.
- [Norm99] *Norman, D. A.*: The Invisible Computer. MIT Press, Cambridge, MA, 1999.
- [Park04] *Parkinson, J.*: RFID: U.S. consumer research findings., <http://www.ftc.gov/bcp/workshops/rfid/>. Retrieved July 17, 2006.
- [Roge95] *Rogers, E. M.*: The diffusion of innovations. New York Free Press, New York, 1995.
- [Shef04] *Sheffi, Y.*: RFID and the Innovation Cycle. In: International Journal of Logistics Management, International Logistics Research Institute 15 (2004) 1.
- [Stoc48] *Stockman, H.*: Communication by Means of Reflected Power. IRE, 1948.
- [Want06] *Want, R.* An Introduction to RFID Technology. In: Pervasive Computing (2006) 6. P. 25-33.
- [Wehr00] *Wehrheim, J.*: CCTV - Closed Circuit Television. In: Forum Wissenschaft (2000) 2.
- [Weis91] *Weiser, M.*: The Computer for the 21st Century. In: Scientific American 265 (1991) 3: P. 94-104.
- [Weis93] *Weiser, M.*: Some Computer Science Issues in Ubiquitous Computing. In: Communication of the ACM 36 (1993) 7.
- [WiDe04] *Wilding, R. Delgado, T.*: RFID Demystified: Company Case Studies. Logistics & Transport Focus 6 (2004) 5. P. 31-42.
- [WCPR06] *Wray, R., Chong, R., Phillips, J., Rogers, S., Walsh, B.*: A Survey of Cognitive and Agent Architectures. <http://ai.eecs.umich.edu/cogarch0/common/prop/closedloop.html>. Retrieved July 9, 2006

Perceived Usefulness of RFID-enabled Information Services – A Systematic Approach

Hanna Krasnova¹, Matthias Rothensee, Dr. Sarah Spiekermann

Institute of Information Systems
Humboldt-Universität zu Berlin
10178 Berlin

krasnovh@wiwi.hu-berlin.de, rothensm@cms.hu-berlin.de, sspiek@wiwi.hu-berlin.de

Abstract

Even though RFID technology is currently gaining importance mainly in logistics, usage areas, such as shopping or after-sales enhancements beyond the supply chain are envisioned. Yet, while RFID hits the street it is questioned if it may undermine one's privacy while providing few customer benefits. Meeting this criticism this paper investigates RFID-enabled information services and the drivers of their usefulness for consumers. The article claims that the more risk one associates with a product the more benefit from RFID-enabled information services is perceived. We show empirically that the nature of product risk provides a useful framework to decide on the types of RFID information services a marketer should offer to create RFID usefulness perceptions and increase technology acceptance.

1 Introduction

RFID (Radio Frequency Identification) is becoming an important factor in logistic chains. Wireless object-to-object identification without a line of sight, which promotes increased transparency in supply chains, is often mentioned as its main benefit. However, the uses of RFID are not only limited to industrial applications. In fact, industry would be happy to see RFID chips go beyond the supply chain and create additional end-user markets. Slowly, retail leaders and brands are making first moves in this direction. METRO Group, for example, has opened an

¹ authors in alphabetic order

Innovation Centre in Neuss (Germany) where myriad after-sales and shop floor solutions are displayed, many of which were first introduced to the public at the CeBIT tradeshow in 2006.

However, many consumers seem to believe that RFID chips have the potential to invade their privacy; they fear to lose control and therefore threaten to reject the technology [GüSp04, 245]. For this reason Benetton's attempt to introduce RFID on clothing was despised and hastily cancelled [Albr03]. On the other hand, some RFID applications already introduced to the market seem to be very welcomed and have never been subject to any privacy critique. These include ski-passes, entry tickets to major events and public transportation, tags used in sports clubs, during marathon events or in public libraries. This contradiction in public feelings has been one of the main impetuses for this study. The main goal of this paper is to determine which RFID-enabled services consumers find useful and whether there are factors systematically influencing usefulness perceptions. To achieve this goal we focus on RFID-enabled information services in contrast to tracking and access services. Information is relevant for consumers prior to purchase, because it allows for reducing purchase uncertainties. RFID allows for the provision of richer information than what is available on packaging surfaces (or using a bar code) as well as qualified recommendations. Therefore, RFID may be used as an important lever for consumers in order to make purchase decisions in a more informed way. The question arises, however, where in particular consumers will appreciate RFID for this purpose.

The present article is structured as follows: first, we summarize product risk theory and then relate this to RFID-enabled information services. Using this theoretical basis, we then present an empirical study we conducted to investigate where RFID-enabled information services can potentially be interesting for consumers. We analyze whether perceived product risk can systematically explain differences in consumers' perceived usefulness of RFID. We also add citations throughout the text from a number of focus groups we conducted on the subject at the Humboldt University in winter 2006. A detailed discussion of the obtained results, conclusions for the industry and study limitations are provided at the end of the paper.

2 Theoretical Background

2.1 Product Risk and Information Search

Marketing theory has accumulated a large number of models regarding the way consumers make product choices and factors which are important for their purchase decisions [Chis94, 193]. One of the most popular models, first developed by [Cunn67], [Baue60] and [CoxD67], suggests that perceived risk attributed to products influences the way consumers make their product choices and seek product information prior to purchase. [Cunn67, 84] suggests that perceived product risk can be measured using a multiplicative relation of two elements: uncertainty associated with a product purchase (measured as the probability of loss) and consequences (measured as the importance of the respective loss). In turn, [MuEn86, 31; based on JaKa72] consider five product risk dimensions:

- „Financial risk – the risk that the product will not be worth the financial price;
- Psychological risk – the risk that a poor product choice will harm a consumer’s ego;
- Physical risk – the risk to the buyer’s or other’s safety in using products“;
- Functional risk – the risk that the product will not perform as expected;
- Social risk – the risk that a product choice may result in embarrassment before one’s friends/ family/ work group“.

The main inference from the perceived risk model is that consumers will try to minimize their purchase risks before they buy something and systematically seek for information that reduces purchase uncertainty [Kosc97, 1658]. Therefore, most of the literature on consumer behavior [KASW03, 325; AnRa98, 256; ScKa97, 189] highlights information search about a product as a risk-reducing strategy.

Information search and availability is also the focus of another well-known product classification suggested by [DaKa73; Nels70] and then further developed by [WeAd95a; WeAd95b]. This classification categorizes products according to the amount of information available to the consumer prior to purchase. It distinguishes between search, experience and credence goods:

- Search goods are goods with dominant search qualities, „which can be fully evaluated by the consumer through product inspection or information collection be-

fore purchase“. Examples include clothing, cell phones, music CDs, apparel, furniture, shoes, memory stick;

- Experience goods are goods with dominant experience qualities, „which can be evaluated by the consumer only after purchase“. Examples include boxed products, medicines, garments, personal items, wine, liquor, food, watch;
- Credence goods are goods with dominant credence qualities, „which can be fully evaluated by the consumer neither before nor after the purchase“ [WeAd95b, 54, translated by the author]. Examples include “vitamins”, bio-labeled food, “water purifier” [GiSK02].

This classification makes plain that uncertainties related to a product may reach beyond the immediate sales situation and into after-sales scenarios where consumers use and experience the product they bought. Here, equally, RFID-enabled information services may play a role.

2.2 RFID Information Services and the Role of Product Risk

Generally, three groups of RFID-enabled consumer services can be distinguished: information services, tracking services and access services. For reasons of complexity reduction the present paper concentrates exclusively on *RFID-enabled information services* which can in turn be subdivided into *information on authenticity, recommendations, warranty information and additional information*.

Ensuring product *authenticity* is vital for many products. In order to „combat counterfeiting“ [Pfiz06] has started tagging its Viagra products in the USA in 2006. Now a consumer can check the authenticity of Viagra pills directly in the pharmacy or using a smart medicine cabinet at home to make sure his health will not be damaged by a fake product, thus reducing physical/health risks. As one participant in a focus group noticed: „*If there is a possibility to make them (medications) safer then I find it good*“. In shopping environments one can check the authenticity of packaged meat or medication using e-info-points by simply placing a product close to an info-reader. Repair shops for expensive goods (musical instruments, watches) can also be supplied with authenticity check devices in order to ensure that a product is not claimed stolen. A portable home reader can ensure the original taste of a French wine ordered online. Financial and functional risks for future owners could thus be reduced.

Recommendations on compatibility or suitability can be provided in various sales contexts as well as after-sales settings. Smart medicine cabinets, smart change closets and mirrors or smart fridges are examples. These applications are supplied with readers registering RFID-tagged products, e.g. medications, clothing or food placed into them. They then „recommend“ buying well fitting clothes, removing expired medicines, etc. Thus, by following electronic wardrobe recommendations one may be able to diminish social risk by ensuring that a selected jacket suits to formal office style and current fashion.

RFID-enabled *warranty* includes expiration dates or information relevant for subsequent repair, resale or recall. For example, RFID-enabled warranty for a laptop ensures that there will be no bad surprises if a customer accidentally loses the receipt, thus reducing financial risk. Equally, RFID pro-actively informs that clothes may be returned to the store until a certain date. Consumers may, as a result, feel a relief in social and psychological purchase risks.

Finally, RFID can provide *additional information* on products. The ability to easily access electronic manuals and set-up support as well as upgrades on the Internet might reduce the functional risk of a complex technical product (such as an espresso machine or a copy machine) by ensuring that the right sequence of actions will be undertaken. Naturally, users will be more likely to use such RFID-enabled features as they do not require any effort from their side as opposed to the bar code, where a line of site is important.

All in all, by using RFID-enabled information services a consumer can alleviate certain risks attributed to a product. On this basis we now narrow down and explore *how* consumers evaluate the usefulness of RFID-enabled information and if perceived risks associated with a product can have an effect on the perceived usefulness of RFID-enabled information services.

3 Method

3.1 Procedure

An online questionnaire was passed to 149 study participants in order to collect empirical results on the relation of risk and perceived RFID usefulness. The recruiting of the study subjects was done as a convenience sample by sending an invitation to a university mailing list. Each participant was rewarded EUR 10 for a completed survey. The questionnaire was accessible online from March 11th 2006 until April 2nd 2006.

Out of the 149 participants 63% (93) were female and 37% (55) were male. Most of the participants were young people (73% were in the age of 20-29 years) and had either a university (35%) or a high school degree (59%). They were also familiar with technology since 129 (88%) were preparing half or more of their work with computers.

3.2 Products Tested and Questionnaire Structure

In order to test the relationship between perceived product risk and usefulness of RFID the questionnaire always related to a constant selection of nine products. These products were chosen with a view to the classification into search, experience and credence goods [WeAd95a; WeAd95b]. Equally, the products were expected to display different levels of product risk. Table 1 gives an overview of the chosen products.

Classification category	Selected products
Search goods	Alarm clock, digital camera, gloves
Experience goods	Anti-dandruff shampoo, espresso machine, migraine pills
Credence goods	Bio-meat, vitamins, diamond

Table 1: Selected products for each category

The questionnaire consisted of three question blocks. In the first block we checked if the participants would categorize the products as theoretically expected. The participants had to rate three statements on a 5-point Likert scale from very improbable (1) to very probable (5): The first question tested the search qualities attributed to products: *„Before purchasing product X I can fully judge on important quality characteristics“*. The second question aimed to test the experience qualities of the products *„After purchasing and using product X I can fully judge on quality characteristics important for me“*. The third negative question tested the credence qualities of the products: *„Even after purchasing product X I cannot fully judge on quality characteristics important for me“*.

The second part of the questionnaire was dedicated to the identification of the levels of perceived risk associated with the selected products. The evaluation was done in two steps. First the participants were asked to put themselves in the position of a potential buyer and from this perspective to rate the probability of financial, functional, psychological, social or physical damages (*„How probable is it that the purchase of the product X will have negative financial consequences for you?“*) from very improbable (1) to very probable (5). At the second step they

had to assess the amount of the resulting damage for them („How high do you estimate financial loss...?“) from very low (1) to very high (5).

The third part of the questionnaire dealt with measuring the level of perceived usefulness of four RFID-enabled information services (authenticity, recommendations, warranty/expiration dates and additional information) for the selected products both before and after purchase. The participants were asked to rate the usefulness of these information services from not useful (1) to very useful (5) for each product first before purchase and then after purchase. The questions were stated in the following way: first a specific RFID-enabled information service was presented to the participants (e.g. „imagine you could electronically get additional recommendations for a product, i.e. recipes for food, right on the shopping floor“), then the way how to access this service was described (e.g. „you simply hold a product to your shopping cart and read information on the screen“) and finally the participants were asked „how useful“ they would find this RFID-enabled information service.

Even though the participants were asked about the usefulness of each information service both before and after purchase, only the results on before purchase evaluations will be considered here for the sake of brevity.

4 Results

4.1 Product Classification in Terms of Perceived Risk

Assessment of the perceived risk profiles for the selected products was carried out using the multiplicative relation described above, thus the least risky product would get the value of 1 and the most risky one the value of 25. Table 2 summarizes the obtained results averaged across all the participants. The values of the three most risky products in each individual risk category are selected in bold.

Averaged risk levels indicate that financial and functional risks were the most dominant ones across all products. These risks were evaluated as especially high for such complex technical goods as an espresso-machine or a digital camera and a diamond. Overall, these goods illustrate the same perceived risk profile: the highest three levels of financial, functional, psychological and social risks and very low levels of the physical/health risk. On the basis of their risk profile we will generally call this group of products „complex“ goods. Such products as vitamins, bio-

meat and migraine pills excelled in the highest physical/health risk. For this reason we will refer to them as „high physical/health risk“ products. Alarm clock, anti-dandruff shampoo and gloves were perceived as relatively risk free products and can be labeled as „simple“ goods. Overall, the *social risk* played only a marginal role.

Products	Financial Risk	Function. Risk	Psycho-logical Risk	Physical Risk	Social Risk	OPR* _j
Diamond	17,3	11,3	12,0	3,2	8,5	52,3
Espresso machine	14,8	13,2	11,0	6,0	7,1	52,1
Digital camera	13,8	11,7	12,1	3,3	6,6	47,5
Vitamins	6,6	9,0	6,8	13,3	4,3	40,0
Bio-meat	5,6	7,2	5,0	12,0	4,8	34,6
Migraine pills	5,0	7,8	5,1	14,7	4,0	36,6
Alarm clock	4,3	6,5	5,0	3,3	3,3	22,4
Anti-D. shampoo	4,0	6,8	5,8	8,6	3,9	29,1
Gloves	4,0	4,8	5,4	4,0	4,1	22,2
Averaged risk levels						
	8,4	8,7	7,6	7,6	5,2	

Table 2: Summary of perceived risk level results for the selected products
 *OPR_j - overall perceived risk for product j is calculated as a sum of five individual risk values attributed to product j.

Finally, throughout all risk types (except for the psychological risk) we have got the same ranking order with respect to the perceived risk level for the three products that best represent the respective product categories: vitamins as a typical credence good are always followed by anti-dandruff shampoo as a typical experience good, which is always followed by gloves as a typical search good. Seen that the three products can be purchased at comparable prices this risk correspondence to the product classification is interesting and helpful for subsequent generalizations. In Table 3 we have averaged all risk values for all products in a respective classification category. As one can see there is – beyond individual products - a general tendency for the highest risk to be attributed to credence goods, followed by experience goods and then search goods.

Product categories	Products in the category	Averaged level of risk for a product category	SD*
Search goods	Alarm clock, dig. camera, gloves	6.1	2.0
Experience goods	Espresso machine, migraine pills, anti-dandruff shampoo	7.8	2.4
Credence goods	Bio-meat, vitamins, diamond	8.4	2.5

Table 3: Summary of risk level results averaged over product categories
 *SD =standard deviation

4.2 Perceived Usefulness of RFID-enabled Information Services

At this stage of the questionnaire the participants were asked to rate the usefulness of four RFID-enabled information services. Averaged results for each product and each RFID-enabled information service before purchase are presented in Table 4. Here, the top three perceived usefulness values for the products are selected in bold for each individual information service.

As follows from Table 4, *authenticity* was mainly perceived as the most useful information service out of the four RFID-enabled information types and this is true for almost all products before purchase.

Products	Authenticity	Recommendation	Additional Information	Warranty Expiration Dates	Sum of usefulness values	The most useful RFID-enabled information service
Dig. camera	4,3	3,7	3,9	4,2	16,0	authenticity
E-machine	3,9	3,3	3,7	4,2	15,1	warranty
Migr. pills	4,0	3,5	3,4	3,4	14,3	authenticity
Vitamins	4,0	3,4	3,3	3,3	14,1	authenticity
Bio-meat	4,3	3,3	2,6	3,6	13,8	authenticity
Alarm clock	3,0	2,4	2,9	3,5	11,8	warranty
Shampoo	3,4	3,1	2,7	2,7	11,8	authenticity
Diamond	3,9	2,5	2,1	2,9	11,5	authenticity
Gloves	2,8	2,3	2,0	2,7	9,7	authenticity
Averaged usefulness levels						
	3,7 (quite useful)	3,1 (fairly useful)	3,0 (fairly useful)	3,4 (fairly/quite useful)		authenticity

Table 4: Perceived usefulness of RFID-enabled information services for the products before purchase

The participants rated authenticity higher than 3,0 on a scale from 1 to 5 for most of the products. Foremost, the participants were interested in the authenticity information for high physical/health risk products such as bio-meat, migraine pills and vitamins. Authenticity information was also important for complex goods, such as digital cameras, espresso machines or diamonds. The participants equally rated the usefulness of *warranty and expiration dates* higher than 4,0 for complex technical products and higher than 3,3 for high physical/health risk products. Some value was attributed to RFID-enabled *recommendations*, with the answers fluctuating around the level of 3,0 with complex technical goods and medications having been placed at the top of the perceived usefulness of recommendations list. However, for technical products, such as digital cameras, espresso machines or alarm clock recommendations were the least useful RFID-enabled information service of all four. For products from a personal use domain, such as

anti-dandruff shampoo, migraine pills and vitamins, it was the second most important information. *Additional information* was not popular among the participants with the attributed usefulness values surpassing the level of 3,0 only for complex technical goods and medications. For such products as bio-meat, migraine pills, diamond and gloves it was the least useful information service of all four.

Column six in Table 4 shows summed averaged perceived usefulness values of the four information services for every product sorted in a descending order. As one can see, RFID-enabled information services are valued highest when it comes to complex technical goods (here digital camera and espresso machine). The second highest valuation of the information is attributed to products high on physical/health risks (here migraine pills, vitamins and bio-meat).

4.3 Multiple Regression Analysis

Sections 4.1 and 4.2 have prepared the ground for investigating in a next step whether perceived risks associated with a product can systematically affect perceived usefulness of RFID-enabled information services. This effect can be evaluated by looking at the multiple regression coefficients (b_{ijt}) resulting from a multiple regression analysis, where perceived usefulness of an RFID-enabled information service (U_{jt}) for a product is a dependent variable and the five perceived risks (R_{ij}) associated with the product in question are independent variables. Equation (1) summarizes this relationship.

$$U_{jt} = \sum_{i=1}^n b_{ijt} R_{ij} \quad (1)$$

where

U_{jt} - perceived usefulness of RFID-enabled information service t of product j before purchase;

R_{ij} - perceived risk of dimension i attributed to product j ;

b_{ijt} - regression weight of risk dimension i attributed to product j on the perceived usefulness of RFID-enabled information service t ;

n - risk dimensions.

Since we collected measurement points from nine products and four RFID-enabled information services, our analysis involved 36 regression models. As expected not all of them resulted in a high level of explained variance since product risk can only be *one* variable driving usefulness perceptions. However, we did find that regression equations on vitamins, migraine pills and

anti-dandruff shampoo have relatively high levels of R^2 with these models explaining 10-17% of the usefulness variance of RFID-enabled information services. Furthermore, risk had a slightly higher explanatory value for the usefulness of RFID-enabled recommendations before purchase than after purchase. This makes sense since risk-reducing recommendations before purchase are more valuable to consumers than when the products have already been bought. This difference is however statistically non significant.

Looking into the detail of regression functions 30% of b_{ijt} -coefficients in the regression equations were significant. However, this percentage number is relatively less important than an understanding *where* beta coefficients were indeed significant. Here an interesting pattern arises. First, functional risk seems to be the main driver of RFID usefulness perceptions. 41% of all significant b_{ijt} -coefficients refer to functional risk. The next highest scoring risk dimension is physical risk. 26% of significant b_{ijt} -coefficients fall into this category. In comparison, social (17%) and financial (13%) risks seem negligible. And psychological risk adds no explanatory value at all. Short: RFID-enabled information services are valued across all products in order to reduce functional risk. Physical risk also plays a role, even though few products imply such a risk. Once a product does imply physical risk from a consumer's perspective almost any RFID-enabled information type is welcomed. Finally, social risk has an impact on RFID usefulness perception. It seems to be important when it comes to recommendations about a product or checking for warranty information. Table 5 gives an aggregated view to support this analysis by providing the average of b_{ijt} -coefficients (\bar{b}_{it}) over the nine products for RFID-enabled information service t and risk dimension i before purchase. The values of \bar{b}_{it} higher or equal to 0,1 are selected in bold. This aggregated view shows that warranty information takes a lead among the various available information types, closely followed by information on authenticity.

Information service	Financial risk		Functional risk		Psychological risk		Social risk		Physical risk	
	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
Authenticity	0,05	0,13	0,18	0,09	-0,01	0,07	0,01	0,11	0,11	0,06
Recommendation	0,02	0,18	0,08	0,07	-0,01	0,10	0,11	0,06	0,06	0,12
Additional info.	0,15	0,10	0,05	0,08	-0,04	0,10	0,05	0,10	0,13	0,13
Warranty	0,04	0,12	0,10	0,09	+0,04	0,07	0,10	0,07	0,11	0,10

Table 5: Regression coefficients (\bar{b}_{it}) and standard deviations for RFID-enabled information services and risk dimensions, averaged over all products

5 Discussion

The data collected give us some interesting insights into the areas where RFID-enabled information services are useful from a consumer perspective. Three angles can be used to discuss the findings: the type of product, associated product risk as well as the type of information provided by RFID.

The analysis of product perceptions and averaged service usefulness has shown first that complex technical goods imply the highest usefulness for RFID-enabled information. This group of products is closely followed by sensitive products in the physical/health domain. For simple products RFID-enabled information services, in contrast, seem to be perceived as little useful by the majority of the study participants. Another angle to grasp this finding is to argue that there may be a slight tendency for preferring RFID-enabled information if experience or credence goods are concerned (in contrast to search goods).

The study of perceived usefulness of various RFID-enabled information services has also revealed that the participants were interested most in information on authenticity and warranty. Authenticity seems to be particularly important for health/physical risk products. The focus groups have shown that in case of high physical/health risk goods the participants equated product authenticity with safety. In this vein one focus group participant stressed that *„(fake medicines) have disastrous consequences for consumers“*. *Warranty/expiration information* was especially important for complex technical goods: *„...one can check on the personal computer if the promises of the producer are really true“*.

Finally, the analysis of risk dimensions and their explanatory value for the appreciation of RFID has shown that functional and physical risks are the most important drivers of information needs. This is a valuable insight, because marketers may want to start enhancing their information activities with the help of RFID for products in this particular area. An example could be to offer interactive manuals and clips that could be downloaded from the chip or accessed with the help of the chip's product code. Equally, safety related issues, such as the potential of food to trigger allergies, or the sources of meat and milk products may be appreciated by consumers. In contrast, RFID-enabled information was less appreciated to relieve psychological risk, even though RFID could be used to trigger quite a few services addressing the coolness and suitability of products. First prototypical implementations of RFID-enabled information services for consumers also focus in this area, such as the intelligent change closet offered by GERRY WEBER. Both accessibility and „habitude“ may offer an explanation for why the relief of psy-

chological risks through RFID may be less appreciated than the relief of functional or physical risks. Marketers already today put a great effort into reducing functional risk when offering manuals and how-to-use information. Equally, they provide information on ingredients or safety notices so that physical or health risks are addressed. Social risk is typically reduced today by interacting with store personnel. On the other hand, psychological risk is alleviated by internal mechanisms such as self-reflection and self-confidence. Thus, it is reasonable to expect that psychological risk will be most resistant to change by an external device or service, because the device or service cannot directly access these mechanisms. As one participant in the focus group put it: *„I don't know whether I need a smart wardrobe that makes suggestions to me, which I don't want at all. It goes too far for me. I see no sense in this. I have my own taste“*. Furthermore, the usefulness of RFID-enabled information displays is new and may not be appreciated as much due to its unfamiliarity to the user. It is what Donald Norman refers to as a „disruptive technology“ [Norm98, 231-247] and it lacks compatibility to traditional norms [Roge03], both arguments standing in the way of a rapid dissemination of a technology. This argument has consequences for the current work in so far as it emphasizes that the appreciation of RFID-enabled information in different risk contexts may be apt to change over time. Also, it cannot be excluded that it is differing between cultures. Therefore the study results should be considered applicable at the current point only to the German market as a relatively uneducated market as far as RFID is concerned.

However, in the current situation industry needs a way to systematically prioritize the introduction of RFID on products and it needs to know what to focus on when introducing RFID-enabled information services. For this purpose Table 6 provides some good guidance, summarizing the findings reported on in this article.

Risk dimensions	Potential risk relievers provided by RFID-enabled information services		
Financial risk	Additional info.		
Functional risk	Authenticity	Warranty	
Psychological risk			
Social risk	Recommendations	Warranty	
Physical risk	Authenticity	Additional info.	Warranty

Table 6: Perceived risks and their potential relievers for customers

6 Conclusion and Limitations

This article has shown how product risk can be used to provide for a systematic approach in offering RFID-enabled information services to a final consumer and thus drive usefulness perceptions and acceptance. The insights allow for some valuable recommendations for the industry. In their marketing agenda, companies should first of all emphasize the possibility to check warranty/expiration dates especially for complex technical and high physical/health risk products, such as electronics, medications or food. The possibility to access authenticity information should also be put in the forefront of advertising campaigns for high physical/health risk products, especially those susceptible to counterfeiting. Identification of the risk perceptions associated with a given product may generally be considered a strategy for companies to determine in which directions the technology should be advertised.

Yet, despite these interesting results, there are also some limitations to this research. First of all because recruiting of the participants for the questionnaire was done as a convenience sample, the sample is not representative. The study also involved a limited selection of nine products which does not allow for a broad generalization. Because of hypothetical nature of many survey questions, it could have been difficult for some participants to realistically assess part of the proposed situations. Furthermore, only for a small part of models the R^2 exceeded 10% of explained variance. Also, previous studies have shown that information search (shopping) is not always the most preferred method for consumers to reduce their risk [Rose67, 58]. Instead, product brand, specialized retailers and personal traits could be often more influential for product choice. Understanding this relative importance of future RFID-enabled information services in comparison other marketing activities is an interesting area of future research.

Bibliography

- [Albr03] Albrecht, Katherine: Press Release.
http://www.boycottbenetton.com/PR_030407.html, 2003-04-09, download on 2006-04-10.
- [AnRa98] Antonides, Gerrit; van Raaij, W. Fred: Consumer Behaviour. A European Perspective. John Wiley and Sons Ltd, Chichester 1998.

- [Baue60] Bauer, Raymond: Consumer Behavior as Risk Taking. In: Hancock, R. S. (ed.): Dynamic Marketing for a Changing World, Proceedings of the 43rd Conference of the American Marketing Association 1960, p. 389- 398.
- [Chis94] Chisnall, Peter M.: Consumer Behaviour. 3rd rev. ed., McGraw-Hill, London 1994.
- [CoxD67] Cox, Donald: Risk Taking and Information Handling in Consumer Behavior. In: D.Cox (eds.): Risk Taking and Information Handling in Consumer Behavior. Harvard University Press, Boston 1967, p. 604-639.
- [Cunn67] Cunningham, Scott M.: The Major Dimensions of Perceived Risk. In: Cox, D. (eds.): Risk Taking and Information Handling in Consumer Behavior. Harvard University Press, Boston 1967, p. 82-108.
- [DaKa73] Darby, Michael R.; Karni, Edi: Free competition and the optimal amount of fraud. In: Journal of Law and Economics 16 (1973), p. 67-86.
- [GiSK02] Girard, Tulay; Silverblatt, Ronnie; Korgaonkar, Pradeep: Influence of Product Class on Preference for Shopping on the Internet. In: Journal of Computer-Mediated Communication 8 (2002) 1, <http://jcmc.indiana.edu/vol8/issue1/girard.html>, download on 2006-04-10.
- [GüSp04] Günther, Oliver; Spiekermann, Sarah: RFID vs. Privatsphäre ein Widerspruch? In: Wirtschaftsinformatik 46 (2004) 4, p. 245-246.
- [JaKa72] Jacoby, Jacob; Kaplan, Leon B.: The Components of Perceived Risk. In Venkatesan, M. (ed.): Proceedings, Third annual conference of the Association of Consumer Research. Association for Consumer Research, Ann Arbor, MI 1972, <http://www.acrwebsite.org/volumes/display.asp?id=12016>, download on 2006-04-10.
- [Kosc97] Koschnick, Wolfgang J.: Lexikon Marketing M-Z. 2nd rev. ed., Schäffer-Poeschel, Stuttgart 1997.

- [KASW03] Kotler, Philip; Armstrong, Gary; Saunders, John; Wong, Veronica: Grundlagen des Marketing. 3rd ed., Pearson Education Deutschland, München 2003.
- [MuEn86] Murphy, Patrick E.; Enis, Ben M.: Classifying Products Strategically. In: Journal of Marketing 50 (1986), p. 24-42.
- [Nels70] Nelson, Phillip: Information and consumer behavior. In: Journal of Political Economy 78 (1970) 2, p. 311-329.
- [Norm98] Norman, Donald A.: The invisible computer. MIT Press, Cambridge 1998.
- [Pfiz06] Pfizer: Pfizer Introduces Radio Frequency Identification Technology to Combat Counterfeiting, Protect Patient Health.
http://www.pfizer.com/pfizer/are/investors_releases/2006pr/mn_2006_0106.jsp, 2006-01-06, download on 2006-04-10.
- [Roge03] Rogers, Everett M.: Diffusion of innovations. 5. ed., Free Press, New York 2003.
- [Rose67] Roselius, Ted: Consumer Ranking of Risk Reduction Methods. In: Journal of Marketing 35 (1967), p. 56-61.
- [ScKa97] Schiffman, Leon G.; Kanuk, Leslie L.: Consumer Behavior. 6th ed., Prentice-Hall, Englewood Cliffs 1997.
- [WeAd95a] Weiber, Rolf; Adler, Jost: Ein Einsatz von Unsicherheitsreduktionsstrategien im Kaufprozeß: Eine informationsökonomische Analyse. In: Zeitschrift für betriebswirtschaftliche Forschung, Sonderheft 35 (1995), p. 61-77.
- [WeAd95b] Weiber, Rolf; Adler, Jost: Informationsökonomisch begründete Typologisierung von Kaufprozessen. In: Zeitschrift für betriebswirtschaftliche Forschung 47 (1995), p. 43-65.

Zum Einsatz von RFID in der Filiallogistik eines Einzelhändlers

Ergebnisse einer Simulationsstudie

Frédéric Thiesse, Elgar Fleisch

Institut für Technologiemanagement (ITEM-HSG)

Universität St. Gallen

9000 St. Gallen, Schweiz

{frederic.thiesse, elgar.fleisch}@unisg.ch

Abstract

Vor dem Hintergrund des bis heute bestehenden Problems der unzureichenden Regalverfügbarkeit im Handel eröffnet RFID durch eine automatische Erfassung von Warenbewegungen in der Filiale die Möglichkeit, den Prozess der Regalnachbefüllung neu zu gestalten. Dieser Beitrag vergleicht im Rahmen einer Simulationsstudie den herkömmlichen Prozess der Nachbefüllung mit manueller Bestandskontrolle mit einem RFID-gestützten Prozess und analysiert den Einfluss einzelner Kostenfaktoren auf das Gesamtergebnis. Dabei zeigt sich, dass der neue Prozess in Abhängigkeit von Leserate und Hardwarekosten in der Lage ist, die Leistungsfähigkeit des bisherigen Prozesses bezüglich Kosteneffizienz und Warenverfügbarkeit zu übertreffen.

1 Einleitung

1.1 Hintergrund und praktische Relevanz

Trotz umfassender Anstrengungen des Handels und seiner Lieferanten in Form von Initiativen wie „Efficient Consumer Response (ECR)“ konnte eine ganze Reihe von Ineffizienzen in der logistischen Kette bislang nur unzureichend gelöst werden. Zu den aktuellen Problemen der Branche zählt neben Schwund [HoDa02], unverkäuflichen Produkten aufgrund von Beschädigung oder Verderb [Ligh03], Lagerbestandsungenauigkeiten [Rama00] und Fehlern bei der Rechnungsstellung [GMA02] auch das Phänomen der unzureichenden Warenverfügbarkeit in der Filiale, der sog. „Stockouts“ [GrCB02].

Während die Servicegrade in der Retail-Supply-Chain bei den Herstellern bzw. Verteilzentren üblicherweise über 98 % liegen, werden letztlich auf der Verkaufsfläche nur ca. 90-93 % erreicht [ECR03]. Die Ursachen hierfür sind vielfältig und sind sowohl in der Filiale selbst als auch in den vorgelagerten logistischen Prozessen zu suchen [GrCB02]. Wie in Abb. 1 dargestellt ist dabei insb. bei europäischen Einzelhändlern der Prozess der Regalnachbefüllung überdurchschnittlich stark für Stockouts verantwortlich. Ein ähnliches Bild ergab sich auch in einer aktuellen Studie zum Supply Chain Management des Handels in Europa, bei der Regalverfügbarkeiten zwischen 90 und 98,7 % ermittelt wurden [TBKM05]. Bei einzelnen der untersuchten Unternehmen konnte an Tagen mit starker Nachfrage sogar ein Absinken auf nur 80 % beobachtet werden. Dabei wurde die Filiallogistik als wichtigster Ansatzpunkt zur Verbesserung identifiziert, wobei die befragten Retailer gleichzeitig angaben, hier bei der Umsetzung im Vergleich zu anderen Aspekten der logistischen Kette am wenigsten weit fortgeschritten zu sein.

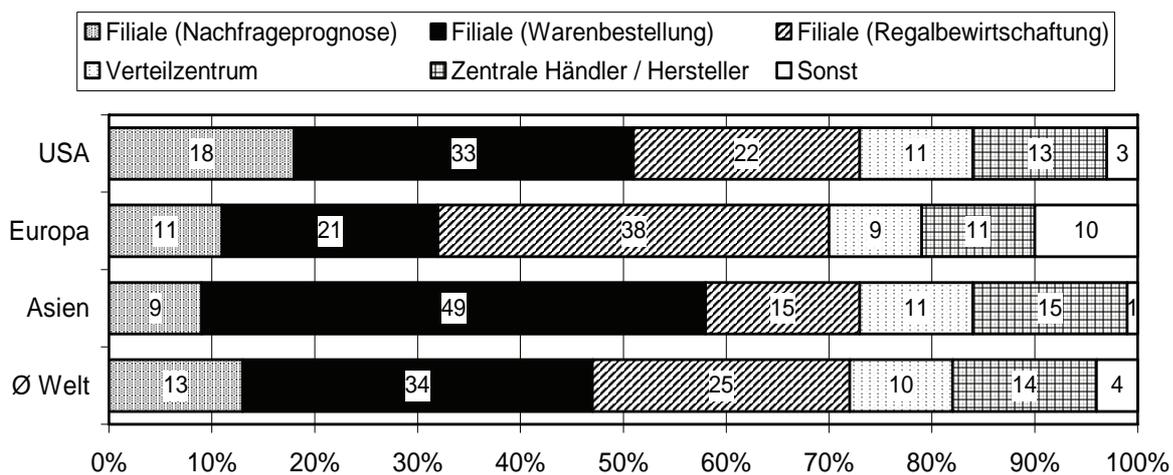


Abb. 1: Ursachen für unzureichende Regalverfügbarkeit [GrCB02]

Von der Reaktion des Konsumenten auf Stockouts sind Handel und Hersteller gleichermaßen betroffen. Je nach Art des Produkts, der Marke oder der konkreten Kaufsituation des Kunden reicht die Palette möglicher Handlungen von einem Aufschub des Kaufs über einen Wechsel von Marke oder Laden bis zum völligen Verzicht auf das Produkt [EmSE91; CoGr04]. Die Folgen sind einerseits verlorene Umsätze – der Kunde entscheidet sich meist für ein kleineres und/oder billigeres Produkt – sowie erhöhte Kosten für unvorgesehene Nachbestellungen, aber andererseits auch langfristige Schäden für das Markenimage und eine Erosion der Kundenbindung [Ange04; CoGr04].

Vor diesem Hintergrund erklärt sich das in den letzten Jahren rapide gewachsene Interesse am Einsatz der Radiofrequenzidentifikation (RFID) zur Erhöhung der Warenverfügbarkeit. Die grundsätzliche Idee besteht hier darin, durch eine verbesserte Sichtbarkeit von Produkten und Lagerbeständen den Nachbefüllungsprozess in der Filiale genauer steuern zu können [CDGG03]. In den zahlreichen Studien und White Papers zum RFID-Einsatz im Handel finden sich unterschiedlichste Abschätzungen über das sich daraus ergebende Potenzial: Bspw. nennt A.T. Kearney eine Reduktion von Stockouts in Höhe von 0,07 % des Umsatzes [ATKe04]. SAP vermutet eine Verbesserung der Warenverfügbarkeit um 5-10 %, die sich in eine Umsatzsteigerung von 3-7 % übersetzen lässt [SAP03]. Accenture erwartet eine Umsatzsteigerung von 1-2 % [Boot03]; eine andere Untersuchung geht hingegen von 2-3 % aus [Pise04]. Die überwiegende Mehrheit derartiger Zahlen hat den Charakter von „educated guesses“, denen keine genauere Analyse der Auswirkungen von RFID auf Prozesse und der relevanten Einflussfaktoren zugrunde liegt [LeÖz05]. Die Verlässlichkeit der genannten Schätzungen ist somit eher gering und lässt keine Aussage zur Übertragbarkeit auf den realen Einzelfall zu.

1.2 Verwandte Arbeiten

In der aktuellen wissenschaftlichen Literatur finden sich verschiedene Arbeiten, die sich mit den Potenzialen und Auswirkungen der RFID-Technologie in der Filiallogistik auseinandersetzen. Dabei handelt es sich einerseits um qualitative Beiträge, die auf der Grundlage von Fallstudien, Experteninterviews oder Umfragen den RFID-Einsatz in der Filiale strukturieren und bewerten. Andererseits liegen auch quantitative Beiträge vor, die erste Ergebnisse aus Pilotprojekten vorstellen oder RFID-gestützte Prozesse anhand mathematischer Modelle untersuchen.

Die Untersuchung von Wong und McFarlane [WoMc03] ist ein Beispiel für die erstgenannte Kategorie. Die Autoren beschreiben zunächst den Ablauf des traditionellen Regalbefüllungsprozesses, wobei zwischen einer „Pull Policy“ mit Prüfung der Regalbestände und einer „Push Policy“ mit Prüfung der Bestände im Lagerraum der Filiale unterschieden wird. Anschließend werden die wesentlichen Einflussfaktoren analysiert, die in der Realität zu einer suboptimalen Prozessperformance führen, u. a. Verzögerungen bei der Bestandskontrolle oder veraltete Pick-Listen. Vor diesem Hintergrund identifizieren die Autoren Ansatzpunkte zur Verbesserung mittels RFID wie bspw. die automatische Identifikation von Regalbeständen und Warenbewegungen sowie die automatische Generierung von Pick-Listen auf mobilen Endgeräten.

In der Simulationsstudie von Lee, Cheng und Leung [LeCL04] wird anhand eines einfachen Supply-Chain-Modells analysiert, welche Effekte sich durch (a) die Vermeidung von Bestandsungenauigkeiten, (b) die Neugestaltung des Regalbefüllungsprozesses und (c) den Austausch von Bestandsinformationen zwischen Hersteller und Händler ergeben. Im zweiten Fall wird der bisherige Prozess des periodischen Bestandsmanagements mit einer kontinuierlichen Kontrolle, z.B. durch RFID-Leser in den Verkaufsregalen, verglichen. Die Ergebnisse der Studie zeigen auf, dass durch RFID der Befüllungsprozess häufiger und besser an die Nachfrage angepasst durchgeführt wird, wobei gleichzeitig geringere Bestände in den Regalen erforderlich sind. Die Aussagekraft der Studie wird dadurch relativiert, dass für beide Prozesse die wesentlichen Parameter „Regalfläche“ und „Sicherheitsbestand“ nicht optimal, sondern willkürlich gewählt werden, so dass keine eindeutige Bewertung zur Vorteilhaftigkeit abgegeben werden kann. Darüber hinaus verwenden die Autoren ausschließlich Bestände und Stockouts als Kenngrößen ihrer Analyse und verzichten auf eine Gesamtbetrachtung von Kosten und Nutzen.

Hardgrave, Waller und Miller [HaWM05] berichten über die Ergebnisse eines Pilotprojekts bei Wal-Mart, bei dem von Februar bis September 2005 in 12 Filialen unterschiedlicher Formate 4554 verschiedene Produkte auf der Kartonebene „getaggt“ wurden, um eine Erfassung der Warenverräumung beim Übergang vom Lagerraum zur Verkaufsfläche zu ermöglichen. Der RFID-Einsatz führte hier zu einer durchschnittlichen Reduktion der Stockouts um 16 % im Vergleich zu einer Kontrollgruppe aus 12 weiteren Filialen, wobei im besten Fall bis zu 62 % Reduktion bei Produkten mit einem täglichen Absatz von 6 bis 15 Einheiten erzielt werden konnten [Coll06]. Als wesentlichen Treiber für diese Verbesserung identifizieren die Autoren die automatisch generierte Pick-Liste, die die manuelle Bestandskontrolle ersetzt.

1.3 Zielsetzung und Vorgehen

Ziel dieses Beitrags ist es, die Leistungsfähigkeit des traditionellen Prozesses der Regalbefüllung in der Filiale mit einem RFID-gestützten Prozess zu vergleichen und Schlussfolgerungen für den Einsatz in der Praxis abzuleiten. Da europäische Händler meist auf Einzelproduktebene kommissionieren, gehen wir von einer RFID-Kennzeichnung einzelner Verkaufseinheiten aus. Im Gegensatz zu den zuvor genannten Arbeiten vergleichen wir dabei nicht nur Hilfsindikatoren, sondern die effektiven Kosten und Nutzen. Aufgrund der Komplexität des Prozesses greifen wir dazu im nächsten Abschnitt auf das Werkzeug der Simulation zurück (vgl. [AlMa92]). Der Beitrag endet mit einer Zusammenfassung und Vorschlägen für die weitere Forschung.

2 Simulationsstudie

Im Folgenden entwickeln wir in einem ersten Schritt die Struktur des zugrunde liegenden Simulationsmodells und der beiden Nachbefüllungsprozesse. Anschließend präsentieren und diskutieren wir Ergebnisse verschiedener Simulationsläufe inklusive einer Sensitivitätsanalyse. Die dabei verwendeten Modellparameter sind in Tabelle 1 zusammengefasst. Die Implementierung des Simulationsmodells und die anschließende Outputanalyse erfolgten in Microsoft Excel bzw. Visual Basic for Applications (VBA).

R	Zeitlicher Abstand zwischen zwei Kontrollen
s	Sicherheitsbestand im Regal („Base Stock“)
S	Reservierter Regalplatz für das Produkt („Shelf Space“)
p	Marge pro Verkaufseinheit exkl. Kosten für Filiallogistik
c^{Inv}	Lagerkosten pro Verkaufseinheit und Tag
c^{Rev}	Personalkosten pro Bestandskontrolle („Review“)
c^{Rep}	Personalkosten pro Nachbefüllungsvorgang („Replenishment“)
c^{Tag}	Kosten eines Transponder-Labels („Tag“)
N	Anzahl Simulationsläufe
T	Länge des Untersuchungszeitraums
V^{Sales}	Anzahl verkaufter Produkte im Untersuchungszeitraum
V^{OOS}	Anzahl Stockouts im Untersuchungszeitraum
V^{Rev}	Anzahl Bestandskontrollen im Untersuchungszeitraum
V^{Rep}	Anzahl Regalnachbefüllungen im Untersuchungszeitraum
OOS	Stockout-Quote im Untersuchungszeitraum
d	Zeitliche Verzögerung zwischen Bestandskontrolle und Nachbefüllung
I_t^P	Physischer Regalbestand zum Zeitpunkt t
I_t^{IS}	Regalbestandsinformation im Warenwirtschaftssystem zum Zeitpunkt t
φ	Leserate der RFID-Hardware
r_i	Erkennung des Produkts i durch die RFID-Hardware beim Warenübergang (ja/nein)
t_k	Zeitpunkt der Ankunft des Kunden k am Regal
λ	Kundenankunftsrate
F^{PR}	Zielfunktion für den Prozess mit periodischer Bestandskontrolle
F^{RFID}	Zielfunktion für den RFID-gestützten Prozess

Tab. 1: Referenzliste der verwendeten Modellparameter

2.1 Modellformulierung

Das in diesem Abschnitt beschriebene Simulationsmodell dient als Grundlage für die Analyse der RFID-gestützten Regalbewirtschaftung. Es stellt insofern eine Vereinfachung der realen Gegebenheiten dar, als wir eine fehlerfreie Identifikation am Wareneingang und am POS voraussetzen, so dass der Bestand eines Produkts für die Gesamtfiliale stets bekannt ist. Dies schließt insb. auch den Schwund durch Diebstahl und andere Faktoren, die zu Fehlern bei den Lagerbestandsinformationen im Warenwirtschaftssystem führen, aus. Für eine genauere Betrachtung des RFID-Einsatzes im Zusammenhang mit Bestandsungenauigkeiten in der Lieferkette sei der Leser auf [AtLÖ05; FITe05; KöSh06] verwiesen.

2.1.1 Prozess auf Grundlage periodischer Bestandskontrolle

Der traditionelle Prozess der Regalnachbefüllung entspricht einem Bestandsmanagement mittels periodischer Kontrolle der Verkaufseinheiten im Regal („Periodic Review“). Dabei prüft ein Mitarbeiter in regelmäßigen zeitlichen Abständen R , ob noch genügend Produkte im Regal vorhanden sind. Unterschreitet der Regalbestand einen bestimmten Schwellwert s , wird eine Nachbefüllung ausgelöst, die den Bestand mit einer zeitlichen Verzögerung d wieder auf den Maximalwert S – d.h. bis zur für das Produkt reservierten Regalfläche – auffüllt. Dieses Vorgehen wird in der Literatur als „(R, s, S)-Regel“ bezeichnet [SiPP98, 240].

Durch die Bestandskontrolle und die Nachbefüllung entstehen jeweils Personalkosten in Höhe c^{Rev} bzw. c^{Rep} . Darüber hinaus sind Lagerkosten c^{Inv} für die Reservierung von Regalfläche pro Verkaufseinheit zu berücksichtigen. Andererseits verdient der Händler vor Abzug der erwähnten Kosten für die Filiallogistik einen Betrag p pro verkaufter Einheit. Ziel ist es nun, durch die Wahl geeigneter Werte für R , s und S den Nachbefüllungsprozess so zu steuern, dass der sich ergebende Profit F^{PR} für den Händler maximal wird. Während R unmittelbar die Zahl der Bestandskontrollen V^{Rev} in einem Untersuchungszeitraum T bestimmt, wird die Zahl der notwendigen Nachbefüllungen V^{Rep} bzw. der verkauften Einheiten V^{Sales} auch von s und S beeinflusst. Das Zielkriterium setzt sich somit wie folgt aus den Verkaufserlösen abzüglich der Kosten für Bestandskontrollen, Nachbefüllungen und Regalfläche im Zeitraum T zusammen:

$$\max_{R,s,S} F^{PR} = \max_{R,s,S} \left(pV^{Sales} - c^{Rev}V^{Rev} - c^{Rep}V^{Rep} - c^{Inv}S \cdot T \right) \quad (1)$$

2.1.2 RFID-gestützter Prozess

Die Idee hinter dem Konzept des RFID-gestützten Prozesses ist es, durch eine automatische Erfassung von Warenbewegungen zwischen Lagerraum und Verkaufsfläche die manuelle Bestandskontrolle überflüssig zu machen. Mit Hilfe der RFID-generierten Information über Produkte, die zum Regal gebracht werden, und den Abverkaufsdaten aus dem POS-System kann zu jedem Zeitpunkt t im Warenwirtschaftssystem eine Schätzung I_t^{IS} für den tatsächlichen Regalbestand I_t^P errechnet werden, auf deren Grundlage eine Nachbefüllung ausgelöst wird.

Die Qualität dieser Schätzung ist abhängig von der Lesegenauigkeit der RFID-Infrastruktur am Ausgang des Lagerraums, die real nur selten 100 % erreicht. So können sich in der Praxis je nach Transponder- und Produkttyp bei der Pulkerfassung auch Leseraten von weniger als 30 % ergeben [GS1F05]. Aus diesem Grund beinhaltet unser Modell explizit die RFID-Leserate in Form des Parameters $\varphi \in [0 \dots 100]$. Die Zufallsvariable $r_i \in [0,1]$ bezeichnet dementsprechend die Erfassung ($r_i = 1$) oder Nicht-Erfassung ($r_i = 0$) einer Verkaufseinheit i beim Übergang vom Lagerraum auf die Verkaufsfläche, wobei die Wahrscheinlichkeit der Erfassung durch

$$p(r_i = 1) = \frac{\varphi}{100} \quad (2)$$

gegeben ist. Bezeichnen t' und t die Zeitpunkte unmittelbar vor bzw. nach dem Wechsel eines Produkts i auf die Verkaufsfläche, so wird die Regalbestandsinformation im System folgendermaßen aktualisiert:

$$I_t^{IS} := I_{t'}^{IS} + r_i \quad (3)$$

Umgekehrt wird bei der Erfassung an der Kasse ein Produkt i auch nur dann vom Regalbestand I_t^{IS} abgezogen, falls zuvor $r_i = 1$ galt, damit im System nicht kontinuierlich mehr Einheiten aus- als eingebucht werden. Zu diesem Zweck ist es notwendig, dass im Warenwirtschaftssystem nicht nur zwei getrennte Lagerbestände, sondern auch die ID-Codes der erfassten Artikel im Regal bis zum Check-Out vorgehalten werden. Somit gilt zu jedem Zeitpunkt t :

$$0 \leq I_t^{IS} \leq I_t^P \leq S \quad (4)$$

Beim RFID-gestützten Prozess entfallen manuelle Bestandskontrollen und damit auch die entsprechenden Personalkosten vollständig. Andererseits fallen für die Ausstattung jeder Verkaufseinheit Transponderkosten c^{Tag} an, die die Marge des Händlers schmälern. Der Entscheidungsraum beschränkt sich daher auf die Parameter s und S , durch deren Wahl der Profit F^{RFID} maximiert werden soll:

$$\max_{s,S} F^{RFID} = \max_{s,S} \left((p - c^{Tag}) V^{Sales} - c^{Rep} V^{Rep} - c^{Inv} S \cdot T \right) \quad (5)$$

2.2 Simulationsergebnisse

Ausgangspunkt unserer Simulationsexperimente ist der Fall eines typischen „Mitteldrehers“ mit einer Nachfrage von 10 Einheiten pro Tag bei einem Verkaufspreis von 10 €, z.B. einer Packung Babywindeln. Wir betrachten pro Simulationslauf eine Filiale, die $T = 365$ Tage à 10 Stunden geöffnet ist. Die Ankunft der Kunden modellieren wir entsprechend als Poisson-Prozess mit Ankunftsrate $\lambda = 10$ und berechnen für jeden Simulationslauf eine Folge von Kundenankunftszeiten t_k für alle im Untersuchungszeitraum eintreffenden Kunden ($0 \leq t_k \leq T$).

Für diese Filiale nehmen wir Kosten pro Bestandskontrolle von $c^{Rev} = 0,5$ €, Kosten pro Nachbefüllung von $c^{Rep} = 2$ € und Regalkosten von $c^{Inv} = 0,1$ € pro Einheit und Tag an. Wir gehen weiter davon aus, dass vor Abzug der Kosten für die Filiallogistik dem Händler ein Gewinn von $p = 1$ € verbleibt. Wie wir weiter unten sehen werden, führt dies letztlich zu einer realistischen Marge von 6,3 %. Außerdem unterstellen wir, dass wie in der Realität die Nachbefüllung nicht in Nullzeit, sondern mit einer Verzögerung von 30 Minuten ($d = 0,05$ Tage) geschieht.

2.2.1 Prozess auf Grundlage periodischer Bestandskontrolle

Um die kostenoptimalen Werte für R , s und S zu ermitteln, wurden alle Kombinationen für $0 < R \leq 3$ und $0 < s \leq S \leq 50$ mit jeweils $N = 50$ Durchläufen simuliert. Wie in Tabelle 2 dargestellt ergibt sich für $R = 1,6$, $s = 15$ und $S = 20$ der maximale Durchschnittswert der Zielfunktion von 2254,82 €. Bei durchschnittlich 3555,86 verkauften Einheiten pro Jahr bleibt dem Händler somit ein Gewinn von 0,63 € pro Stück. Zum Vergleich geben wir in der Tabelle auch die Ergebnisse für leicht variierte Werte von R , s und S an sowie für ganzzahlige R , da in

die Ergebnisse für leicht variierte Werte von R , s und S an sowie für ganzzahlige R , da in der Praxis der rechnerisch korrekte Kontrollzyklus von 1,6 Tagen kaum implementierbar wäre.

R	s	S	\overline{F}^{PR} [€]	\overline{V}^{Sales}	\overline{V}^{Rev}	\overline{V}^{Rep}	\overline{V}^{OOS}	OOS [%]	$\overline{F}^{PR}/\overline{V}^{Sales}$ [€]
1,6	15	20	2254,82	3555,86	228,00	227,52	85,18	2,33	0,63
1,6	16	20	2254,62	3556,26	228,00	227,82	84,78	2,33	0,63
1,6	14	20	2254,10	3553,70	228,00	226,80	87,34	2,40	0,63
1,7	15	21	2251,56	3553,62	214,00	213,28	87,42	2,40	0,63
1,6	15	19	2249,14	3514,28	228,00	227,82	126,76	3,48	0,64
1,5	15	20	2248,72	3586,22	243,00	242,00	54,82	1,51	0,63
1,6	15	21	2247,52	3583,62	228,00	226,80	57,42	1,58	0,63
2,0	20	24	2216,74	3549,74	182,00	182,00	91,30	2,51	0,62
1,0	11	21	2180,62	3567,80	364,00	218,34	73,24	2,01	0,61
1,2	19	24	2018,68	3641,04	304,00	296,00	0,00	0,00	0,55

Tab. 2: Resultate der (R,s,S)-Regel für verschiedene Parameterkombinationen

Wie sich an dieser Stelle zeigt, führt der kostenoptimale Fall keineswegs zu einer 100-%igen Regalverfügbarkeit, sondern zu einer Stockout-Quote OOS von durchschnittlich 2,33 %. Diese ist definiert als

$$OOS = 100 \cdot \frac{V^{OOS}}{V^{Sales} + V^{OOS}} \quad (6)$$

Eine Vermeidung der Stockouts wäre zwar durch mehr Regalfläche oder häufigere Kontrolle möglich, der daraus resultierende zusätzliche Umsatz würde aber durch die ebenfalls höheren Lager- oder Personalkosten mehr als wettgemacht: Der bestmögliche Wert für \overline{F}^{PR} unter der Bedingung $OOS < 0,01$ % beträgt 2018,68 €, d.h. 10,5 % weniger als im optimalen Fall.

2.2.2 RFID-gestützter Prozess

Die Simulationen des RFID-gestützten Prozesses wurden für verschiedene Kosten und Fähigkeiten der Technologie durchgeführt: Einerseits wurde die Leserate φ von 0 bis 100 % in 10er-Schritten variiert, andererseits wurden unterschiedliche Preise für Transponder-Labels c^{Tag} an-

genommen. Den heute am Markt üblichen Kosten von 0,2 € pro Tag wurden die für die nächsten Jahre zu erwartenden Preise von 0,1 € und 0,05 € sowie der mit heutiger Chip- und Fertigungstechnik noch nicht erzielbare Preis von 0,01 € gegenübergestellt. Die übrigen Parameter s und S wurden wie zuvor variiert.

Die für jede Wertekombination jeweils maximalen Werte für \overline{F}^{RFID} sind im Vergleich zum Optimum für \overline{F}^{PR} in Abbildung 2 dargestellt. Wie sich hier zeigt, schneidet der RFID-Prozess bei $c^{Tag} = 0,2$ € und $c^{Tag} = 0,1$ stets schlechter ab als der manuelle Prozess. Eine Verbesserung ist demgegenüber für niedrigere Tagkosten möglich. Bspw. ergibt sich für eine 90-%ige Leserate und $c^{Tag} = 0,05$ € ein Optimum in Höhe von 2319,07 € (d.h. +2,85 % gegenüber Periodic Review) bei $s = 4$ und $S = 17$.

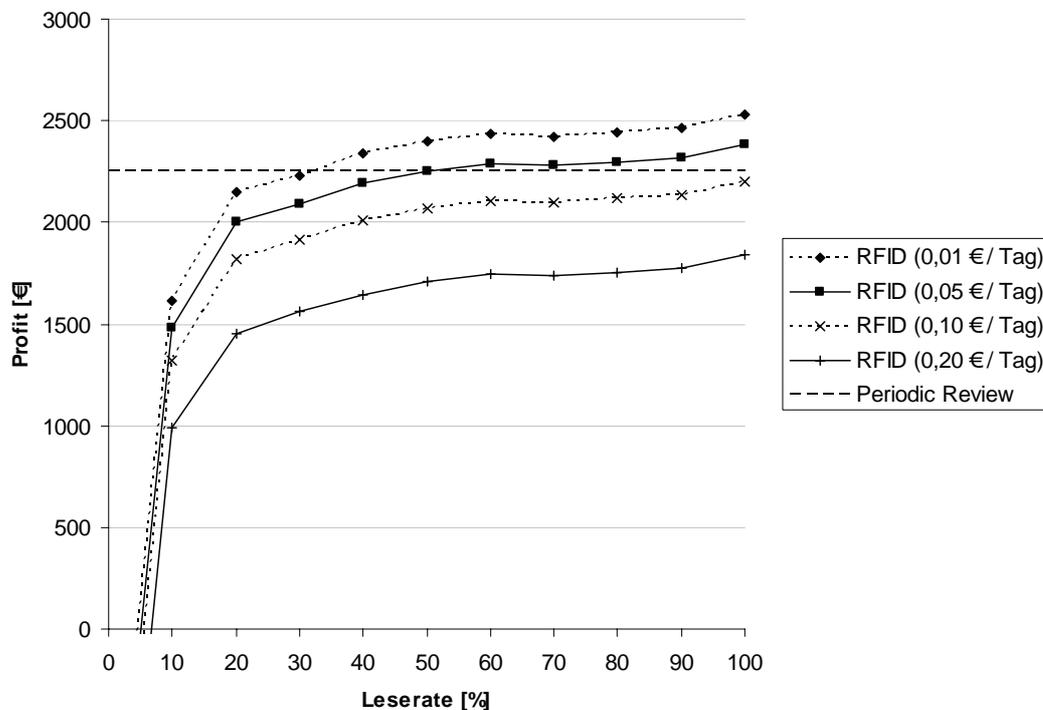


Abb. 2: Resultate des RFID-gestützten Prozesses in Abhängigkeit der Leserate

Bezüglich der Auswirkungen schlechter werdender Leseraten ergeben die Simulationen ein eher kontraintuitives Resultat: Während scheinbar eine deutliche Abhängigkeit der Prozessleistung von der Anzahl der bei der Warenverräumung erfassten Einheiten zu erwarten gewesen wäre, verhält sich F^{RFID} im Modell vergleichsweise insensitiv gegenüber verändertem φ . Erst für Leseraten unter 50 % lässt der Profit des Händlers deutlich nach, um schließlich bei weniger als 20 % ganz einzubrechen.

Wie in Tabelle 3 dargestellt, führen schlechte Leseraten zu einem Optimum mit erhöhtem Regalplatzbedarf als Ausgleich. So steigt bei $\varphi = 50\%$ der Regalplatz S gegenüber $\varphi = 100\%$ von 15 auf 19 Einheiten während die Umsätze pV^{Sales} und Nachbefüllungskosten $c^{Rep}V^{Rep}$ annähernd konstant bleiben. Die Minderung des Profits bei sinkendem φ wird also praktisch ausschließlich durch die höheren Lagerkosten $c^{Inv}S \cdot T$ verursacht.

φ [%]	s	S	$\overline{F^{RFID}}$ [€]	$\overline{pV^{Sales}}$ [€]	$\overline{c^{Rep}V^{Rep}}$ [€]	$\overline{c^{Inv}S \cdot T}$ [€]	\overline{OOS} [%]	$\overline{F^{PR}/V^{Sales}}$ [€]
0	1	3	-1550,17	3422,53	4863,20	109,50	1.55	-0.64
10	1	33	1484,20	3111,02	422,32	1204,50	10.06	0.45
20	2	27	2000,47	3458,17	472,20	985,50	0.02	0.55
30	2	22	2089,41	3332,01	439,60	803,00	3.67	0.60
40	3	21	2192,56	3458,74	499,68	766,50	0.01	0.60
50	3	19	2250,06	3445,88	502,32	693,50	0.38	0.62
60	3	18	2287,57	3446,05	501,48	657,00	0.37	0.63
70	3	19	2278,60	3420,86	448,76	693,50	1.10	0.63
80	4	18	2299,06	3458,82	502,76	657,00	0.00	0.63
90	4	17	2319,07	3458,65	519,08	620,50	0.01	0.64
100	2	15	2382,53	3433,07	503,04	547,50	0.75	0.66

Tab. 3: Umsatz und Kosten in Abhängigkeit der Leserate bei einem Tagpreis von 0,05 €

Der Grund für dieses überraschende Verhalten ist in der Eigenschaft von RFID als Technik zur eindeutigen Identifikation zu suchen. Während über den klassischen Barcode nur eine Erkennung des Produkttyps – und damit nur eine Mengenerfassung – möglich ist, erlaubt RFID die individuelle Erkennung jeder Verkaufseinheit (vgl. Abschnitt 2.1.2). Die beim Übergang zur Verkaufsfläche und später an der Kasse erfassten Produkte bilden damit eine Stichprobe, über deren Entwicklung eine auch bei schlechten Leseraten hinreichend gute Aussage über den Verlauf des Gesamtbestands getroffen werden kann. Abbildung 3 zeigt beispielhaft für $\varphi = 50\%$, wie sich der physische Regalbestand I_t^P und der im Warenwirtschaftssystem abgebildete Bestand der mittels RFID erfassten Einheiten I_t^{IS} im Zeitverlauf entwickeln.

Nicht zuletzt lässt sich an den Ergebnissen auch ablesen, dass die Stockout-Quoten für den RFID-Prozess auch für suboptimale Leseraten fast durchweg deutlich besser ausfallen als für den Fall der manuellen Bestandskontrolle. Die Technologie hat somit das Potenzial, die Regalverfügbarkeit bei gleichzeitig niedrigeren Kosten zu erhöhen.

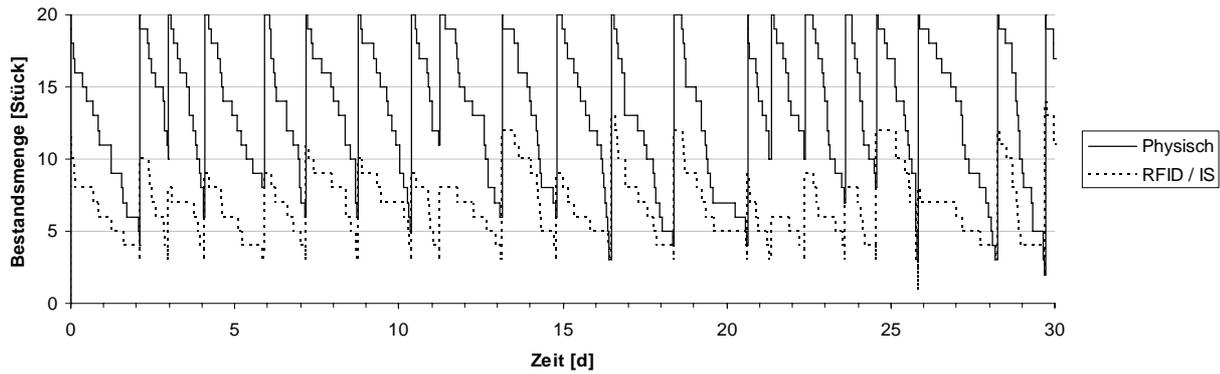


Abb. 3: Bestandsmengen im Regal und im Warenwirtschaftssystem bei 50-%iger Leserate (Beispiel)

Wie wir zuvor gesehen haben, profitiert der RFID-gestützte Prozess von der geringeren Regalfläche, so dass weniger Lagerkosten anfallen. In der Praxis ist eine solche dynamische Anpassung von Regalfläche jedoch meist nur begrenzt möglich. Aus diesem Grund stellt sich die Frage, inwieweit sich RFID auch sinnvoll einsetzen lässt, wenn die Regalfläche S als konstant angenommen wird. Abbildung 4 zeigt für das Beispiel $c^{Tag} = 0,05$ €, dass sich die Durchschnittswerte für F^{RFID} bei guten Leseraten nicht drastisch unterscheiden. So ist $\overline{F^{RFID}} = 2299,80$ € bei $\varphi = 90$ % und $S = 20$ und damit 0,83 %-Punkte schlechter als im optimalen Fall.

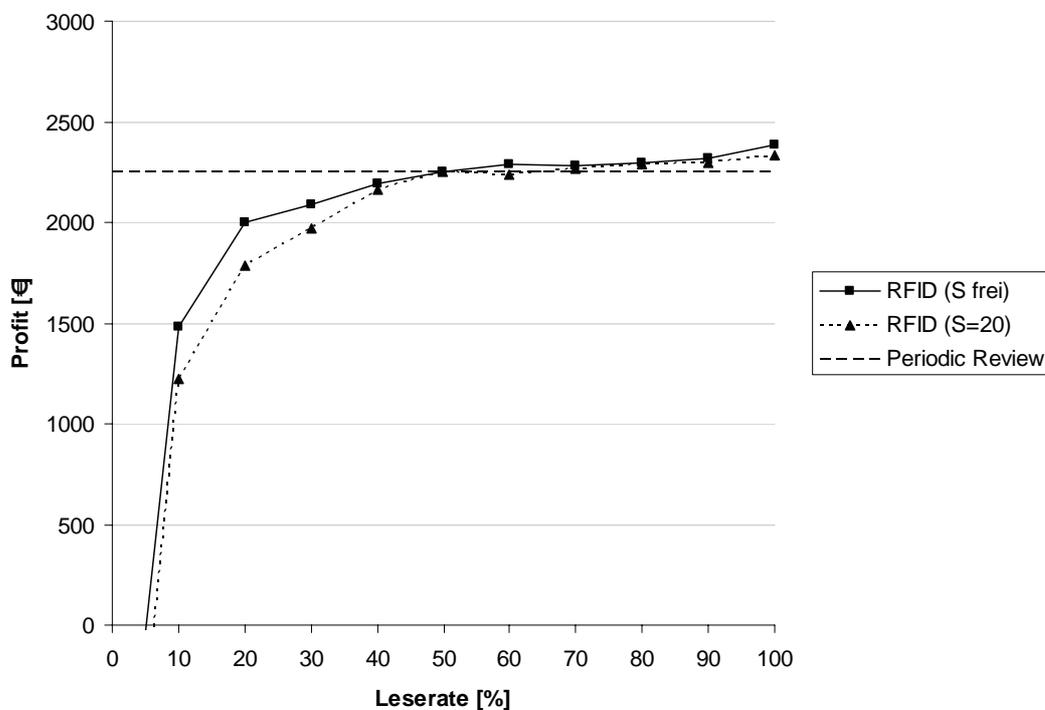


Abb. 4: Vergleich des RFID-gestützten Prozesses bei fixer und bei optimierter Regalfläche

2.3 Sensitivitätsanalyse

Die Leistung der beiden zuvor betrachteten Prozesse wird von verschiedenen Kostengrößen bestimmt, die wir bisher als konstant vorausgesetzt haben. In der Praxis unterscheiden sich Personal- und Lagerkosten jedoch deutlich je nach Ladenformat (Hypermarkt vs. Tante-Emma-Laden), Art des Retailers (Discounter vs. Feinkostgeschäft), Lage (Grüne Wiese vs. Innenstadt), Region (Großstadt vs. ländlicher Raum) usw. Aus diesem Grund analysieren wir in diesem Abschnitt die Auswirkungen veränderter Kosten c^{Rev} , c^{Rep} und c^{Inv} auf das sich ergebende Optimum der Zielfunktionen und die jeweiligen Stockout-Quoten. Dabei variieren wir jede Kostengröße separat um den Faktor $\frac{1}{4}$ und $\frac{1}{2}$ bzw. 2 und 4. Dem Prozess der periodischen Bestandskontrolle stellen wir einen RFID-Prozess mit $c^{Tag} = 0,05$ € und $\varphi = 90$ % gegenüber.

2.3.1 Kosten für die Bestandskontrolle

Tabelle 4 fasst die Ergebnisse der Sensitivitätsanalyse bei unterschiedlichen Kosten für die Bestandskontrolle c^{Rev} zusammen. Da der RFID-Prozess keine Reviews benötigt, bleiben hier die Werte jeweils unbeeinflusst. Wie zu erwarten, leidet die Performance des herkömmlichen Prozesses unter den hohen Personalkosten, die zu weniger Kontrollen und Nachbefüllungen führen und durch größere Regalflächen ausgeglichen werden müssen. Die Zahl der Stockouts bewegt sich hingegen in allen Fällen auf einem ähnlichen Niveau.

c^{Rev} [€]	Regel	R	s	S	\bar{F} [€]	\bar{V}^{Sales}	\bar{V}^{Rev}	\bar{V}^{Rep}	\bar{V}^{OOS}	\bar{OOS} [%]
0,125	P.Rev.	0,5	6	17	2369,98	3567,60	729	242,00	73,44	2,02
	RFID		4	17	2319,07 (-2,15 %)	3640,68		259,54	0,36	0,01
0,25	P.Rev.	1,4	13	18	2318,26	3558,66	260	258,20	82,38	2,26
	RFID		4	17	2319,07 (+0,03 %)	3640,68		259,54	0,36	0,01
0,5	P.Rev.	1,6	15	20	2254,82	3555,86	228	227,52	85,18	2,34
	RFID		4	17	2319,07 (+2,85 %)	3640,68		259,54	0,36	0,01
1,0	P.Rev.	1,7	15	21	2144,56	3553,62	214	213,28	87,42	2,40
	RFID		4	17	2319,07 (+8,14 %)	3640,68		259,54	0,36	0,01
2,0	P.Rev.	2,1	19	25	1944,98	3551,36	173	172,94	89,68	2,46
	RFID		4	17	2319,07 (+19,23 %)	3640,68		259,54	0,36	0,01

Tab. 4: Auswirkungen unterschiedlicher Kostensätze für die Bestandskontrolle

2.3.2 Kosten für die Nachbefüllung

Die in Tabelle 5 aufgeführten Resultate demonstrieren erwartungsgemäß, dass beide Prozesse unter steigenden Nachbefüllungskosten c^{Rep} leiden. Dabei zeigt sich, dass interessanterweise der relative Abstand zwischen \overline{F}^{PR} und \overline{F}^{RFID} sowohl für kleinere als auch für größere Werte von c^{Rep} zunimmt. Dieses Ergebnis kann dadurch erklärt werden, dass einerseits der RFID-Prozess von niedrigen Befüllungskosten profitiert, da dort die Anzahl der Nachbefüllungen tendenziell größer ist: So werden im Fall $c^{Rep} = 0,5$ € durchschnittlich 521,58 Befüllungen bei RFID im Vergleich zu 400,90 bei periodischer Kontrolle durchgeführt. Bei hohen Befüllungskosten wiederum profitiert der RFID-Prozess von seiner gezielteren Reaktion auf die Nachfrage, so dass bei weniger Nachbefüllungen dennoch eine bessere Regalbewirtschaftung erreicht wird. Dies schlägt sich in einer höheren Zahl verkaufter Einheiten (3639,16 im Gegensatz zu 3509,96) bzw. einer niedrigeren Stockout-Quote (0,05 % im Gegensatz zu 3,60 %) nieder.

c^{Rep} [€]	Regel	R	s	S	\overline{F} [€]	\overline{V}^{Sales}	\overline{V}^{Rev}	\overline{V}^{Rep}	\overline{V}^{OOS}	\overline{OOS} [%]
0,5	P.Rev.	0,9	11	13	2696,65	3574,60	405	400,90	66,44	1,82
	RFID		4	10	2832,76 (+5,05 %)	3640,58		521,58	0,46	0,01
1,0	P.Rev.	1,2	13	16	2525,64	3565,62	304	302,98	75,42	2,07
	RFID		4	13	2620,61 (+3,76 %)	3640,62		363,48	0,42	0,01
2,0	P.Rev.	1,6	15	20	2254,82	3555,86	228	227,52	85,18	2,34
	RFID		4	17	2319,07 (+2,85 %)	3640,68		259,54	0,36	0,01
4,0	P.Rev.	2,1	18	25	1856,70	3550,98	173	172,82	90,06	2,47
	RFID		3	23	1903,23 (+2,51 %)	3608,52		171,34	32,52	0,89
8,0	P.Rev.	2,5	18	28	1249,70	3509,96	145	144,72	131,08	3,60
	RFID		3	30	1324,92 (+6,02 %)	3639,16		129,66	1,88	0,05

Tab. 5: Auswirkungen unterschiedlicher Kostensätze für die Nachbefüllung

2.3.3 Kosten für die Regalfläche

In Tabelle 6 sind die Ergebnisse der Sensitivitätsanalyse für verschiedene Kosten der Regalfläche pro Einheit c^{Inv} zusammengefasst. Bei niedrigen Lagerkosten erweist sich der herkömmliche Prozess als effizienter, da höhere Regalbestände als Puffer aufgebaut werden können. Die bei hohen Lagerkosten zwangsläufig kleineren Regalbestände hingegen führen zu geringeren Umsätzen bzw. zu einer höheren Stockout-Quote im Fall der periodischen Bestandskontrolle.

c^{Inv} [€]	Regel	R	s	S	\bar{F} [€]	\bar{V}^{Sales}	\bar{V}^{Rev}	\bar{V}^{Rep}	\bar{V}^{OOS}	\bar{OOS} [%]
0,025	P.Rev.	2,5	26	33	2952,60	3618,22	145	145,00	22,82	0,63
	RFID		3	30	2924,13 (-0,96 %)	3639,16		129,66	1,88	0,05
0,05	P.Rev.	2,1	21	27	2672,85	3599,98	173	172,94	41,06	1,13
	RFID		4	22	2674,08 (+0,05 %)	3640,78		191,58	0,26	0,01
0,1	P.Rev.	1,6	15	20	2254,82	3555,86	228	227,52	85,18	2,34
	RFID		4	17	2319,07 (+2,85 %)	3640,68		259,54	0,36	0,01
0,2	P.Rev.	1,1	10	14	1669,58	3507,28	331	324,1	133,76	3,67
	RFID		4	13	1782,63 (+6,77 %)	3640,62		363,48	0,42	0,01
0,4	P.Rev.	0,8	7	9	868,10	3306,78	456	447,34	334,26	9,18
	RFID		4	10	955,39 (+10,06 %)	3640,58		521,58	0,46	0,01

Tab. 6: Auswirkungen unterschiedlicher Kostensätze für die Regalfläche

2.4 Diskussion

Wie auch bei anderen Simulationsstudien sind die in den vorangegangenen Abschnitten vorgestellten Ergebnisse bezüglich Ihrer Übertragbarkeit auf die Praxis mit Vorsicht zu interpretieren. Einerseits beinhaltet das zu Anfang entwickelte Modell eine Reihe von Vereinfachungen gegenüber dem realen Prozess. Ein Beispiel hierfür ist der bereits erwähnte Schwund in der Filiale durch Verderb, Diebstahl oder „Product Misplacement“, d.h. Fehler bei der Warenverräumung. Ein anderes Beispiel ist die Abhängigkeit der Nachfrage nach einem Produkt vom aktuellen Regalbestand bzw. der Präsentation der Ware im Regal [LaDe99]. Andererseits werden im Modell verschiedene Randbedingungen und Einflussfaktoren ausgeblendet, die bei einer Implementierung zu berücksichtigen wären. Hierzu zählen vertragliche Vereinbarungen über Regalflächen mit den Lieferanten, vorgegebene Layouts der Filialräume sowie Aspekte der Sortimentsgestaltung und der Konsumentenpsychologie.

Nichtsdestotrotz lassen die Simulationsergebnisse einige Schlussfolgerungen zum RFID-Einsatz zu, die auch praktisch relevant sind. Es konnte gezeigt werden, dass RFID eine Neugestaltung von Filialprozessen ermöglicht, die in Abhängigkeit der Kostengrößen und der RFID-Leserate als wesentlichem Indikator für die Leistungsfähigkeit der Technologie effizienter arbeiten kann als der bisherige Prozess. Dabei erwies sich auch, dass die Leserate keine so große Rolle für den Prozess spielt, wie intuitiv zu erwarten gewesen wäre. Darüber hinaus deuten die Ergebnisse darauf hin, dass der maximale Nutzen aus Technologieeinsatz erst gezogen werden kann, wenn parallel auch eine Optimierung der Regalflächen vorgenommen wird.

3 Zusammenfassung

Ziel dieses Beitrags war es, am Beispiel eines typischen Supermarktprodukts mittels eines Simulationsmodells eine Aussage zur Vorteilhaftigkeit eines RFID-gestützten Prozesses zur Regalnachbefüllung in der Filiale im Vergleich zum heutigen Prozess periodischer Bestandskontrolle abzuleiten. Dabei konnte gezeigt werden, dass auch beim kostenintensiven Tagging von Einzelprodukten RFID das Potenzial zur Effizienzsteigerung in der Filiallogistik bei gleichzeitiger Verbesserung der Regalverfügbarkeit bietet. Für den Praktiker bietet das vorgestellte Modell einen ersten Ansatz zur Bewertung der Sinnhaftigkeit des RFID-Einsatzes in der Filiallogistik, in die auch weitere Faktoren wie Infrastrukturkosten, Fragen der Sortimentspolitik usw. miteinbezogen werden müssen. Aus Sicht der Forschung ergeben sich aus unseren Ergebnissen verschiedene Ausgangspunkte für die weitere Arbeit. Dies betrifft zunächst eine Verfeinerung des Modells, z.B. durch Miteinbeziehung von Schwund. Daneben bietet sich eine Analyse für weitere Produkttypen und unterschiedliche Nachfrageverläufe an sowie ein Vergleich mit der bereits heute in Pilotprojekten praktizierten RFID-Kennzeichnung auf Kartonebene.

Literaturverzeichnis

- [AlMa92] Alstrøm, Poul; Madsen, Per: Simulation of inventory control systems. *International Journal of Production Economics* 26 (1992) 1-3, S. 125-134.
- [Ange04] Angerer, Alfred: Out of Stock: Ausmaß, Ursachen und Lösungen. *Logistik Inside* 2 (2004) 7, S. 16.
- [ATKe04] RFID/EPC: Managing the Transition (2004-2007). A.T. Kearney, Chicago (IL) 2004.
- [AtLÖ05] Atali, Aykut; Lee, Hau; Özer, Özalp: If the Inventory Manager Knew: Value of RFID Under Imperfect Inventory Information. *MSOM Conference*, Evanston (IL) 2005.
- [Boot03] Booth-Thomas, C.: The See-It-All Chip. *Time Magazine*, September 14, 2003.

- [CDGG03] Chappell, Gavin; Durdan, David; Gilbert, Greg; Ginsburg, Lyle; Smith, Jeff; Tobolski, Joseph: Auto-ID in the Box: The Value of Auto-ID Technology in Retail Stores. Auto-ID Center, MIT, Cambridge (MA) 2003.
- [CoGr04] Corsten, Daniel; Gruen, Thomas W.: Stock-Outs cause Walkouts. Harvard Business Review 82 (2004) 5, S. 26-28.
- [Coll06] Collins, Jonathan: RFID's Impact at Wal-Mart Greater Than Expected. RFID Journal, May 4, 2006.
- [ECR03] Optimal Shelf Availability: Increasing Shopper Satisfaction at the Moment of Truth. ECR Europe, Brüssel 2003.
- [EmSE91] Emmelhainz, Margaret A.; Stock, James R.; Emmelhainz, Larry W.: Consumer Responses to Retail Stockouts. Journal of Retailing 67 (1991) 2, S. 138-147.
- [FITe05] Fleisch, Elgar; Tellkamp, Christian: Inventory Inaccuracy and Supply Chain Performance: A Simulation Study of a Retail Supply Chain. International Journal of Production Economics 95 (2005) 3, S. 373-385.
- [GMA02] Invoice Accuracy: 2002 Industry Survey & Benchmarks. Grocery Manufacturers of America, Washington DC 2002.
- [GrCB02] Gruen, Thomas W.; Corsten, Daniel; Bharadwaj, Sundar: Retail Out-of-Stocks: A Worldwide Examination of Extent, Causes and Consumer Responses. Grocery Manufacturers of America / The Food Marketing Institute / CIES - The Food Business Forum, Washington DC 2002.
- [HaWM05] Hardgrave, Bill C.; Waller, Matthew; Miller, Robert: Does RFID Reduce Out of Stocks? A Preliminary Analysis. Information Technology Research Institute, University of Arkansas, Fayetteville (AR) 2005.
- [HoDa02] Hollinger, Richard C.; Davis, Jason L.: National Retail Security Survey 2001. Department of Sociology and the Center for Studies in Criminology and Law, University of Florida. Gainesville (FL) 2002.

- [KöSh06] Kök, A. Gürhan; Shang, Kevin H.: Inspection and Replenishment Policies for Systems with Inventory Record Inaccuracy. Fuqua School of Business, Duke University, Durham (NC) 2006.
- [LaDe99] Larson, Paul D.; DeMarais, Robert A.: Psychic stock: an independent variable category of inventory. *International Journal of Physical Distribution & Materials Management* 20 (1999) 7, S. 495-507.
- [LeCh04] Lee, Young M.; Cheng, Feng; Leung, Ying T.: Exploring the Impact of RFID on Supply Chain Dynamics. Winter Simulation Conference, Washington DC 2004.
- [LeÖz05] Lee, Hau; Özer, Özalp: Unlocking the Value of RFID. Graduate School of Business, Stanford University, Palo Alto 2005.
- [Ligh03] Lightburn, Anne: Unsaleables Benchmark Report. Joint Industry Unsaleables Steering Committee, Food Distributors International / Food Marketing Institute / Grocery Manufacturers of America. Washington DC 2003.
- [Pise04] Pisello, T.: The Three Rs of RFID: Rewards, Risk and ROI. *Technology-Evaluation.com*, September 27, 2004.
- [Rama00] Raman, Ananth: Retail-Data Quality: Evidence, Causes, Costs, and Fixes. *Technology in Society* 22 (2000) 1, S. 97-109.
- [SAP03] SAP Auto-ID Infrastructure. SAP AG, Walldorf 2003.
- [SiPP98] Silver, Edward A.; Pyke, David F.; Peterson, Rein: Inventory management and production planning and scheduling. Wiley, New York (NY) 1998.
- [TBKM05] Thonemann, Ulrich; Behrenbeck, Klaus; Küpper, Jörn; Magnus, Karl-Hendrik: Supply Chain Excellence im Handel. Gabler, Wiesbaden 2005.
- [WoMc03] Wong, C.Y.; McFarlane, Duncan: The Impact of Auto-ID on Retail Shelf Replenishment Policies. Auto-ID Center, MIT, Cambridge (MA), 2003.

Wahrgenommene strategische Bedeutung von RFID aus Sicht von IT-Entscheidern in Deutschland

Eine empirische Analyse

Uta Knebel, Jan Marco Leimeister, Helmut Krcmar

Lehrstuhl für Wirtschaftsinformatik (I17)
Technische Universität München
85748 Garching b. München
{knebel, leimeister, krcmar}@in.tum.de

Abstract

RFID wird aktuell vielerseits als Top IT-Thema gehandelt. Die Technologie zur automatisierten kontaktlosen Datenerfassung wird von Beratungshäusern, Anbietern und Teilen der Forschung oft als Schlüssel für die „real time enterprise“ - das Echtzeitunternehmen - oder das „Internet der Dinge“ bezeichnet. Wenig bekannt ist jedoch, wie das Thema auf Unternehmensseite wahrgenommen wird. Auf Basis einer branchenübergreifenden quantitativen Befragung unter IT-Entscheidungsträgern in Deutschland liefert der vorliegende Artikel hierzu erste Erkenntnisse. Im Vordergrund stehen die Verbreitung und Nutzung von RFID, die wahrgenommene strategische Wichtigkeit, die Relevanz der Technologie im Vergleich zu anderen Top IT-Themen und die Investitionsabsichten der Entscheider.

1 Einführung

Radiofrequenzidentifikation (RFID) wird gegenwärtig in wissenschaftlichen und nicht-wissenschaftlichen Medien heftig diskutiert. Obwohl die Technologie nicht neu ist – die erste Publikation erschien bereits 1948 – ist sie erst vor kurzem in das Bewusstsein der breiten Öffentlichkeit gerückt. Die Notwendigkeit höherer Effizienz und Sicherheit in der Supply Chain, Kostendruck, Standardisierungsinitiativen und prominente Technologiepioniere wie Wal-Mart, Metro oder Tesco haben die Aufmerksamkeit vieler Unternehmen auf neue Auto-ID-Technologien, allen voran RFID, gelenkt [Sarm04]. In der Umfrage der Society of Information

Management (SIM) unter IT-Führungskräften im Jahr 2005 [LuKN06] wurde RFID als eine der 20 wichtigsten Anwendungs- und Technologieentwicklungen eingestuft.

Die vorliegende Studie soll Anhaltspunkte darüber geben, wie IT-Führungskräfte RFID wahrnehmen und einschätzen. Nach einem kurzen Überblick über technische Grundlagen und mögliche Einsatzgebiete werden die Ergebnisse einer quantitativen Befragung unter IT-Entscheidungsträgern in Deutschland präsentiert, dabei wird insbesondere auf Aussagen über die wahrgenommene strategische Bedeutung von RFID abgezielt.

1.1 Technologie

RFID ist eine Technologie zur automatischen elektronischen Identifikation und Datenerfassung (Auto-ID). Über induktive Kopplung können Objekte oder Personen berührungslos identifiziert werden [Finke02]; [Want04]. Im Vergleich zu anderen weit verbreiteten Auto-ID-Technologien wie z.B. dem Barcode weist RFID eine Reihe vorteilhafter Eigenschaften auf [Agar01]:

- Eindeutige Identifikation: Mit dem Standard „Electronic Product Code“ (EPC) können über RFID-Tags nicht nur Produktgruppen, sondern auch einzelne Objekte identifiziert werden.
- Sichtkontaktlosigkeit: RFID-Tags können auch ohne Sichtkontakt zwischen Tag und Lesegerät ausgelesen werden, beispielsweise wenn der Tag verdeckt oder verschmutzt ist.
- Pulkfähigkeit: Innerhalb der Reichweite eines Lesegerätes können mehrere Tags gleichzeitig ausgelesen werden.
- Speicherkapazität: Auf RFID-Tags können weit mehr Informationen als eine Identifikationsnummer hinterlegt werden.
- Dynamische Information: Wiederbeschreibbare RFID-Tags ermöglichen die Aktualisierung der hinterlegten Informationen.

Vollständig ausgereift sind viele am Markt erhältliche RFID-Lösungen jedoch nicht. Für etliche Themen steht eine Lösung noch aus. Problemfelder hierbei sind:

- Beeinflussung durch Materialien in unmittelbarer Umgebung: Metalle oder Flüssigkeiten führen zur Dämpfung oder Ablenkung von Funksignalen und können damit die Funktion von RFID-Tags beeinträchtigen.
- Vielzahl an Standards: Das von der Standardisierungsorganisation EPC Global vorgelegte Rahmenwerk [Epcg04] hat sich noch nicht ausreichend etabliert. Es werden nach wie vor diverse unterschiedliche Standards verwendet.
- Datenmenge: Die Sammlung und Übermittlung erweiterter Objektinformationen führt unweigerlich zu steigenden Datenvolumen. Zum einen ist die technische Integration dieser Informationen in die Informationssysteme von Unternehmen, üblicherweise über Middleware, noch problematisch. Zum anderen ist bei vielen Unternehmen noch unklar, wie die gewonnenen Daten effizient verwertet und gewinnbringend eingesetzt werden können, z.B. durch das Angebot neuer Dienstleistungen am Markt oder durch neue bzw. optimierte Kontroll-, Planungs- und Steuerungsprozesse.

1.2 RFID als Wegbereiter für das Echtzeitunternehmen und innovative neue Lösungen

RFID birgt das Potential, Informationsflüsse zu beschleunigen, anzureichern, zu automatisieren und somit grundlegend zu verändern. Ansätze zur Verbesserung von Prozesseffizienz und -effektivität lassen sich direkt aus den Eigenschaften der Technologie erkennen. Sichtkontaktlosigkeit macht eine Auslesung möglich, ohne dass das Objekt erst nach dem Lesegerät ausgerichtet werden muss. Dies erspart manuelle Intervention, wie sie bei Barcodes zur Herstellung des Sichtkontakts häufig nötig ist. Eine Auslesung ist sogar bei bereits verbauten Teilen möglich. Pulklesung kann die Durchlaufzeit verkürzen. Höhere Speicherkapazität erlaubt die Vorhaltung und Anpassung erweiterter Informationen über den gesamten Produktlebenszyklus hinweg. In Verbindung mit Sensoren könnte der Tag zusätzlich aktuelle Umweltinformationen abspeichern, beispielsweise Temperatur, Luftfeuchtigkeit oder Druck [Hall02], und so die betriebliche Prozesstransparenz weiter erhöhen.

Über die Effizienzsteigerung bestehender Prozesse hinaus kann RFID die Basis für neue Produkte und Dienstleistungen sein [StFl05]. Denkbar ist ein Einsatz z.B. für Waren(rück)verfolgung, Lagerverwaltung, Produktmontage, Fälschungsschutz, Umweltüberwachung, Ortung, Diebstahl – oder Sabotagesicherung. [WiDe04]. Dies soll im Folgenden durch einige ausgewählte Beispiele verdeutlicht werden:

- Healthcare: Erhöhte Fälschungssicherheit für Medikamente. Eindeutige Identifizierung und lückenlose Verfolgung könnten die Authentizität von Medikamenten sicherstellen. Dadurch würden zum einen Konsumenten besser vor schädlichen oder wirkungslosen Medikamenten geschützt, zum anderen würde es Fälschern erschwert, unechte Medikamente in Umlauf zu bringen [Usfo04].
- Automobilbranche: Selektive Rückrufaktionen. Bisher müssen Automobilhersteller bei vermutetem Defekt einer im Auto verbauten Komponente für gewöhnlich alle Fahrzeuge zurückrufen, die im kritischen Zeitraum produziert wurden. Würde jede einzelne Komponente im Produktionsprozess verfolgt, könnte der Hersteller die betroffenen Fahrzeuge exakt benennen und die kosten- und zeitintensive Inspektion aller anderen Fahrzeuge vermeiden. Unternehmen wie z.B. BMW sehen hier eine große Chance für Kosteneinsparungen bei Rückrufaktionen durch RFID.
- Handel: Der intelligente Einkaufswagen. Ohne zeitintensiven Rechercheaufwand ist es für Konsumenten oft schwierig, sich detailliert über Eigenschaften und Inhaltsstoffe von Produktangeboten des Handels zu informieren. Diese Informationen sind jedoch von großer Bedeutung, wenn der Käufer an Krankheiten, beispielsweise Allergien oder Diabetes leidet. Hilfreich wäre es, wenn die Produktinformationen vor Ort über den Einkaufswagen ausgelesen und angezeigt werden könnten, möglicherweise in Verbindung mit Allergiehinweisen oder Zubereitungsvorschlägen. Die Metro Group experimentiert zur Zeit in ihrer RFID-Pilotfiliale mit intelligenten Einkaufswagen, sogenannten Smart-Carts.

In den oben beschriebenen sowie in vielen weiteren Anwendungsbereichen kann RFID dazu beitragen, die Kluft zwischen der Realwelt und ihrem Abbild in Informationssystemen zu überbrücken [Hall02] [FIMa05]. Informations- und Warenfluss nähren sich weiter an. Dies kann der erste Schritt zur „real time enterprise“ sein [KuTh05], von dem sich Unternehmen unter anderem Prozessoptimierung, höhere Entscheidungsqualität durch präzisere und schnellere Information sowie eine höhere Integration aller Partner in der Supply Chain erhoffen.

2 Literaturanalyse

2.1 Wahrgenommene strategische Bedeutung von RFID

Michael Porter beschreibt Strategie als „performing different activities from rivals’ or performing similar activities in different ways“, und betont, dass Effizienz im operativen Geschäft zwar ein wesentlicher Erfolgsfaktor für die Profitabilität eines Unternehmens ist, deshalb allein aber noch nicht das Merkmal einer Strategie erfüllt [Port96]. Demzufolge können der Erfolg der Metro bei der Verbesserung betrieblicher Abläufe und Kosteneinsparungen durch RFID [Coll05] sowie positive ROIs von RFID-Systemen in der Automobilindustrie [KrSc05] noch nicht als strategisch eingestuft werden. Die Anwendungen entwickeln sich jedoch kontinuierlich weiter. Wie oben erläutert, kann RFID Unternehmen in die Lage versetzen, neue Services anzubieten, und sich so von der Konkurrenz abzuheben. Verschiedene große Unternehmensberatungen betonen die strategische Bedeutung von RFID. Folgt man der Darstellung von Gartner, so könnte RFID nicht nur die Verfolgung von Objekten entlang der gesamten Lieferkette revolutionieren [RoFL05], sondern birgt darüber hinaus großes Potential, die Geschäftsstrategien von Unternehmen entscheidend zu verändern [WoPH03].

Doch wie wird das strategische Potential von (zukünftigen) Anwendern selbst eingeschätzt? Die Forschungsfragen (FF) 1 und 2 greifen diese Thematik auf:

FF1: Wie weit ist RFID in Unternehmen verbreitet?

FF2: Wie schätzen CIOs die strategische Bedeutung von RFID ein?

2.2 Determinanten der wahrgenommenen strategischen Bedeutung

Zunächst wurde RFID vor allem im militärischen Bereich genutzt [Schm04], Stück für Stück etablierte es sich auch in anderen Branchen. Unternehmen erhoffen sich von RFID zumeist die Optimierung und Rationalisierung des Supply Chain Management [Lang04]. ABI research sagt voraus, dass sich bestimmte Branchen besonders im RFID-Umfeld engagieren werden, darunter Konsumgüter und Handel, Automobil, Militär und Verteidigung [Mase03]. Dies lässt vermuten, dass manche Branchen mehr zur Verwendung von RFID neigen, und soll in Forschungsfrage 3 behandelt werden:

FF3. Beeinflussen die Eigenschaften eines Unternehmens wie Branche und Größe die wahrgenommene strategische Wichtigkeit von RFID?

Die Theorie der Innovationsdiffusion nach Rogers beschreibt unter anderem drei Eigenschaften einer Innovation, die ihre Akzeptanz und Verbreitung beeinflussen: *relativer Vorteil*,

Erprobbarkeit und *Beobachtbarkeit* [Roge95]. Unter der Annahme, dass eine Innovation nur angenommen wird, wenn ihm der Einzelne eine gewisse Bedeutung oder Nützlichkeit zuschreibt, könnten diese Faktoren auch die Wahrnehmung der strategischen Wichtigkeit einer Innovation, in vorliegendem Fall RFID, beeinflussen. *Erprobbarkeit* ist der Grad, zu dem eine Innovation getestet oder erfahren werden kann. Je besser das Individuum versteht, wie die Innovation für seine Person und für sein Umfeld funktioniert, desto eher wird er sie annehmen. *Beobachtbarkeit* ist der Grad, zu dem die Auswirkungen einer Innovation für andere sichtbar sind. Die momentane Verbreitung von RFID ist gering [Lang04], jedoch haben viele Unternehmen Pilotprojekte initiiert, was eine gewisse Erprobbarkeit schafft. In Kombination mit Demonstrationssystemen von RFID-Anbietern und anderer verfügbarer Information stellen diese Pilotprojekte Beobachtbarkeit für andere potentielle Anwender dar. Folglich lautet Forschungsfrage 4:

FF4: Beeinflusst der Grad der Erfahrung mit RFID die wahrgenommene strategische Bedeutung?

Relativer Vorteil beschreibt den Grad, zu dem eine Innovation als besser wahrgenommen wird als das Konzept, das sie verdrängt. Einen ähnlichen Ansatz legt Davis seinem Technology Acceptance Model (TAM) zu Grunde, in dem er die wahrgenommene Nützlichkeit einer Innovation als Determinante ihrer Annahme beschreibt [Davi89]. Wahrgenommener Nutzen wurde auch von Iacovou und Benbasat in ihrer Studie über die Adoption von EDI als relevant identifiziert [IaBe95]. In ihrer Arbeit zur Adoption von RFID berücksichtigen Sharma und Citrus diese Größe ebenfalls [ShCi05]. In dieser Studie bezeichnet relativer Vorteil den potentiellen Nutzen und Verbesserungen durch RFID im Vergleich zur Barcode-Technologie. Forschungsfrage 5 behandelt die wahrgenommenen Potentiale von RFID:

FF5: Beeinflussen wahrgenommene Potentiale von RFID die wahrgenommene strategische Bedeutung?

2.3 Investitionsbereitschaft in RFID

Seit Großhändler wie Wal-Mart in den USA, Metro in Deutschland oder Tesco in Großbritannien, sowie Behörden wie das US Verteidigungsministerium RFID zur Schlüsseltechnologie erklärt haben, überbieten sich die Marktvorhersagen gegenseitig [Lang04]. Frost & Sullivan sagen ein Marktwachstum von 20% bis 2010 im Vergleich zu 2003 voraus [FrSu04], Accenture sogar 40% [Acce05]. In einer aktuellen Studie von AMR research

geben 69 % der Befragten an, die Evaluation, Pilotierung oder Einführung von RFID zu planen. Es wird ebenfalls ein Marktwachstum von 40% innerhalb der nächsten zwei Jahre vorausgesagt [Reil05]. In Forschungsfrage 6 soll überprüft werden, ob potentielle Anwender diese Sicht teilen, Forschungsfrage 7 verbindet die oben behandelte wahrgenommene strategische Bedeutung mit der Investitionsbereitschaft.

FF 6: Planen CIOs, in RFID zu investieren?

FF 7: Beeinflußt die wahrgenommene strategische Wichtigkeit die Investitionsentscheidung?

2.4 RFID im Vergleich zu anderen IT-Themen

In der SIM-Befragung unter IT-Entscheidungsträgern über die wichtigsten Trends der Anwendungs- und Technologieentwicklung aus dem Jahr 2005 nahm RFID den 16. Platz ein. Dies bescheinigt dem Thema zwar Relevanz, jedoch scheint es nicht von primärem Interesse zu sein. Um die Ergebnisse der oben genannten Forschungsfragen ins Verhältnis zu setzen, werden in Forschungsfrage 8 die Top IT-Themen aus Sicht der CIOs erhoben.

FF 8: Wie relevant ist RFID im Vergleich zu anderen IT-Themen?

3 Forschungsdesign

Die aus der Literatur abgeleiteten Forschungsfragen wurden in 15 semistrukturierten Interviews mit IT-Führungskräften aus verschiedenen Branchen auf ihre Relevanz überprüft. Die Ergebnisse der Interviews flossen neben der Literaturanalyse in die Konstruktion des Fragebogens ein. Die erste Version des Fragebogens wurde in zwei Pre-Tests mit 8 IT-Experten aus Wissenschafts- und Unternehmensumfeld getestet und weiterentwickelt. Die Einladung zur Teilnahme an der Befragung erfolgte postalisch oder per E-Mail, eine Nachfassaktion schloss sich nach drei Wochen an. Design und Durchführung der Umfrage folgen mit leichten Anpassungen dem Modell der Phasen der Informationsgewinnung nach Nieschlag, Dichtl und Hörschgen [NiDH02] (Abbildung 1). Tabelle 1 fasst die wichtigsten Gestaltungsparameter der Studie zusammen.

Forschungsansatz	Explorative Studie
Forschungsmethode	Quantitative Online-Umfrage
Laufzeit	11. November 2005 – 2. Januar 2006
Art der Stichprobe	convenience sample (nicht repräsentativ)
Zugang zur Umfrage	Zugang auf Einladung mit persönlichem Code

Zielgruppe	IT-Entscheidungsträger aller Branchen in Deutschland
Versendete Einladungen	3995 (davon 3546 postalisch, 449 per E-Mail)
Rücklaufquote	Ca. 12 %

Tabelle 1: Gestaltungsparameter der Studie (Quelle: eigene Darstellung)

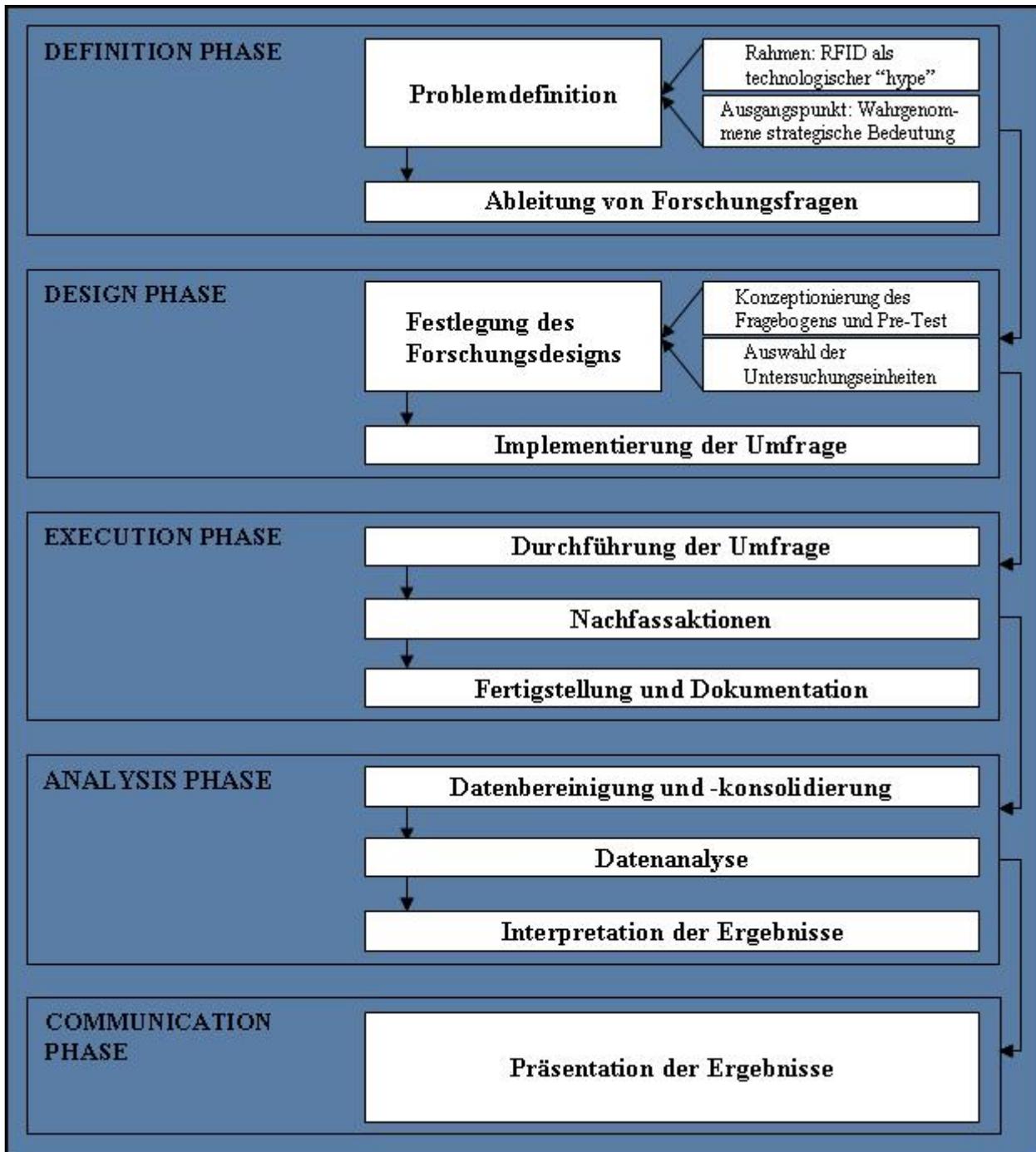


Abbildung 1: Durchführung der Umfrage (Quelle: bearbeitet nach [NiDH02])

4 Emprische Ergebnisse

4.1 Merkmale und Soziodemographie der Teilnehmer

Alle vollständigen Datensätze wurden darauf überprüft, ob unter den Teilnehmern ein Unternehmen mehrmals vertreten war. War dies der Fall, so wurde der Datensatz des CIO ausgewählt und die restlichen verworfen. War der CIO nicht vertreten, so wurde in der Reihenfolge der Nennung der CEO bzw. IT-Leiter / Bereichsleiter in die Auswertung aufgenommen. In die Auswertung flossen 463 Datensätze ein, aufgrund der Auswahl handelt es sich dabei um Vertreter von 463 unterschiedlichen Unternehmen. Alle außer 10 der teilnehmenden Personen waren männlich, die Altersgruppe von 41-50 Jahre dominierte. 70 % der Befragten sind CIO oder üben eine andere leitende Position aus. Als Hauptgeschäftsfeld ihres Unternehmens¹ nannten die Befragten Produktion / Herstellung (36.9%), Automobil (18.4%), Dienstleistung (16.0%), Handel (14.3%), Konsumgüter (11.7%), IT (11.7%), Logistik (11.0%), Pharma (8.4%) und sonstiges (22.9%). Fast 75% kommt aus Unternehmen mit weniger als 2500, 11 % aus Unternehmen mit mehr als 10000 Mitarbeitern.

4.2 Ergebnisse: Verbreitung von RFID (FF 1)

Ihre Sachkenntnis über RFID und dessen Anwendungsmöglichkeiten schätzen die meisten Befragten als gut oder sehr gut ein (43 %), 36% als mittel und 20% als gering. Erfahrung mit RFID-Systemen liegt wenig vor. Nur bei ca. 10 % der Unternehmen ist ein RFID-System bereits im Einsatz oder in der Implementierung, die meisten davon sind Großunternehmen (> 10000 Mitarbeiter). Ein Drittel führt aktuell Potentialanalysen durch. 14% haben sich gegen eine RFID-Einführung entschieden, zumeist aus Kostengründen und wegen fehlender Anwendungspotentiale. Die verbleibenden 38% haben sich noch nicht mit der Thematik beschäftigt, viele nehmen eine abwartende Haltung ein bis das Thema von Geschäftspartnern angestoßen wird.

4.3 Ergebnisse: Wahrgenommene strategische Bedeutung von RFID (FF 2)

Die Teilnehmer wurden gebeten, auf einer fünfstufigen Skala einzuschätzen, ob RFID für ihr Unternehmen strategische Bedeutung hat (1=stimme voll und ganz zu bis 5=stimme überhaupt nicht zu). Der Mittelwert zeigt keine deutliche Tendenz (3,24), die 283 Antwortenden scheinen die strategische Bedeutung beinahe neutral einzuschätzen. Um einen tieferen Eindruck zu gewinnen, wurden die Teilnehmer anschließend gefragt, ob RFID Einfluss auf die

¹ Mehrfachangaben möglich, Summe der Prozentsätze kann daher über 100% liegen.

Kernkompetenzen ihres Unternehmens haben könnte. Ähnlich wie bei der strategischen Bedeutung äußern die Teilnehmer weder besondere Zustimmung noch Ablehnung, ob RFID eine bessere Ausnutzung bestehender Kernkompetenzen ermöglichen könne (Mittelwert=3,13). In Bezug zum Aufbau neuer Kernkompetenzen zeigt sich eine leicht negative Tendenz (Mittelwert=3,53). Die Erschließung von Wettbewerbsvorteilen wird ebenfalls neutral bewertet. (Mittelwert=2,87). In den folgenden Abschnitten wird nur die wahrgenommene strategische Bedeutung weiter analysiert, da sie als inhaltliche Zusammenfassung der anderen genannten Größen angenommen wird.

Mit Blick auf die Zukunft zeigt sich ein deutlicher Trend. Die große Mehrheit der Teilnehmer (82%) ist der Überzeugung, dass die Bedeutung von RFID steigen wird. 60% glauben sogar, dass die Technologie für ihr Unternehmen innerhalb der nächsten 6 Jahre erfolgskritisch wird.

4.4 Ergebnisse: Determinanten der wahrgenommenen strategischen Bedeutung (FF 3-5)

4.4.1 Eigenschaften des Unternehmens und des Befragten

Abbildung 2 zeigt die wahrgenommene strategische Bedeutung von RFID nach Branchen.

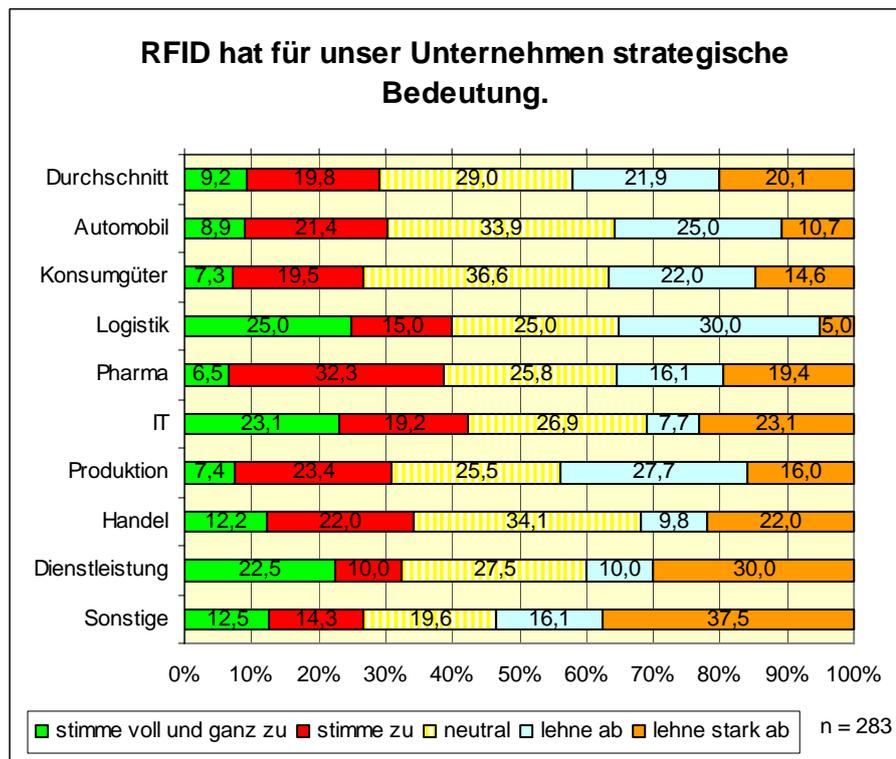


Abbildung 2: Wahrgenommene strategische Bedeutung von RFID nach Branchen (Quelle: eigene Darstellung)

Auch wenn die Unterschiede nicht drastisch sind, so zeigt sich doch, dass die wahrgenommene strategische Bedeutung über die Branchen hinweg variiert. Vor allem in Logistik, Pharma und IT sind die Zustimmungsraten überdurchschnittlich

Wird die wahrgenommene strategische Wichtigkeit nach Unternehmensgröße ausgewertet, so entsteht der Eindruck, dass sie mit steigender Unternehmensgröße als höher eingestuft wird. Dies bestätigt sich in einer Korrelationsanalyse, die eine schwache negative Beziehung zwischen „RFID hat für unser Unternehmen strategische Bedeutung“² und der Unternehmensgröße anzeigt (Rangkorrelation nach Spearman; $r=-0,221$; $p<0,001$). Zwischen den Eigenschaften des Antwortenden wie Alter und RFID-Kenntnis und der wahrgenommenen strategischen Bedeutung scheint keine Beziehung zu bestehen.

4.4.2 Erfahrung mit RFID

Vergleicht man die Mittelwerte der wahrgenommenen strategischen Bedeutung nach RFID-Erfahrung, ergeben sich unterschiedliche Werte für alle Erfahrungsgruppen (Tabelle 2). Unternehmen mit laufenden RFID-Tests beurteilen die strategische Wichtigkeit neutral. Jene, die nach Tests keine Einführung planen, messen der Technologie auch wenig strategische Bedeutung zu. Im Gegensatz dazu sehen Firmen, die eine Einführung beabsichtigen, durchaus strategische Bedeutung in RFID. Am meisten überzeugt ist die Gruppe derer, die zum Zeitpunkt der Umfrage gerade ein RFID-System implementieren. Sie wird sogar höher beurteilt als bei der Gruppe, die bereits RFID im Einsatz hat.

Erfahrung mit RFID	Strategische Bedeutung von RFID	
	Mittelwert	Std. Abweichung
RFID im Einsatz	2,7	1,291
RFID wird realisiert	2,21	1,122
Test abgeschlossen, Einführung geplant	2,38	1,115
Tests laufend	3,09	1,028
Test abgeschlossen, Einführung nicht geplant	4,40	0,853

Anmerkung: 5-stufige Skala von „stimme voll und ganz zu“=1 bis „lehne stark ab“=5; n=463

Tabelle 2: Kreuztabelle über RFID Erfahrung & wahrgenommene strategische Bedeutung (Quelle: eigene Darstellung)

4.4.3 Wahrgenommene Potentiale von RFID

Im Durchschnitt sind die Befragten der Meinung, dass RFID einige Verbesserungspotentiale für Unternehmensabläufe birgt, allen voran Fehlerreduzierung, Automatisierung und eine höhere

² Fünfstufige Skala mit 1=stimme voll und ganz zu bis 5=lehne stark ab.

Konsistenz der Datenbestände über verschiedene Anwendungen hinweg. Eine Rangkorrelationsanalyse nach Spearman zeigt einen hoch signifikanten positiven Zusammenhang zwischen allen abgefragten Potentialen und der wahrgenommenen strategischen Bedeutung. Die Stärke der Korrelationen ist jedoch als schwach einzustufen (Tabelle 3).

RFID hat Potential zur...	Mittelwert	Std. Abweichung	Spearman's Rho	Signifikanz -wert
Reduzierung von Fehlern (n=285)	1,89	0,84	0,219	0,000
Automatisierung / weniger Personaleinsatz (n=285)	2,17	0,98	0,254	0,000
Konsistenter Datenbestand in allen relevanten Unternehmensanwendungen (n=278)	2,19	0,95	0,197	0,001
Reduktion von Fehlbeständen (n=284)	2,24	0,95	0,157	0,009
Beschleunigung der Warenflüsse (n=282)	2,24	1,03	0,210	0,000
Verbesserter Kundenservice (n=282)	2,35	1,04	0,252	0,000
Qualitätssteigerung (n=282)	2,38	1,08	0,291	0,000
Optimierung der Lagerhaltung (n=279)	2,47	1,02	0,192	0,001
Reduzierung von Fälschungen (n=263)	2,76	1,28	0,198	0,001

Anmerkung: 5-stufige Skala von "stimme voll und ganz zu"=1 bis "lehne stark ab"=5

Tabelle 3: Wahrgenommene Potentiale von RFID und deren Zusammenhang mit wahrgenommener strategischer Bedeutung von RFID (Quelle: eigene Darstellung)

4.5 Ergebnisse: Investitionsbereitschaft in RFID (FF 6-7)

Die Teilnehmer der Befragung wurden gebeten, die Aussage „RFID ist ein Thema, in das wir investieren werden“ anhand einer fünfstufigen Skala („stimme voll und ganz zu“=1 bis „lehne stark ab“=5) für ihr Unternehmen zu bewerten. Der Mittelwert zeigte eine positive Tendenz (2,70; S= 1,21). Eine Korrelationsanalyse zwischen „RFID ist ein Thema, in das wir investieren werden“ und „RFID hat für unser Unternehmen strategische Bedeutung“ ergibt einen starken und hoch signifikanten positiv gerichteten Zusammenhang zwischen den beiden Aussagen (Rangkorrelation nach Spearman, $r=0,630$; $p<0,001$).

Auf die Frage hin, wie sich das RFID-Budget des jeweiligen Unternehmens über die nächsten Jahre hinweg entwickeln wird, gaben die Befragten an, dass eine Steigerung zu erwarten sei, in

erster Linie in einem Zeithorizont von drei bis fünf Jahren. Die höchsten Steigerungsraten werden in der Pharmaindustrie erwartet (Abbildung 3).

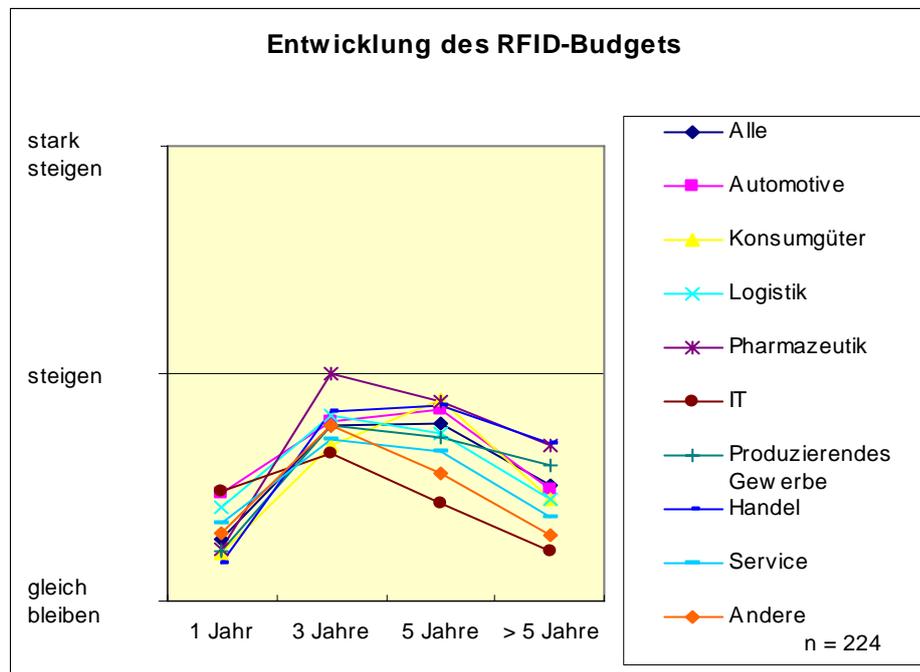


Abbildung 3: Entwicklung der RFID Budgets (Quelle: eigene Darstellung)

4.6 Ergebnisse: RFID und Top IT Themen (FF 8)

Die Aussage „RFID ist eines unserer Top IT-Themen“ wurde im Durchschnitt abgelehnt (Mittelwert=3,87, S=1,05). Dies bestätigte sich auch im Antwortspektrum auf eine offene Frage nach den drei wichtigsten IT-Themen im Unternehmen. RFID, Tracking & Tracing und verwandte Begriffe wurde hier nur von ca. 10% der Teilnehmer erwähnt.

4.7 Blick in die Zukunft: RFID-Visionen

In einer offenen Frage wurden die Teilnehmer anschließend nach der RFID-Vision gefragt, die sie mit ihrem Unternehmen verfolgen möchten. Fast alle Stellungnahmen konnten in folgende 5 Inhaltskategorien gruppiert werden:

Typ 1: Das gläserne Unternehmen (18,6%). Dieser Visionstyp strebt nach Echtzeitinformationen und lückenloser Transparenz aller Prozesse über die gesamte Lieferkette hinweg, geäußert wurde z.B. „durchgängige Informationskette durch RFID, Lieferanten und Partner eingebunden - die 'gläserne' Fabrik“.

Typ 2: Konkrete Anwendungsbereiche (9,5%). Vertreter dieser Gruppe zielen auf die Optimierung sehr konkreter Bereiche oder Prozesse ab. Beispielhaft sind die Formulierungen

„Lagerverwaltung mit Positionsbestimmung der Stapler über Transponder im Boden“ oder „Einsatz von RFID in der Instandhaltung um wartungsrelevante Informationen an Maschinenteile zu bekommen“.

Typ 3: Barcode ersetzen (3%). RFID soll bestehende Barcodeanwendungen ersetzen.

Typ 4: Kundengesteuerte Vision (3,2%). Diese Gruppe macht ihre RFID-Aktivitäten vollständig von Kundenwünschen abhängig und plant keine eigenen Vorstöße.

Typ 5: Keine Vision (26,8%). Diese Gruppe gab explizit an, keine RFID-Vision zu haben oder diese nicht zu benötigen. Geäußert wurde z.B. „wir haben keine und benötigen keine“ oder „Wir beobachten die Technologie und entscheiden bei Bedarf“.

Ca. 3,9% der Antworten konnten keine der oben genannten Kategorien zugeordnet werden. Die verbleibenden 37,2% machten keine Angaben. Es ist unklar, ob in diesen Fällen keine Vision vorliegt (vgl. Typ 5), oder ob diese nicht offengelegt werden wollte.

5 Fazit

Die vorliegende Arbeit präsentiert quantitatives Datenmaterial über die Wahrnehmung von RFID bei IT-Entscheidungsträgern. Die Verbreitung von RFID ist sehr gering. Viele Entscheidungsträger haben sich bereits mit der Thematik beschäftigt, sind aber von einer Implementierung noch weit entfernt. Es wird erwartet, dass die Bedeutung von RFID für Unternehmen innerhalb der nächsten Jahre stark ansteigt. Aktuell wird der Technologie jedoch wenig strategische Bedeutung und eher geringe Priorität zugemessen. Die Wahrnehmung der strategischen Bedeutung scheint im vorliegenden Fall beeinflusst zu sein durch:

- **Branche:** In einigen Branchen, insbesondere Logistik, Pharma und IT wird RFID von einem deutlich höheren Anteil als strategisch bedeutsam eingestuft.
- **Unternehmensgröße:** Vertreter größerer Unternehmen messen dem Thema RFID tendenziell mehr Bedeutung zu als Vertreter kleinerer Unternehmen.
- **Eigenschaften:** Verbinden die Befragten bestimmte Eigenschaften (z.B. Qualitätsverbesserung, Fehlerreduktion) stärker mit RFID, so fällt auch die Beurteilung der strategischen Bedeutung höher aus.

Wie erwartet, geht eine hohe wahrgenommene strategische Bedeutung auch mit einer erhöhten Investitionsbereitschaft einher. Da in den nächsten Jahren Budgetsteigerungen für RFID erwartet werden, ist eine tatsächliche Investition wahrscheinlich. Abbildung 4 fasst diese Ergebnisse zusammen.

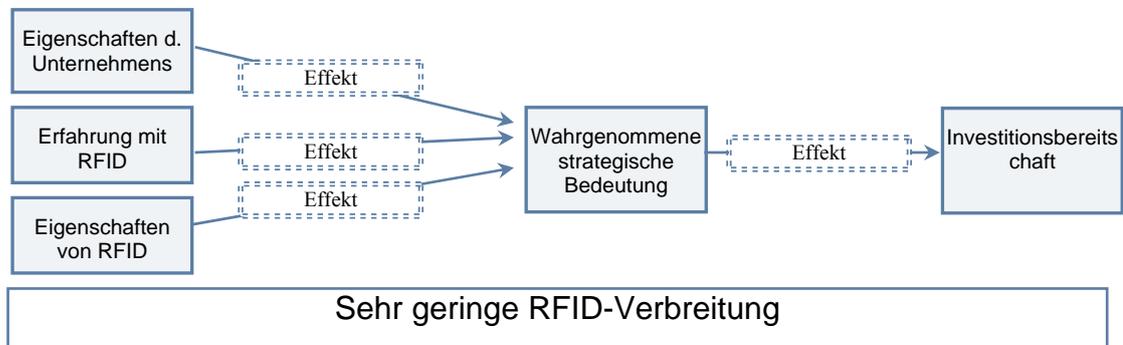


Abbildung 4: Ergebnisse der Studie (Quelle: eigene Darstellung)

6 Handlungsempfehlungen für potentielle RFID-Anwender

Aus den Ergebnissen der Umfrage sowie der begleitenden Experteninterviews lassen sich Handlungsempfehlungen für potentielle RFID-Anwender ableiten. Für viele Firmen ist keine sofortige Aktion nötig, da u. U. keine interne Anwendungsmöglichkeit besteht. Wesentliche Vorteile von RFID, beispielsweise in der Warenverfolgung, werden jedoch über Unternehmensgrenzen hinweg realisiert. In solchen Verbänden liegt die Entscheidung über Technologieeinführungen oft beim stärksten Partner der Kette. Prominentes Beispiel hierfür sind die Großhandelsunternehmen Metro, Wal-Mart und Tesco, die ihre Lieferanten zur RFID-Verwendung zwangen. Hiervon betroffenen Unternehmen ist deshalb anzuraten, die Aktivitäten direkter Partner in der Lieferkette zu beobachten und sich unabhängig über die Technologie und mögliche Einsatzbereiche speziell in ihrem Unternehmen zu informieren, sowie Planungen zur Einführung vorzubereiten. Dies ermöglicht ihnen, auch im Falle einer oktroyierten Einführung gestaltend mitzuwirken und durch gezielte Erweiterung Prozessverbesserungen oder Wettbewerbsvorteile für sich selbst erzielen zu können. Die Betrachtung sollte sich nicht auf die reine Hardware beschränken, sondern insbesondere auch mögliche Auswirkungen auf die IT-Strategie und bestehende Anwendungen einbeziehen. Beim Test erster Anwendungen ist Geduld zentral, da eine Amortisation der Lösung meist in Zeiträumen von mehr als einem Jahr zu erwarten ist. Unternehmen erwarten durchweg, dass die Bedeutung von RFID in den

nächsten Jahren steigen wird, gute Vorbereitung ist also umso wichtiger. Möglicherweise entpuppt sich die Technologie als schlafender Riese.

7 Grenzen der Untersuchung und weiterer Forschungsbedarf

Grenzen der Untersuchung finden sich insbesondere in zwei Aspekten. Erstens sind die im convenience sample erhobenen Daten nicht zwangsläufig repräsentativ für die Gesamtheit deutscher IT-Entscheidungsträger. Zweitens erfasst die Befragung individuelle Wahrnehmungen, die unter Umständen nicht die Einschätzung aller am RFID-Thema involvierten Entscheider in einem Unternehmen widerspiegeln. In weiteren Arbeiten könnte versucht werden, „strategische Bedeutung“ als latentes Konstrukt zu operationalisieren und zu explorieren, inwieweit sich Ergebnisse so von der direkten Abfrage unterscheiden. Desweiteren könnten Kausalitäten überprüft sowie weitere Einflussfaktoren und intermediierende Konstrukte einbezogen werden. Unabhängig hiervon besteht erheblicher Forschungsbedarf im Bereich der RFID-Technologie-Komponenten, insb. im Bereich der Middleware. Gleiches gilt für die Entwicklung und Umsetzung RFID-basierter IT-Strategien, die ebenfalls noch weit von einer konkreten Umsetzung im Unternehmenskontext entfernt zu sein scheinen.

Literatur

- [Acce05] Accenture. Pushing the Pace - How Leaders Are Putting RFID to Work. Accenture, München 2005.
- [Agar01] Agarwal, V. Assessing the benefits of Auto-ID Technology in the Consumer Goods Industry. Cambridge University Auto ID Centre, Cambridge 2001.
- [Coll05] Collins, J..Metro Group Reaps Gains From RFID. In: RFID Journal, 24.01.2005.
- [Davi89] Davis, F.D.. Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology. MIS Quarterly 13 (1989) 3, 319-339.
- [Epcg04] EPC Global (2004). The EPC Global Network. http://www.epcglobalinc.org/news/EPCglobal_Network_Overview_10072004.pdf, Abruf am 24.07.2006

- [Fink02] Finkenzeller, K.. RFID-Handbuch, 3 Aufl.. Hanser, München, Wien 2002.
- [FlMa05] Fleisch, E.; Mattern, F.. Das Internet der Dinge - Ubiquitous Computing und RFID in der Praxis. Springer, Berlin, 2005.
- [FrSu04] Frost & Sullivan. World RFID-based Application Markets (A686-11). Frost & Sullivan, Palo Alto 2004.
- [Hall02] Haller, S.; Hodges, S. The Need for a Universal SmartSensor Network. Cambridge University Auto ID Centre, Cambridge 2002.
- [IaBe95] Iacovou, C.; Benbasat, I. Electronic Data Interchange and Small Organizations: Adoption and Impact of Technology. In: MIS Quarterly, 19 (1995) 4 , 465-485.
- [KrSc05] Kraft, J.; Schauler, C.. RFID setzt sich durch. Information Week, 14.04.2005.
- [KuTh05] Kuhlmann, B.; Thielmann, H. Real Time Enterprise in der Praxis, Springer 2005
- [Lang04] Lange, V.. Perspektiven für die Nutzung der RFID-Technologie in Supply Chain Management und Logistik. In: IM, 19 (2004) 4, 20-26.
- [LuKN06] Luftman, J.; Kempaiah, R.; Nash, E. (2006): Key Issues for IT Executives 2005 In: MIS Quarterly Executive, Vol. 5 (2006) Nr. 2, S. 27-45.
- [Mase03] Maselli, J. (2003). ABI: RFID Market Poised for Growth. In: RFID-Journal, 18.07.2003.
- [NiDH02] Nieschlag, R.; Dichtl, E.; Hörschgen, H.. Marketing, 19., überarbeitete und ergänzte Auflage Aufl., Duncker&Humboldt, Berlin 2002
- [Port96] Porter, Michael E. What is Strategy? Harvard Business Review, 74(1996)6, 61 - 78.
- [Reil05] Reilly, K.. AMR Research Survey Finds 69% of Respondents Plan to Evaluate, Pilot, or Implement RFID in 2005: AMR research 2005.

- [RoFL05] Roussel, A.-M.; Fenn, J.; Linden, A.. Gartner's Radar Screen Pinpoints Hot Technologies Likely to Grab Investors' Attention. Gartner Research, Stanford 2005.
- [Roge95] Rogers, E.. Diffusion of Innovations, 4 Aufl., The Free Press, New York, 1995.
- [Sarm04] Sarma, S.. Integrating RFID. QUEUE (October 2004), 50-57.
- [Schm04] Schmid, F. Radio Frequency Identification: Standards - Produkte - Vorgehen. In: IM 19 (2004) 4, 6-11.
- [ShCi05] Sharma, A.; Citurs, A.. Radio Frequency Identification (RFID) Adoption Drivers - A radical Innovation Adoption Perspective. In: Proceedings of the Eleventh Americas Conference on Information Systems, Omaha, NE, USA, August 11th - 14th 2005.
- [StFl05] Strassner, Martin; Fleisch, Elgar: Innovationspotenzial von RFID für das Supply-Chain Management. In: Wirtschaftsinformatik 47 (2005) 1, S. 45-54.
- [Usfo04] U.S. Food and Drug Administration. COMBATING COUNTERFEIT DRUGS: U.S. Food and Drug Administration 2004.
- [Want04] Want, R. (2004). The Magic of RFID. In: QUEUE (October 2004), 40-48.
- [WiDe04] Wilding, R.; Delgado, T. (2004). RFID Demystified: Supply-Chain Applications. In: Logistics & Transport Focus, 6 (2004)4.
- [WoPH03] Woods, J.; Peterson, K.; Hirst, C.. Maturing Open RFID Applications Will Reshape SCM. Gartner Research, Stanford 2003.

Einführung in den Track

eGovernment - effektiv, wirtschaftlich, transparent

Prof. Dr. Jörg Becker

Universität Münster

Willi Kaczorowski

Cisco Systems

Prof. Dr. Helmut Krcmar

TU München

Prof. Maria A. Wimmer

Universität Koblenz-Landau

Ein großes Potenzial für Bürokratieabbau und Kosteneinsparungen in der öffentlichen Verwaltung liegt in der Reorganisation existierender Abläufe und deren Unterstützung durch moderne Informations- und Kommunikationstechnologien - kurz Electronic Government. Allerdings wird dieses Potenzial bisher nur in sehr geringem Maße ausgeschöpft, da die Verantwortlichen häufig vor den notwendigen Investitionen zurückschrecken. Dies liegt daran, dass sich der Nutzen von Informationstechnologie bisher nicht hinreichend nachweisen lässt: Sowohl die Wirkung von IT auf Geschäftsprozesse als auch die Geschäftsprozesse selbst sind für Entscheider nicht transparent. Daher lassen sich sinnvolle Investitionen in größerem Umfang gerade vor dem Hintergrund umstrittener Großprojekte wie der ALGII-Software oder des Internetauftritts der Bundesagentur für Arbeit vor der Öffentlichkeit und der Politik nur noch schwer rechtfertigen.

Schwerpunkt dieses Tracks bildet daher die Fragestellung, wie man mit Hilfe von Methoden und Werkzeugen aus der Wirtschaftsinformatik (z.B. der Nutzung von Informationsmodellen) die Informationen für IT-Entscheider in öffentlichen Verwaltungen verbessern kann.

Model-Driven Integration Engineering in der E-Government-Domäne Meldewesen

Stefan Kühne, Maik Thränert
Institut für Informatik, Universität Leipzig
04103 Leipzig
{kuehne,thraenert}@informatik.uni-leipzig.de

Werner Rotzoll, Jan Lehmann
DVZ Datenverarbeitungszentrum Mecklenburg-Vorpommern GmbH
19059 Schwerin
{rotzoll,lehmann}@dvz-mv.de

Abstract

Im Rahmen des E-Government integrieren IT-Dienstleister der öffentlichen Verwaltung bestehende Anwendungen von Bund, Ländern und Kommunen in komplexen Infrastrukturen. Durch domänenspezifische Modellierungssprachen lässt sich die Komplexität der Integrationsvorhaben reduzieren. Am Beispiel der E-Government-Domäne Meldewesen wird in diesem Beitrag der Ansatz Model-Driven Integration Engineering (MDIE) vorgestellt. Er beschreibt u. a. die Entwicklung kompakter, intuitiver Modellierungssprachen für Integrationsvorhaben in einer fachlichen Domäne und fokussiert dabei die direkte Weiterverwendung der modellierten Sachverhalte über Modelloperationen.

1 Integrationsproblematik im E-Government

Derzeit werden in allen Bundesländern große Anstrengungen in der E-Government-Domäne Meldewesen unternommen, eine Infrastruktur aufzubauen und erste Dienste, wie *einfache Melderegisterauskunft* [Osci06, S. 352] oder *Rückmeldung* [Osci06, S. 133] anzubieten. Allgemein betrachtet, entwickeln IT-Dienstleister der öffentlichen Verwaltung E-Government-Dienste, indem sie bestehende Anwendungen von Bund, Ländern und Kommunen in komplexen Infrastrukturen vernetzen und integrieren. Dabei sind sie mit typischen Herausforderungen von In-

tegrationsprojekten konfrontiert, die sich aus der Vielfältigkeit und Heterogenität bestehender Systeme ergeben. Zusätzliche Schwierigkeiten ergeben sich aus Änderungen z. B. hinsichtlich Kundenanforderungen, Gesetzesvorgaben oder verfügbaren Implementierungstechnologien.

Die Komplexität der Integrationsvorhaben zeigt sich bereits in aus fachlicher Sicht vergleichsweise einfachen Prozessen, wie der *einfachen Melderegisterauskunft*. Deren technische Umsetzung basiert auf verschiedenen Infrastruktur- und Basiskomponenten einer E-Government-Plattform, wie Portal, Virtuelle Poststelle oder Verzeichnisdiensten. Die Orchestrierung dieser Komponenten erfolgt unter Nutzung von Workflows in einem Business Process Management System.

Für eine effizientere Umsetzung derartiger Integrationsprojekte im E-Government wird in diesem Beitrag der Ansatz *Model-Driven Integration Engineering (MDIE)* vorgeschlagen, welcher die Anwendung modellgetriebener Techniken im Anwendungsbereich des Integration Engineering (IE) [Raut93, FäTW05, SDTK05] fokussiert. Der MDIE-Ansatz wird am Beispiel der E-Government-Domäne Meldewesen vorgestellt.¹ Es wird gezeigt, wie Konzepte dieses Bereichs identifiziert und analysiert, das gewonnene Domänenwissen in eine kompakte, spezifische Modellierungssprache überführt und Modelle der Sprache über Transformationen weiterverarbeitet werden können. Die Anwendung des MDIE-Ansatzes ermöglicht die (Integrations-)projektübergreifende Wiederverwendung von Domänenwissen sowie Komplexitätsreduktionen, womit Verbesserungen des Entwicklungsprozesses, insbesondere hinsichtlich Flexibilität, Entwicklungsgeschwindigkeit, Entwicklungskosten, Qualität sowie Transparenz für beteiligte Personengruppen (z. B. Auftraggeber), adressiert werden.

2 Der Ansatz des Model-Driven Integration Engineering

2.1 Modellgetriebene Entwicklungsansätze

Modellgetriebene Entwicklungsansätze wie das Model-Driven Engineering (MDE) [Bézi05, FaNg05, Schm06], Model-Driven Software Development (MDSD) [StVö06] oder Generative Software Development (GSD) [Czar04, CzEi00] betonen die zentrale Rolle formaler Modelle im Lebenszyklus eines Softwaresystems bzw. einer Softwaresystemfamilie. Ein Modell dient hierbei als stellvertretende Beschreibung eines zugrunde liegenden Systems (Original), in wel-

¹ Beispielhaft wurden hierfür die Ergebnisse des Projektvorhabens Meldewesen für das Bundesland Mecklenburg-Vorpommern [LeRo06] in die Untersuchungen einbezogen.

chem ausschließlich zweckdienliche Informationen repräsentiert sind. Da ein Modell selbst als System aufzufassen ist, kann vom technischen Informationssystem in mehreren Stufen hinsichtlich verschiedener Sichten und Abstraktionsebenen abstrahiert werden.

Ein zentraler Bestandteil modellgetriebener Ansätze ist die Definition von Modelloperationen, wie Modelltransformationen, Modellvalidierungen oder Modelldifferenzen, auf Basis formaler Modelle zur *Automatisierung* des Entwicklungsprozesses. Modelltransformationen ermöglichen beispielsweise die Anreicherung von Modellen mit zusätzlichen Informationen (z. B. technischen), die informationserhaltende Formüberführung von Modellen oder die Extraktion und Aggregation bestimmter Eigenschaften zur Bewertung von Modellen. Modellvalidierungen überprüfen, ob modellierte Sachverhalte gültig hinsichtlich definierter Anforderungen sind. Die Modelloperationen rechtfertigen somit die zentrale Rolle formaler Modelle im Entwicklungsprozess, da sie direkt zu signifikanten Mehrwerten, wie effizientere Entwicklung oder bessere Qualität, beitragen können.

In den modellgetriebenen Ansätze MDE, MDSD und GSD werden die Konzepte technologieunabhängig definiert. Für den praktischen Einsatz existieren zahlreiche Konkretisierungen, wie beispielsweise die Model Driven Architecture [ObjeoJ], das Model-integrated Computing [KSLB03], das architecture-centric Model-Driven Software Development [StVö06, S. 21 ff.] oder Microsofts Software Factories [GrSh04]. Dabei werden die Konzepte für verschiedene Anwendungsbereiche mit Methoden, Werkzeugen und Standards hinterlegt.

Erfahrungen zeigen, dass die Ansätze in sehr unterschiedlichen Anwendungsbereichen, wie bei der Entwicklung verteilter eingebetteter Systeme oder komplexer Enterprise-Anwendungen, erfolgreich eingesetzt werden und zu Verbesserungen im Entwicklungsprozess führen können [Schm06]. Ein noch wenig betrachteter Anwendungsbereich modellgetriebener Ansätze ist das Integration Engineering.

2.2 Integration Engineering

Beim Integration Engineering (IE) handelt es sich um eine Engineering-Disziplin, „*die Prinzipien, Modelle, Methoden und Werkzeuge entwickelt, bewertet und zur Verfügung stellt, sodass diese in einem ingenieurmäßigen Prozess der Integrierung eingesetzt werden können.*“ [ThKü06, S. 14] Vereinfacht formuliert, liefert das IE „*das »Handwerkszeug« mit zugehörigen Einsatzanweisungen, damit Integrierung durchgeführt werden kann.*“ [Thrä05, S. 21]

Dabei betrachtet das IE Informationssysteme (IS) als relevante Integrationsobjekte, sodass diese zu Integrierten Informationssystemen (IIS) umgestaltet werden. Da es sich bei einem Informati-

onssystem um ein „sozio-technisches System“ [Birk05] handelt, spielen neben den technischen auch die organisatorischen Aspekte einer Integration eine wichtige Rolle, obgleich sich die Ausführungen in diesem Beitrag primär den technischen Integrationsproblemen zuwenden.

Grundsätzlich sind „sowohl die Neuentwicklung von integrierten Informationssystemen („Integration ex ante“) als auch die Zusammenführung bereits existierender Informationssysteme („Integration ex post“)“ [KuRa96, S. 170] Aufgabenbereiche des IE. Zur ex-post-Integration gehören z. B. die Anpassung und Integrierung eines Informationssystems in eine bestehende IT-Infrastruktur, die Integrierung von IS oder deren Komponenten untereinander sowie die Anbindung bestehender IS an eine zentrale Plattform. Zur ex-ante-Integration gehört auch die Desintegration eines IS bzw. IIS mit dem Ziel eines Refactoring, sodass eine bessere Integrierbarkeit dieses IS bzw. IIS erreicht wird. Ebenso zur ex-ante-Integration gehört die Harmonisierung der Anwendungsarchitektur, wie z. B. die Einführung einer einheitlichen technischen Kommunikationsbasis unter Nutzung von z. B. Web Services oder auch die Harmonisierung von Anwendungsmodellen durch Schaffung einer einheitlichen Metamodellinfrastruktur.

Die Anwendungsbereiche des IE sind dabei sehr vielseitig. Dazu gehören beispielsweise der Maschinenbau [WeSo05], E-Government [LeRK06] oder E-Commerce [HäKü06; FFSD06]. Das Integration Engineering versucht dabei ständig, durch neue Methoden, Werkzeuge und Vorgehensweisen eine Effizienzsteigerung zu erzielen. In diesem Sinne wurde auch der im Folgenden dargestellte Ansatz des Model-Driven Integration Engineering entwickelt.

2.3 Model-Driven Integration Engineering

Die Anwendung des Paradigmas modellgetriebener Entwicklung im Anwendungsbereich des IE führt zur Disziplin des *Model-Driven Integration Engineering (MDIE)* [ThKü06]. Dabei werden modellgetriebene Ansätze zur Lösung von IE-Problemen ausgewählt und angepasst. Konkreter formuliert liegt der Fokus des MDIE in

- der Auswahl, Anpassung und ggf. Neuentwicklung
- von Prinzipien, Modellen, Methoden, Werkzeugen und Vorgehensweisen modellgetriebener Ansätze
- zur formalen domänenspezifischen Modellierung und Anwendung von Modelloperationen (wie Modelltransformationen oder Modellvalidierungen)

- für die IE-Aufgabenbereiche Beschreibung der Integrationsform als Ausgangsbasis und Ziel einer Integrierung sowie Überführung von Integrationsformen.

Eine Zielsetzung des MDIE ist die kompakte, problemangepasste und intuitive Modellierung als Basis für ein besseres Verständnis sowie leichtere Handhabbarkeit im Vergleich zum zugrunde liegenden System. Modellierungssprachen mit weit gefasstem Anwendungsbereich (Unified Languages) unterliegen der Problematik, dass die fokussierten Aspekte des Systems nur mit generischen Sprachkonstrukten repräsentiert werden können. Dies führt tendenziell zu feingranularen Beschreibungen. Problemangepasste Modellierungssprachen (Specific Languages) adressieren dagegen einen meist fachlich beschränkten Anwendungsbereich (Domäne). Die spezifischen Gegebenheiten und Charakteristika der Domänen werden berücksichtigt, indem für die Konzepte der Domäne spezifische Sprachkonstrukte zur Verfügung gestellt werden. Dieses Vorgehen ermöglicht eine direkte Repräsentation der Sachverhalte und damit kompakte Modellierung, was zu reduziertem Aufwand bei der Erstellung von Modellen und zu einer verbesserten Kommunikation zwischen den beteiligten Personengruppen führt.

Ein weiteres Ziel des MDIE ist die Definition von Modelloperationen auf Basis formaler domänenspezifischer MDIE-Modellierungen. Entsprechende Transformationen ermöglichen bspw. die Anreicherung von Modellen mit technischen Informationen oder die Extraktion bestimmter Modellaspekte beispielsweise zur Bewertung der Integrationsform. Mithilfe von Validierungen können Modelle bspw. hinsichtlich Konsistenz, Vollständigkeit oder Anforderungskonformität überprüft werden. Die im Rahmen modellgetriebener Ansätze definierten Modelloperationen führen zur Automatisierung von manuellen Arbeitsschritten.

2.4 Anwendung des MDIE im E-Government

Die Potenziale modellgetriebener Entwicklungsansätze übertragen sich auf das MDIE. Insbesondere verspricht dessen Anwendung Verbesserungen hinsichtlich Qualität, Portabilität, Flexibilität und Transparenz der Integrationslösung. Dem gegenüber stehen zusätzliche Aufwände, die sich aus Erstellung und Pflege zusätzlicher MDIE-Artefakte wie domänenspezifische Sprachen, Transformationen und Validierungen ergeben. Inwiefern sich bezüglich des Entwicklungsprozesses langfristige Kostensenkungen und Steigerungen der Entwicklungsgeschwindigkeit ergeben, hängt demnach von den Gegebenheiten des Einsatzgebietes (der Domäne) ab. Hierbei spielen Faktoren wie die Stabilität von Konzepten oder die Anzahl und Umfang der Projekte eine wesentliche Rolle.

Wie in Abschnitt 1 geschildert, stellt die Umsetzung behördenübergreifender E-Government-Prozesse auf Basis existierender Fachverfahren eine aktuelle Herausforderung für IT-Dienstleister der öffentlichen Verwaltung dar. Die effiziente Anwendbarkeit des MDIE-Ansatzes in der Domäne E-Government wird durch folgende Charakteristika begünstigt:

- Der Domäne liegt eine gewachsene und gefestigte Terminologie zugrunde.
- Durch Vorgaben des Gesetzgebers sowie einer Reihe von Standards und Richtlinien von E-Government-Initiativen und Arbeitsgruppen bestehen stabile Konzepte, hinsichtlich des Gegenstandsbereichs der Integrierung und der Integrationsmittel.
- Anzahl und Umfang potenzieller Integrationsprojekte sind beträchtlich. Die einzelnen Integrationsprojekte stehen in einem zeitlichen und logischen Zusammenhang. Demzufolge bestehen Bedarfe hinsichtlich Wiederverwendung von Entwicklungsartefakten, Erfahrungen, usw.

Im Folgenden soll der MDIE-Ansatz anhand der E-Government-Domäne Meldewesen demonstriert werden. Zunächst wird in Abschnitt 3 die Domäne eingegrenzt, die am Entwicklungsprozess beteiligten Personengruppen identifiziert, deren spezifischen Informationsbedarfe in eine Modellierungsarchitektur eingeordnet sowie der für den MDIE-Ansatz relevante Bereich abgegrenzt. In Abschnitt 4 werden die für die ausgewählte Modellierungsebene relevanten Konzepte identifiziert, analysiert und anschließend in eine domänenspezifische Modellierungssprache überführt. Abschnitt 5 zeigt, wie Modelle der entwickelten Sprache auf eine Ausführungsumgebung abgebildet werden können.

3 Die E-Government-Domäne Meldewesen

3.1 Der Nachrichtenstandard OSCI-XMeld

Das Online Services Computer Interface (OSCI) ist ein zweischichtiger Protokollstandard zum sicheren Austausch von Nachrichten über das Internet mit besonderer Eignung für das E-Government [Bund06]. Die sichere und vertrauliche Übertragung digital signierter Dokumente ist in der *Transportschicht* (OSCI-Transport) beschrieben. Hier wird beispielsweise die OSCI-Nachrichtenstruktur definiert, welche sich in die drei Sicherheitsebenen Administrationsebene, Auftragsebene und Inhaltsdaten untergliedert [Osci01]. Die Strukturierung und semantische

Standardisierung von Nachrichteninhalten erfolgt in OSCI in der *Anwendungsschicht*. Hierzu werden in verschiedenen Projekten, wie XMeld, XBau und XJustiz, entsprechende Standards definiert. Für den Bereich Meldewesen ist der Standard XMeld [Osci06] relevant.

Das Meldewesen versorgt eine Vielzahl staatlicher Stellen mit verwaltungsinternen Daten der Einwohner. Rechtliche Grundlage für eine Umsetzung im Sinne des E-Government ist das novellierte Melderechtsrahmengesetz (MRRG), dessen länderspezifische Umsetzung in Landesmeldegesetzen (LMG) und die erste Bundesmeldedatenübermittlungsverordnung (1. BMeld-DÜV). Im Standard XMeld werden Syntax und Semantik von Nachrichten definiert, welche zwischen Meldeämtern und Kommunikationspartnern, wie Bürger, gewerblicher Kunde oder auswärtiges Meldeamt unter bestimmten, festgelegten Umständen ausgetauscht werden. Für die Einordnung der Nachrichten in einen Prozesskontext werden verschiedene Szenarien wie *Anmeldung*, *Rückmeldung* oder *einfache Melderegisterauskunft* definiert.

3.2 Stakeholderanalyse für eine domänenspezifische Modellierung

Für eine domänenspezifische Modellierung nach MDIE sind zunächst relevante Personengruppen und deren Informationsbedarfe zu identifizieren. Zur Umsetzung von E-Government-Diensten gehören folgende Gruppen:

Auftraggeber (Land, Kommune, Meldebehörde): Der Auftraggeber hat einen Informationsbedarf hinsichtlich der Umsetzung von gesetzlichen Grundlagen (Bundesgesetze, Landesgesetze und Richtlinien von Arbeitsgruppen), z. B. in welcher Form behördenübergreifende Prozesse ablaufen. Im Fall des Meldewesens, welches auf Landesebene umgesetzt wird, sind die fachlichen Anforderungen an Komponenten der Meldebehörden relevant.

Entwickler und Systemintegrator: Relevant sind hier funktionale und technische Aspekte (z. B. IT-Prozess, Prozessschnittstellen und fachliche Anforderungen an Komponenten), welche für die Entwicklung neuer bzw. die Integration vorhandener Komponenten erforderlich sind.

Systembetreuer: Der Systembetreuer verwaltet die Systeme im laufenden Betrieb. Er muss Fehler im Wirkbetrieb Prozessen und Komponenten zuordnen können. Der Informationsbedarf ist demnach ähnlich dem der Gruppe Entwickler und Systemintegrator, jedoch aus einer anderen Perspektive.

Service-Manager: Der Service-Manager steht zwischen den Gruppen Auftraggeber, Entwickler/Systemintegrator und Systembetreuer. Er ist für die fachgerechte Umsetzung von Diensten verantwortlich. Demzufolge sind organisatorische Aspekte wie Service-Level-Agreements,

Konfiguration/Change-Management, aber auch funktionale Aspekte wie Prozessablauf und Prozessschnittstellen relevant.

Weitere Gruppen: Weitere Gruppen sind u. a. Marketing, Datenschützer, User Help Desk, Medien, Bürger, Nutzer der Anwendung, Diensteanbieter und Zertifizierungslabor.

Die unterschiedlichen Zielstellungen dieser Personengruppen erfordern eine Modellierung auf verschiedenen Abstraktionsebenen (vgl. Abb. 1). Das Spektrum reicht dabei von fachlich-abstrakten Modellierungsebenen, welche für Auftraggeber und Service-Management relevant sind, bis zu technisch-konkreten Modellierungsebenen für die Gruppen Entwicklung und Systemintegration. Dabei ist die dynamische Prozesssicht, in der statische Sichten, wie Organisations-, Ressourcen- oder Datensicht, miteinander verknüpft sind, für alle beteiligte Personengruppen von hoher Priorität. Auf dieser Basis können bei der Entwicklung von E-Government-Diensten bspw. die Anforderungen an das Integrationssystem oder der systemübergreifende Entwurf spezifiziert werden. Auch im Betrieb ist die Prozesssicht relevant, z. B. bei der Zuordnung von Fehlern zu Systemkomponenten. Neben den gruppenbezogenen Informationsbedarfen ist die Prozesssicht ebenso für die gruppenübergreifende Abstimmung relevant.

In Abb. 1 ist eine mögliche Modellierungsarchitektur dargestellt. Diese setzt auf einer vorhandenen E-Government-Plattform, bestehend aus verschiedenen Infrastruktur- und Basiskomponenten, auf. Die technische Integration der Komponenten wird auf Integrationsebene beschrieben. Hier können z. B. auf Basis von Web-Service-Technologien die Komponenten systemübergreifend orchestriert werden. Die darüber liegenden Modellierungsebenen, wie die technische und fachliche Prozessmodellierungsebene, beschreiben Prozesse mit abnehmend technischem Detaillierungsgrad bzw. zunehmend fachlichem Abstraktionsgrad.

Die Modellierungsebene Technische Prozessmodellierung ist primär für die Personengruppe Systemintegration relevant und liegt daher im Fokus des MDIE-Ansatzes. Sie dient jedoch auch

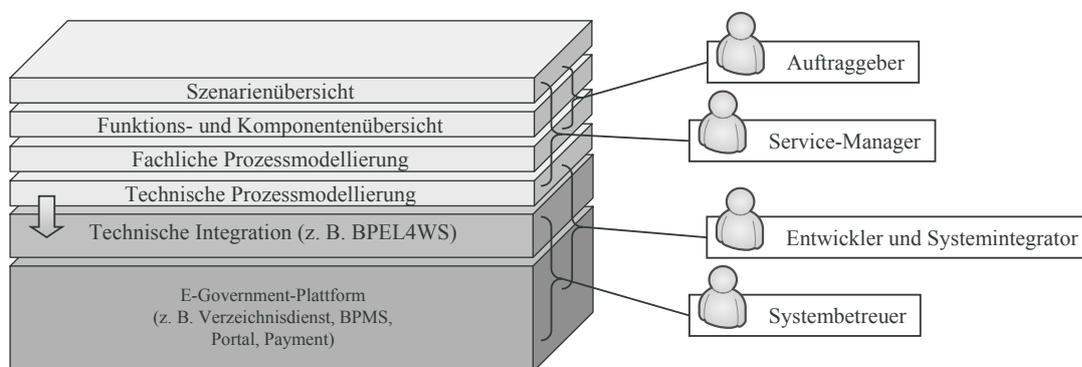


Abb. 1: Modellierungsebenen

zur Abstimmung mit den Gruppen Entwicklung und Service-Management und bildet die Schnittstelle zwischen fachlicher Modellierung und technischer Umsetzung. Sie abstrahiert von technischen Details zugrunde liegender Integrationsprachen und enthält Aspekte der fachlichen Domäne (E-Government, Meldewesen). Zielstellung des MDIE-Ansatzes ist die automatisierte Abbildung von Modellen der technischen Prozessmodellierungsebene auf die technische Integrationsebene (vgl. Pfeil in Abb. 1). Dies wird realisiert über Modelltransformationen, welche die fachlichen Modellierungskonstrukte mit den abstrahierten technischen Aspekten anreichern.

4 Entwicklung einer domänenspezifischen Sprache für die technische Prozessmodellierung im Bereich Meldewesen

4.1 Identifizierung und Analyse relevanter Konzepte

Nach Einschränkung der Domäne und Identifikation relevanter Personengruppen besteht der nächste Schritt in der Identifikation und Analyse relevanter Domänenkonzepte. Für den auf die technische Prozessmodellierungsebene eingeschränkten Bereich wurden hierfür Spezifikationsdokumente aus der Domäne E-Government bzw. der Subdomäne Meldewesen (z. B. SAGA-Architektur [Kbst05], XMeld-Spezifikation [Osci06]) sowie Implementierungsartefakte einer exemplarischen Umsetzung des Szenarios *einfache Melderegisterauskunft* (wie Orchestrierungen und Teilorchestrierungen von am Prozess beteiligten Komponenten auf Basis der Business Process Execution Language for Web Services (BPEL4WS), Schnittstellenbeschreibungen von Diensten auf Basis der Web Service Description Language (WSDL) und Nachrichtenformate auf Basis von XML-Schema-Beschreibungen) ausgewertet. Die Teilergebnisse der Konzeptanalyse wurden sodann in einem iterativen/inkrementellen Prozess mit Domänenexperten aus den Gruppen Systemintegrator, Entwickler und Service-Manager diskutiert und abgestimmt.

Als allgemeine Basiskonzepte für die technische Prozessmodellierung wurden für Prozessbeschreibungssprachen typische Konstrukte wie *Prozess, Aktivität, Nachricht, Attribut, Ereignis, Prozessstart, Prozessende, System* und *Systemschnittstelle* identifiziert. Zwischen diesen allgemeinen Konzepten bestehen vielfältige Beziehungen. So besteht ein Prozess z. B. aus Aktivitäten, eine Nachricht ist Input für eine Aktivität oder ein Prozess terminiert mit dem Prozessende. Diese Konzepte sind elementar und treten mehrfach in verschiedenen Ausprägungen auf.

Auf diesen Basiskonzepten bauen fachdomänenspezifische Konzepte auf, wie *XMeldNachricht*, *VerzeichnisdienstNachricht*, *XMeldAttribut*, *Fehlernachricht*, *Protokollierung*, *Fehlerbehandlung*, *Nachrichtenkonvertierung*, *Verzeichnisdienst*, *Intermediär*. Diese Konzepte haben noch stark technischen Charakter, erweitern die zugrunde liegenden Basiskonzepte jedoch mit domänenspezifischer Semantik. So ist beispielsweise das Konzept *XMeldNachricht* eine Einschränkung des Konzepts *Nachricht* auf die im XMeld-Standard definierten Nachrichtentypen (dargestellt als Vererbungsbeziehungen in Abb. 2).

Neben den Basiskonzepten und domänenspezifischen Konzepten gibt es weiterhin zusammengesetzte Konzepte, welche sich aus Komposition bestehender Konzepte zusammensetzen. Diese repräsentieren Muster, die in zugrunde liegenden Implementierungsartefakten gehäuft und in unterschiedlichen Ausprägungen auftreten. Beispiele hierfür sind *Daten empfangen*, *Daten senden*, *E-Mail versenden*, *Kommunikation mit Verzeichnisdienst*, *Kommunikation mit Web Service* oder *XMeldNachricht bearbeiten*. Das Konzept *Kommunikation mit Verzeichnisdienst* bspw. ist abgeleitet vom Konzept *Aktivität*. Es umfasst neben dem Aufruf einer Verzeichnisdienstoperation durch Senden und Empfangen von Nachrichten an einen Verzeichnisdienst optional auch eine Fehlerbehandlung, in der die Operation nach vorgegebenen Zeitintervallen normal wiederholt wird. Anschließend kann eine Fehlerprotokollierung erfolgen.

In Abb. 3 sind zwei mögliche Ausprägungen des Konzepts *Kommunikation mit Verzeichnisdienst* (auf technischer Integrationsebene) dargestellt. Die Menge der möglichen Ausprägungen eines Konzepts wird durch dessen Variabilität bestimmt. Eine Analyse der Variabilitäten der Konzepte einer Domäne ist Grundlage für die spätere Entwicklung der domänenspezifischen Sprache, da durch sie die Beziehungen der Konzepte, z. B. bestimmte Wechselwirkungen, expliziert werden. Die Erfassung der Variabilitäten kann mithilfe der in [AnCz04] beschriebenen

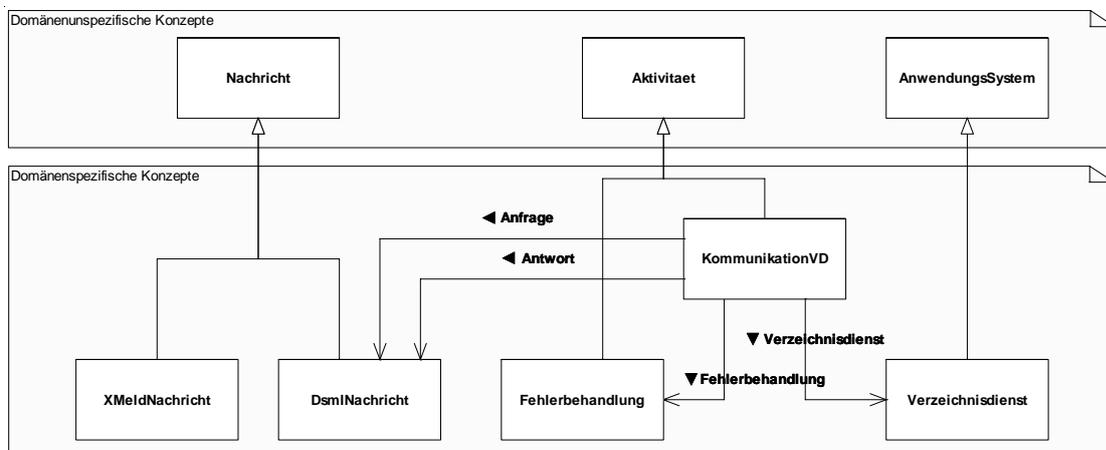


Abb. 2: Konzepte der E-Government-Domäne Meldewesen (Ausschnitt)

Merkmalanalyse durchgeführt werden. Hierfür wird für die Konzepte einer Domäne ein Merkmalmodell erstellt, welches für jedes Konzept ein Merkmaldiagramm enthält. In Abb. 3 (Mitte) ist das zugehörige Merkmaldiagramm für das Konzept *Kommunikation mit Verzeichnisdienst* angegeben. Es enthält das Merkmal Nachrichtenkommunikation, welche aus den Konzepten *Nachrichten senden* und *Nachrichten empfangen* besteht. Abhängig vom Prozesskontext wird dabei eine der aufgeführten Operationen ausgeführt. Ist die optionale Fehlerbehandlung in einer Konfiguration ausgeprägt, so sind hierfür die Anzahl der Anfrageversuche sowie die Wartezeit zwischen den Wiederholungen zu spezifizieren.

Die zusammengesetzten Konzepte sind speziell. Durch die Komposition entstehen Abhängigkeitsbeziehungen von Basiskonzepten, was zur Folge hat, dass zusammengesetzte Konzepte weniger stabil hinsichtlich Änderungen in der Domäne sind. Des Weiteren entsteht Redundanz im Konzeptraum, was zu verringerter Übersichtlichkeit führt. Dennoch wird im MDIE-Ansatz

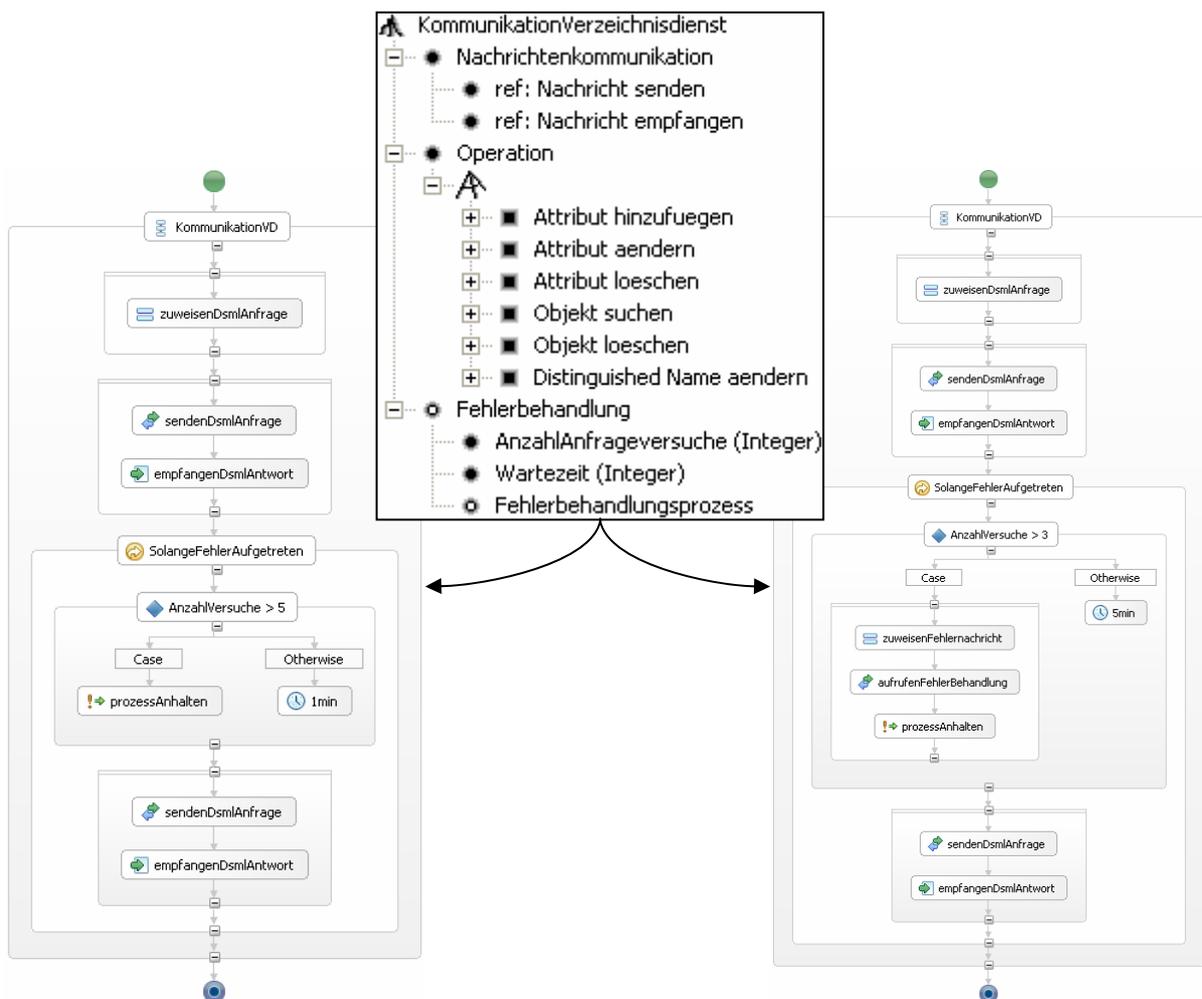


Abb. 3: Merkmaldiagramm des Konzepts *Kommunikation mit Verzeichnisdienst* (erstellt mit dem Eclipse-Plugin fmp [AnCz04]) und zwei verschiedene Ausprägungen (links und rechts)

die Einführung von zusammengesetzten Konzepten bei häufig auftretenden Kombinationen als sinnvoll erachtet. Sie sind die Grundlage für eine kompakte und damit übersichtliche bzw. für Fachexperten intuitive Modellierung.

4.2 Abbildung der Konzepte auf eine domänenspezifische Sprache

Nach der Exploration des Konzeptraums besteht der nächste Schritt in der Definition der domänenspezifische Sprache, welche im Folgenden als EGov_{TP} bezeichnet wird, sowie deren Implementierung in Form eines Editors.

Domänenspezifische Sprachen können in unterschiedlicher Form abgebildet werden. Czarnecki klassifiziert bspw. die Ansätze hinsichtlich Flexibilität und unterscheidet zwischen dem Entscheidungsbaum-basierten Konfigurationsansatz, dem Merkmal-basierten Konfigurationsansatz und der Graph-basierten Modellierungssprache [Czar04].

Für die betrachtete Subdomäne technische Prozessmodellierung im Meldewesen ist die Flexibilität des letztgenannten Ansatzes erforderlich. Für die Definition einer graphischen domänenspezifischen Modellierungssprache existieren leicht- und schwergewichtige Ansätze. Die Grundlage in *leichtgewichtigen Ansätzen* bildet eine universelle Modellierungssprache, welche durch entsprechende Mechanismen erweitert bzw. eingeschränkt werden kann. Durch Vorgabe von Annotationsmöglichkeiten können domänenspezifischen Konzepten in die Sprache eingebracht werden. Ein Beispiel für dieses Vorgehen ist der UML-Profil-Mechanismus, welcher die Definition von Stereotypen und Constraints für Modellierungselemente erlaubt.

Bei *schwergewichtigen Ansätzen* werden grundlegend neue Modellierungssprachen mit Hilfe von Metasprachen definiert. Dieses Vorgehen ist aufwendiger, da hierbei Grundkonzepte nachmodelliert werden müssen, die beim leichtgewichtigen Ansatz bereits durch die Basissprache gegeben sind. Schwergewichtige Ansätze sind jedoch flexibler, da Modellierungselemente über Konstrukte der Metamodellierungssprache (z. B. Vererbungsbeziehungen, Aggregationsbeziehungen) präziser und vielschichtiger definiert werden können. Ein weiterer Vorteil schwergewichtiger Ansätze ist, dass bei der Definition der konkreten Syntax der Modellierungssprache größerer Freiraum besteht. So können beispielsweise für Fachexperten intuitive graphische Symbole definiert werden. Zu den schwergewichtigen Ansätzen gehören bspw. Meta Object Facility 2.0 (MOF 2.0) [Obj06], das Eclipse Modeling Framework (EMF) bzw. das darauf aufbauende Eclipse Graphical Modeling Framework (GMF), MetaGME [BGKS06], MetaEdit+ von MetaCase oder Microsoft DSL Tools. Alle Metamodellierungswerkzeuge ermöglichen in ähnlicher Weise das Editieren domänenspezifischen Sprachen, das Erzeugen zugehöriger do-

mänenspezifischer Modellierungswerkzeuge, sowie deren Anwendung bzw. Validierung [AtKü05].²

Im Fall der Sprache EGov_{TP} fiel die Entscheidung auf den schwergewichtigen Ansatz, da er die direkte Abbildung der Domänenkonzepte sowie deren vielfältigen Beziehungen auf Sprachkonstrukte ermöglicht. Die Implementierung der EGov_{TP} erfolgte auf Basis der Eclipse-Werkzeugplattform (insbes. Eclipse EMF und GMF). Die dem EMF-Framework zugrunde liegende Metasprache Ecore [BSME03] ermöglicht die direkte Repräsentation der identifizierten Domänenkonzepte (vgl. Abschnitt 4.1) in einem entsprechenden Metamodell, daher kann Abb. 2 als Ausschnitt aus dem Metamodell gesehen werden. Basierend auf dem Ecore-Metamodell können verschiedene graphische Editoren definiert werden. Hierzu wird für jedes Metamodellelement spezifiziert, in welcher (graphischen) Form entsprechende Instanzen darzustellen sind. Aus diesen Editor-Spezifikationen können mit GMF als Eclipse-Plugin lauffähige Editoren (vgl. Abb. 4) generiert werden. Das in Abb. 4 dargestellte Prozessdiagramm zeigt einen Ausschnitt aus der Beispielmmodellierung *einfache Melderegisterauskunft*. Das Konzept *Kommunikation mit Verzeichnisdienst* tritt dabei in den Instanzen *MeldebehoerdeAngeschlossen* und *HoleSchnittstellenInformationen* auf. Ein Vergleich der Instanzen mit Ausprägungen desselben Konzepts auf technischer Integrationsebene (vgl. Abb. 3) zeigt, dass sich auf technischer Prozessmodellierungsebene eine deutlich kompaktere und damit übersichtlichere Modellierung erreichen lässt.

4.3 Weiterverwendung von domänenspezifischen Modellen durch Modelltransformationen

Die technische Prozessmodellierung anhand der domänenspezifischen Sprache erlaubt eine kompakte und für Fachexperten verständliche Modellierung der umzusetzenden Integrationsprozesse. Prozessmodelle bilden anschließend die Grundlage für automatisierte Modelloperationen. Die grundlegendste Operation ist die Modelltransformation, da sie die Wiederverwendung der im Modell repräsentierten Informationen in anderen Kontexten ermöglicht. In [Viss01] werden Transformationen hinsichtlich des Abstraktionsgrad der Quell- und Zielsprache in vertikale und horizontale Transformationen untergliedert. Für den betrachteten Anwendungsfall ist die Abbildung der technischen Prozessmodellierung auf Sprachen der Integrationsebene (z. B. BPEL4WS, WSDL) relevant. Es handelt sich dabei um vertikale Modelltransformationen (Refinement) einer High-Level-Modellierungssprache auf Low-Level-Sprachen.

² Atkinson und Kühne verweisen auf prinzipielle Probleme, z. B. Inkompatibilitäten zwischen Modellen und veränderter domänenspezifischer Modellierungssprache. Dennoch ist dieser Ansatz derzeit stark verbreitet.

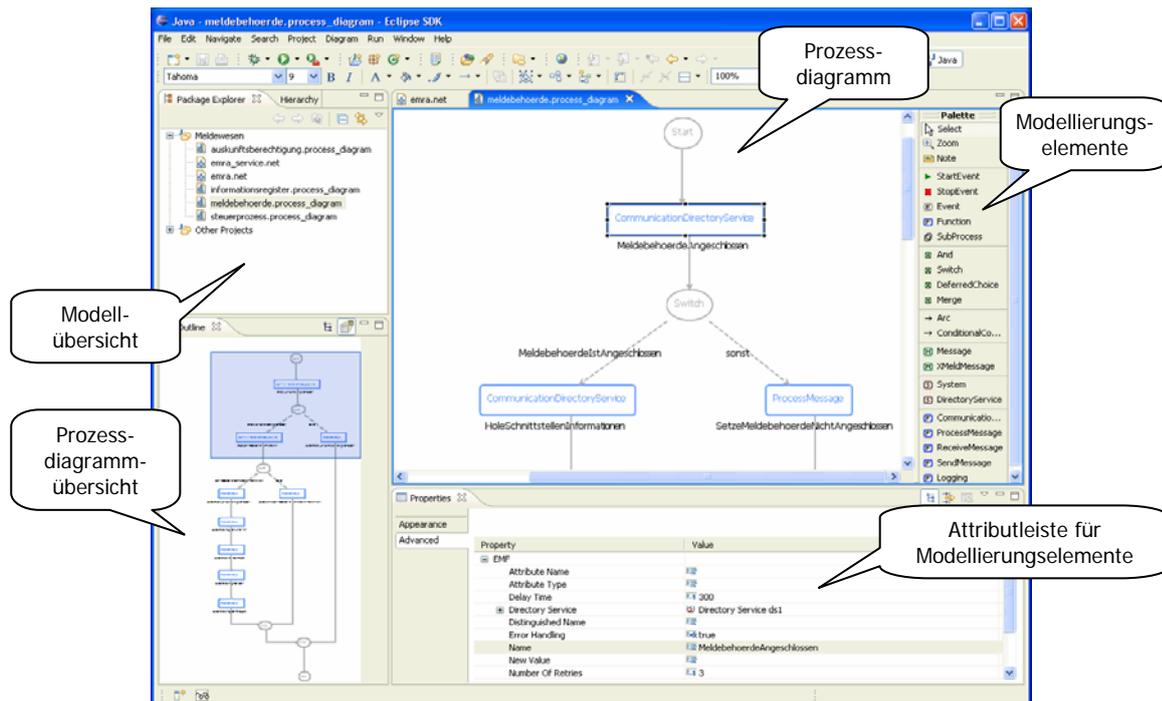


Abb. 4: Graphisches Modellierungswerkzeug für die technische Prozessmodellierungssprache EGov_{TP}

Die Ergebnisse der Domänenanalyse sind die Grundlage für die Definition von Transformationsregeln, welche die Abbildungsvorschriften von EGov_{TP}-Modellen auf Sprachen der Integrationsebene definieren. Im MDIE-Ansatz fließen die Konzepte der Domänenanalyse direkt in die domänenspezifische Modellierungssprache ein. Die Merkmaldiagramme explizieren die den Konzepten zugrunde liegenden Variabilitäten. Den Ausprägungen dieser Merkmaldiagramme (Konfigurationen) werden konkrete Ausprägungen wie in Abb. 3 zugeordnet (vgl. hierzu [CzAn05]). Die Beziehungen der technischen Prozessmodellierungen (z. B. Kontrollflusskanten) geben an, wie diese Konzeptausprägungen in den Sprachen der Integrationsebene zu verknüpfen sind. In [CzHe06, MeGo05] werden verfügbare Transformationstechniken und -werkzeuge für eine konkrete Umsetzung dieser Transformationsstrategie klassifiziert.

5 Zusammenfassung

In diesem Artikel wurde der MDIE-Ansatz vorgestellt, welcher eine Anwendung modellgetriebener Techniken auf den Anwendungsbereich des Integration Engineering darstellt. Der Schwerpunkt lag dabei auf der Integration domänenspezifischer Konzepte in Prozessmodellierungssprachen zur Reduktion der Modellierungskomplexität. Die Anwendung des MDIE-

Ansatzes in der E-Government-Domäne Meldewesen zeigte exemplarisch, dass die angewendeten Abstraktionsmechanismen die Komplexität der zugrunde liegenden Integrations-sprachen reduzieren können. Es wurde eine Transformationsstrategie skizziert, welche die direkte Weiterverwendung repräsentierter Sachverhalte ermöglicht. Andere Modelltransformationen wie die Extraktion bestimmter Modellaspekte (Systeme, Schnittstellen, Nachrichtenformate) zur Analyse bzw. andere Modelloperationen wie Modellvalidierung oder Modelldifferenzen sind gleichfalls denkbar.

Für eine umfassende Bewertung des MDIE-Ansatzes sind jedoch weiterführende Arbeiten erforderlich. So sind für quantitative Kosten-/Nutzenbewertungen empirische Studien durchzuführen. Des Weiteren sind Vorgehensaspekte der MDIE-Artefakte, z. B. der Weiterentwicklung domänenspezifischer Sprachen im betrachteten Bereich sowie der Abstimmung zwischen verschiedenen domänenspezifischen Sprachen (DSL-Management) zu bestimmen.

Literaturverzeichnis

- [AnCz04] Antkiewicz, M.; Czarnecki, K.: FeaturePlugin: Feature Modeling Plug-In for Eclipse. In: OOPSLA'04 Eclipse Technology eXchange (ETX) Workshop, <http://www.swen.uwaterloo.ca/~kczarnec/etx04.pdf>, 2004.
- [AtKü05] Atkinson, C.; Kühne, T.: Concepts for Comparing Modeling Tool Architectures. In: ACM/IEEE 8th International Conference on Model Driven Engineering Languages and Systems, MoDELS / UML 2005, <http://www.mm.informatik.tu-darmstadt.de/staff/kuehne/publications/papers/concepts.pdf>, 2005.
- [Bézi05] Bézivin, J.: On the Unification Power of Models. In: Software and System Modeling (SoSym) 4 (2005) 2, S. 171–188.
- [BGKS06] Balasubramanian, K.; Gokhale, A.; Karsai, G.; Sztipanovits, J.; Neema, S.: Developing Applications Using Model-Driven Design Environments. In: IEEE Computer, Februar 2006, 2006, S. 33–40.
- [Birk05] Birker, K.: Das neue Lexikon der BWL. Cornelsen, 2005.

- [BSME03] Budinsky, F.; Steinberg, D.; Merks, E.; Ellersick, R.; Grose, T. J.: eclipse Modeling Framework. Addison-Wesley (the eclipse series), 2003.
- [Bund06] Bundesamt für Sicherheit in der Informationstechnik (BSI), Projektgruppe EGovernment (Hrsg.): Das E-Government-Glossar: Pragmatische Definitionen, Begriffserläuterungen und Abkürzungsverzeichnis. http://www.bsi.de/fachthem/egov/download/6_EGloss.pdf, 2006-01.04, Abruf am 2006-11-15.
- [CzAn05] Czarnecki, K.; Antkiewicz, M.: Mapping Features to Models: A Template Approach Based on Superimposed Variants. In: Generative Programming and Component Engineering (GPCE 2005), 2005, S. 422–437.
- [Czar04] Czarnecki, K.: Overview of Generative Software Development. In: Banâtre, J.-P.: Unconventional Programming Paradigms (UPP) 2004. Springer, LNCS 3566, 2004, S. 313–328.
- [CzEi00] Czarnecki, K.; Eisenecker, U. W.: Generative Programming: Methods, Tools, and Applications. Addison-Wesley, 2000.
- [CzHe06] Czarnecki, K.; Helsen, S.: Feature-based survey of model transformation approaches. In: IBM Systems Journal 45 (2006) 3, S. 621–645.
- [FaNg05] Favre, J.-M.; NGuyen, T.: Towards a Megamodel to Model Software Evolution Through Transformations. In: Electronic Notes in Theoretical Computer Science 127 (2005) 3, S. 59–74.
- [FäTW05] Fähnrich, K.-P.; Thränert, M.; Wetzel, P. (Hrsg.): Umsetzung von kooperativen Geschäftsprozessen auf eine internetbasierte IT-Struktur: Arbeiten aus dem Forschungsvorhaben Integration Engineering. Leipziger Beiträge zur Informatik. Band III, Eigenverlag Leipziger Informatik-Verbund (LIV), 2005.
- [FFSD06] Fötsch, D.; Feja, S.; Sauer, S.; David, A.: Problemstellungen agiler Schnittstellen am Beispiel des Commerce Management Systems von Truition/AGETO. In: [FKSW06], S. 49–56.

- [FKSW06] Fähnrich, K.-P.; Kühne, S.; Speck, A.; Wagner, J.: Integration betrieblicher Informationssysteme, Problemanalysen und Lösungsansätze des Model-Driven Integration Engineering. Eigenverlag Leipziger Informatik-Verbund (LIV), Leipzig, 2006.
- [GrSh04] Greenfield, J.; Short, K.: Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools. Wiley Publishing, 2004.
- [HäKü06] Hänsgen, P.; Kühne, S.: Modellgetriebene Softwareentwicklung zur Lösung von Integrations- und Migrationsproblemen am Beispiel des E-Commerce-Systems Intershop Enfinity. In: [FKSW06], S. 31–41.
- [Kbst05] KBSt: SAGA Version 2.1: Standards und Architekturen für E-Government-Anwendungen Koordinierungs- und Beratungsstelle der Bundesregierung für Informationstechnik in der Bundesverwaltung (KBSt), <http://www.kbst.bund.de>, 2005, Abruf am 2006-11-15.
- [KSLB03] Karsai, G.; Sztipanovits, J.; Ledeczi, A.; Bapty, T.: Model-integrated development of embedded software. In: Proceedings of the IEEE. 91 (2003) 1, S. 145–164.
- [KuRa96] Kurbel, K.; Rautenstrauch, C.: Integration Engineering: Konkurrenz oder Komplement zum Information Engineering? – Methodische Ansätze zur Integration von Informationssystemen. In: Heilmann, H. (Hrsg.): Information Engineering: Wirtschaftsinformatik im Schnittpunkt von Wirtschafts-, Sozial- und Ingenieurwissenschaften. Oldenbourg, 1996, S. 167–191.
- [LeRK06] Lehmann, J.; Rotzoll, W.; Kühne, S.: Analyse der E-Government-Domäne Meldewesen am Beispiel des Dienstes „einfache Melderegisterauskunft“. In: [FKSW06], S. 20–30.
- [LeRo06] Lehmann, J.; Rotzoll, W.: E-Government-Dienste im Meldewesen Mecklenburg-Vorpommern. Interne Entwicklerdokumentationen, 2006.
- [MeGo05] Mens, T.; Gorp, P. v.: A Taxonomy of Model Transformation. In: International Workshop on Graph and Model Transformation (GraMoT) 2005. 2005, S. 7–23.

- [Obj06] Object Management Group: Meta Object Facility (MOF) Core Specification: Version 2.0, Object Management Group, <http://www.omg.org/cgi-bin/doc?formal/2006-01-01>, 2006-01-01, Abruf am 2006-11-15.
- [Obj0J] Object Management Group: OMG Model Driven Architecture. <http://www.omg.org/mda>, Abruf am 2006-11-15.
- [Osci01] OSCI Leitstelle (Hrsg.): OSCI: Die informelle Beschreibung – Eine Ergänzung zur OSCI Spezifikation. Technischer Bericht, <http://www.osci.de/materialien/summary.pdf>, 2001, Abruf am 2006-11-15.
- [Osci06] Projektgruppe OSCI-XMeld (Hrsg.): OSCI-XMeld 1.3.0. Spezifikation. <http://www.osci.de/xmeld130/xmeld-130.zip>, 2006, Abruf am 2006-11-15.
- [Raut93] Rautenstrauch, C.: Integration Engineering: Konzeption, Entwicklung und Einsatz integrierter Softwaresysteme. Addison-Wesley, 1993.
- [Schm06] Schmidt, D. C.: Model-Driven Engineering. In: IEEE Computer 39 (2006) 2, S. 25–31.
- [SDTK05] Specht, T.; Drawehn, J.; Thränert, M.; Kühne, S.: Modeling Cooperative Business Processes and Transformation to a Service Oriented Architecture. In: 7th International IEEE Conference on E-Commerce Technology, 2005, S. 249–256.
- [StVö06] Stahl, T.; Völter, M.: Model-Driven Software Development: Technology, Engineering, Management. Wiley Publishing, 2006.
- [ThKü06] Thränert, M.; Kühne, S.: Model-Driven Integration Engineering und seine Anwendung im Projekt OrViA. In: [FKSW06], S. 13–19.
- [Thrä05] Thränert, M.: Integration – Eine Begriffsbestimmung. In: [FäTW05], S. 11–22.
- [Viss01] Visser, E.: A Survey of Rewriting Strategies in Program Transformation Systems. In: Electronic Notes in Theoretical Computer Science 57 (2001) 2.
- [WeSo05] Wetzel, P.; Soutchilin, A.: Integration Engineering im Maschinen- und Anlagenbau. In: [FäTW05], S. 23–33.

Support of Harmonisation of Public Processes

Through Modelling of Legal Constraints

Sebastian Olbrich

Institut für Wirtschaftsinformatik
Philipps-Universität Marburg,
35037 Marburg
sebastian.olbrich@zeon.de

Paul Alpar

Institut für Wirtschaftsinformatik
Philipps-Universität Marburg,
35037 Marburg
alpar@staff.uni-marburg.de

Abstract

E-Government can transform and improve the entire scope of administrative actions and political processes. Hence, E-Government is both, vision of a future government and the reality we can experience today. E-government is not an objective per se; it has to be seen more as means in organising public governance for better serving citizens and enterprises. The terms on how citizens and enterprises are served is articulated in the public law. Thus, when analysing public processes, the legal framework must be part of the consideration. We use a semi-formal business process modelling methodology to highlight the legal framework that governs different implementations of an E-Government process and use these models as a basis to create a mathematical model. Such models can be used for simulations, to merge processes or as direct input for a workflow management system. We use the formal models to suggest a harmonised process of car registration in the European Union.

1 Introduction

The fast development of ICT and the benefits realised in the private sector from its use put pressure on the public authorities. E-Government seems to be the perfect answer to the demand for better and more efficient services in times of short public budgets. But the ongoing debates in research and in public show that introducing E-Government systems is not that easy. Many of the technologies and techniques that are successfully used in the private sector cannot be transferred one-to-one to the public institutions [WoKr05]. According to a recent survey of top executives in German public institutions the major barriers are the lack of internal know-how, legal restrictions and the missing documentation of public processes [SKH03]. To show the impact of legal constraints, this paper focuses on an area where E-Government is closest to E-Commerce, the service provision.

Online service provision means that external service structures are adequately mapped to the internal process structures of public authorities [BAD03]. Therefore, the addressee's perspectives have to be complemented by a restructuring of the business processes. Process design has to break new ground by taking into account several aspects such as different locations of service production and delivery, organisational front office / back office connection, combining processes according to life situations or including distinct processes from strict workflows to collaborative decision-making. However, we can observe a trade-off in the ongoing process re-design: on the one hand the local structures need to be strengthened in order to efficiently respond to the individual addressee's needs; on the other hand there is a strong effort to harmonise the processes and centralise the data storage – especially if supra-national players such as the European Union come in play.

Moving ahead implies having an integrated view, clear strategies and concepts that are both innovative and feasible. Often, that strategy seems to be missing. In order to articulate new strategies, the concept of Business Process Modelling (BPM) has been widely accepted - in research and in practice. However, there has been little discussion about the organisational consequences of applying this approach.

Process reorganisation in the public sector may often have to stop short of established structures; but finally they will lead to rethinking the institutional structures of Government [SnZu97]. In many respects the legal framework where the public processes and the institutions itself are defined does not fit the current needs anymore. While at a first look this seems to be a problem of public institutions only, we can observe a comparable trend in the private sector, too [KBK06]. Moreover a design has to consider the very different ways of administrative processes. For each of them, IT support will rather be different. Recurrent and well-structured processes are however the base of such an effort [WiTr03].

As an example, we begin our paper with a description of the registration procedure of motor vehicles in Germany and emphasise the process related regulations formulated in the official registration documents. Afterwards, we translate the regulation document step-by-step into a process notation which allows synthesising the entire reference process model from these fragments. We demonstrate how to implement a concrete workflow on the basis of this framework and make it comparable to different implementations of the same task in some of the other EU countries. We close our paper with an outlook on further applications of the presented approach.

2 Regulations of the registration of motor vehicles

2.1 Process description

More than five million cars are registered in Germany every year, which makes the registration process one of the most executed public processes in the German administration. The legal framework for vehicles traffic in Germany is set in federal law code, the so called *Straßenverkehrsgesetz* (StVG) [StVG03]. Additional regulations on the registration of motor vehicles are specified in the *Fahrzeugregisterverordnung* (FRV) [FRV87]. The information collected here is also the base for the taxation of vehicles in the corresponding tax law (*Kraftfahrzeugsteuergesetz*, KrafStG [KraftStG02]). Although specified in federal law, the registration process is executed on the local level (this can be observed on the German licence plates which carry a local identifier). The following description of the registration process also discusses the ongoing debate about a possible reengineering of the process.

In order to register a car in Germany, the car holder must be a German citizen or resident. Consequently, a validation by the authorities in compliance with §33 II StVG and §1 FRV is an important activity within the registration process. According to §33 I StVG and §2 FRV, the vehicle to be registered must be approved by the technical surveillance service (*Technischer Überwachungsverein*, TÜV). Finally, it is obligatory that every motor vehicle must be insured (§34 I and II StVG). Hence, the cover by an insurance company must be presented.

In some cases, the officials may also ask for a proof of financial means from an applicant in order to guarantee that the car tax will be paid. This can be proven by a bank statement. The tax amount is determined by the technical configuration of a motor vehicle (engine, type of fuel, amount of pollutant production, etc.) in accordance with the motor vehicle tax law (KraftStG). Since the tax is reduced for handicapped people (§3 a KraftStG) or farmers (§3 Nr. 7 KraftStG), the belonging to one of this groups must also be documented.

To sum up: in order to register a motor vehicle in Germany one should present his or her personal ID, the approval by the technical surveillance for re-registration or the papers of the vehicle (European Union Licensing Part II) for a new/first registration, the cover of the insurance and – if needed – the proofs of liquidity and any special circumstances.

2.2 Using BPM methods for modelling of the process

The activities of a business process are modelled in extended Event-driven Process Chains (eEPC) as functions (depicted as rectangles with rounded corners). They are triggered by events (depicted as diamonds) and produce events which indicate that the execution of a function is finished. Beginning with a start event, functions and events are mutually connected by control flow arcs. Additional connectors split and join alternatives and concurrency (using the connector types AND, OR, XOR). Moreover, organisational units and information objects are represented by additional symbols. They are connected with function symbols describing the use of these resources when the corresponding function is executed. Figure 1 illustrates the process described above in an eEPC model [Sche94]. Due to their implementation in the SAP R/3 Analyser and the ARIS Toolset, eEPC diagrams are widely spread – especially in Germany. Hence, the method of eEPC is also predominant in (German) E-Government programs

[VBPO05]. Consequently, we start with an eEPC model that later becomes an input for a more formal model.

Paragraphs and legal regulations are the information objects which control the execution of workflows of public authorities. Alpar and Olbrich, therefore, propose the use of the information object metaphor to represent these legal constructs in models of public processes as depicted in the left graph of figure 1 [AIO105]. Thus, not only the workflow itself is visualised, but also the influence of legal constraints on the workflow. The other way round, laws and legal regulations can be seen as specifications for the workflow as they represent the processes' constraints. Within Figure 1, optional regulations (checking the liquidity and tax reduction) are highlighted with the aid of notes.

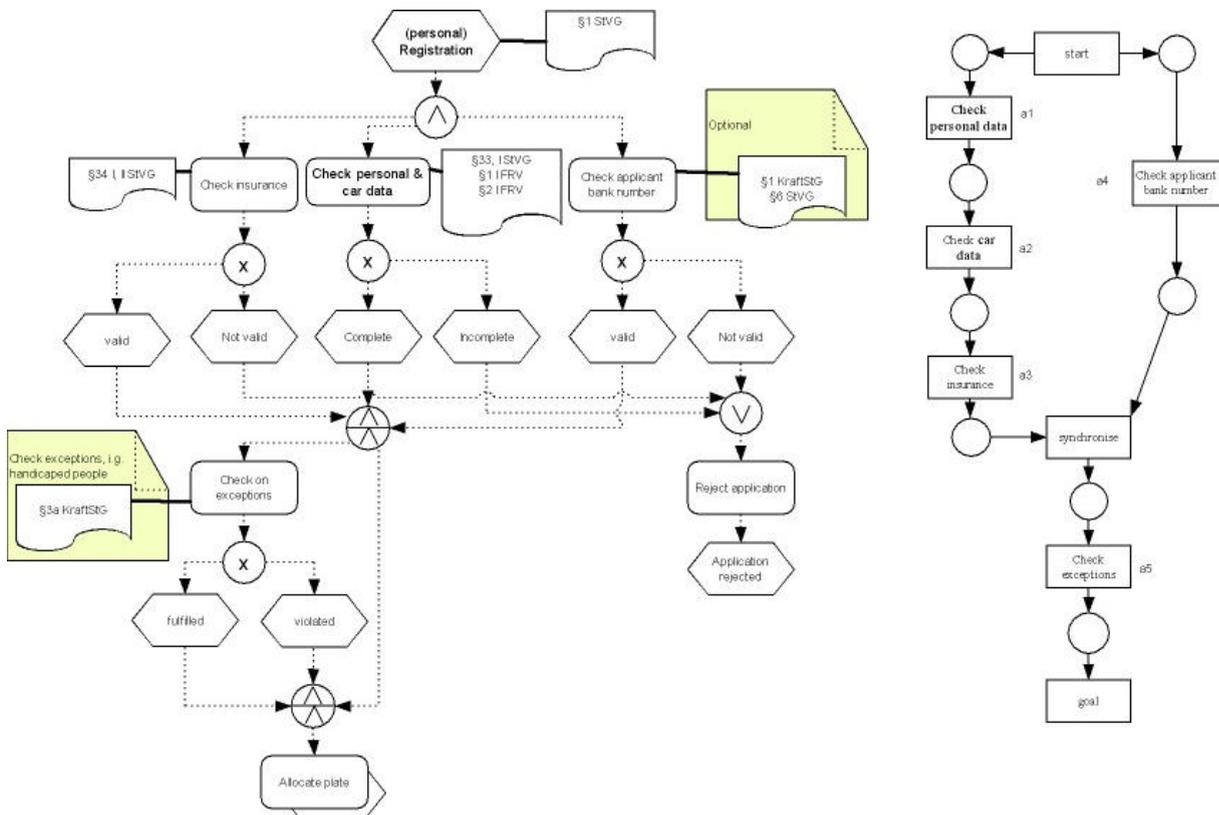


Figure 1: Registration of a motor vehicle according to the current German law

The eEPC model (see Figure 1 – left graph) gives an overview of the application process. Such semi-formal process documentations are increasingly important within the E-Government programs (see *BundOnline 2005* [BuOn05], *E-Government in Österreich* [eÖst03], etc.) because without process documentation there is little chance to successfully introduce E-

Government concepts [SKH03]. Although such semi-formal notations are sufficient for human beings to get an overall comprehension of regulations, they are not, nor were they intended to be, a formal specification against which an actual workflow implementation can be verified. A formal, verifiable model which could be used as a specification of a workflow management system, however, would be a qualitatively higher representation of the process. Moreover, formal methods are prerequisites for analysing processes against specifications and for examining their efficiency [Voss05]. Such a representation is necessary in order to precisely determine the influence of legal constraints. Hence, we take the workflow model of Figure 1 as input and translate it into a formal process specification using the Semantic Process Language [Simo05]. The result is depicted on the right model of figure 1.

The Semantic Process Language is an approach for process specification applied to business process modelling based on Petri nets. The meaning of the formulas of this language, called modules, is defined with the aid of *Module nets* – a kind of Petri nets closely related to relaxed sound *Workflow nets* [Aals98] and which are widely known in business process modelling [Dehn03]. *Module nets* are Petri nets with explicit *start* and *goal* transitions (where the preset of *start* and the postset of *goal* are empty). A *process* of a Module net is defined as a firing sequence reproducing the empty initial marking where *start* and *goal* occur exactly once (at the beginning and the end of the process, respectively). The transitions firing in-between are interpreted by actions and their sequence of occurrence indicates the process. This specific definition of processes allows conclusions about non-trivial properties of a Module net from the net's structure without calculating the full state space. The Module net representation on the right side of figure 1 is based on the following module of the Semantic Process Language (we began with the visualisation because it can be understood more intuitively):

$$M1 := [[a1 < a2 < a3] \wedge a4] < a5$$

The symbol “<” indicates a sequence of (sub-) processes, “^” indicates that processes occur independently.

2.3 Supporting the process under the current legislation

As we said in the beginning of this chapter, the registration of motor vehicles is one of the most executed public processes in Germany. Thus, we can identify a lot of variety in the implementations and a large number of similar process sets; e.g. the cancellation of a

registration of a vehicle, renewal of a registration, temporary registration, registration for transport of goods, and about 40 more. In order to demonstrate the influence of legal constraints and describe the challenges of harmonisation, we focus on the common registration of a car by a private citizen since this is the most common process [ScPe03]. Therefore, all figures in this paper will only describe this single process.

However, the registration process occurs that often because a decree to the StVG dictates that licence plates in Germany are bound to the region the private citizen lives in. Consequently, if a private citizen moves within Germany, she or he must change the licence plate of his/her vehicle, too. Since it was suggested abandoning this decree in June 2006, there is a strong political debate going on about the replacement of the local vehicle plates by individualised plates. Even though this is the most evident way to reduce the number of process occurrences, it does not simplify or change the process itself. The model would consequently represent the same organisational workflow. It is therefore, out of the scope of this paper.

Taking a closer look at figure 1, there is more than one way to reengineer the registration process towards E-Government. Starting with the application itself, some local authorities offer a web form to fill in the personal and vehicle data before actually going to the administration. The quite complex form is verified and the tax calculation can be conducted in advance. Although an applicant still has to appear at the respective office to personally hand over all needed files, the overall process time is reduced. The most obvious goal of E-Government would now be to work on a completely electronic process in order to make the personal appearance obsolete.

Currently, the restrictions on posting documents electronically or on sending the licence plates by post prevent a completely online execution. Yet, by giving someone power of attorney (§167 of German civil law, BGB) the personal appearance can also be taken over by a third party. It is a very common practise that for example car traders or intermediate brokers offer to the actual applicant for a certain fee. Hence, the graphs (we continue with eEPCs and Module Nets) in figure 2 are representing the process with this option.

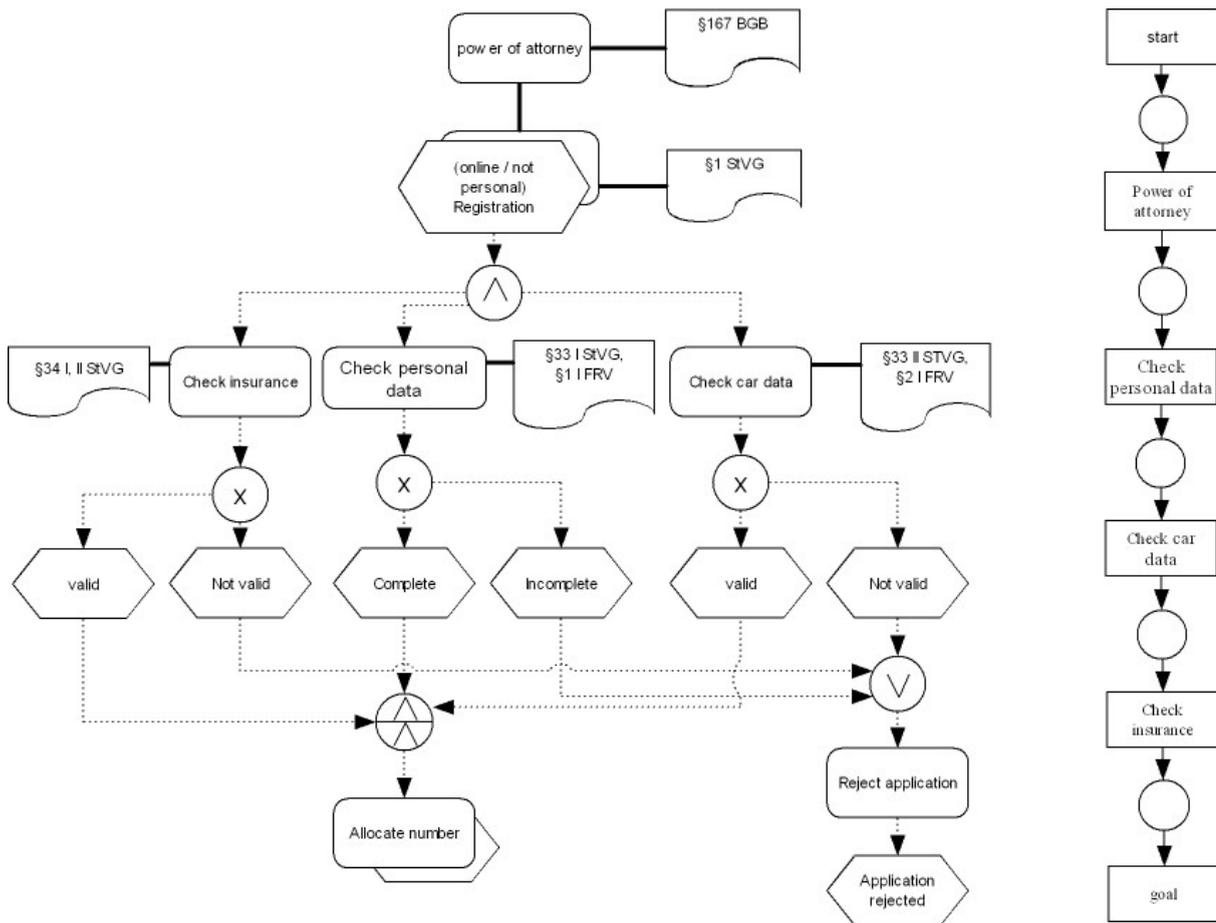


Figure2: Modified process in compliance with the current German law

In order to harmonise and simplify the process in compliance with the current legislation, all the optional process steps that refer to special circumstances or local variety (special permissions for farmers, checking of foreign cars, etc.) are not represented anymore. Since the information on a tax reduction is actually not bound to the vehicle but based on the individual's tax liability (KraftStG is a tax law), the collection of that information should not be part of the core registration process. Thus, the eEPC of figure 2 knows no check on the exceptions anymore. To offer a common process to the citizens (e.g. in the portal of a metropolitan area or a federal state) there must be an agreement, whether the liquidity (bank statement, etc.) of an applicant needs to be checked or not. Since the current law knows no regulation on that point, the checking is purely optional and therefore ignored in the process depicted in figure 2. Even though figure 2 still represents a legal exceptional process, it is very common process. We will, therefore, use it as (equivalent to figure 1) input for joining different process sets.

2.4 Alternative process sets from European Union countries

When we talk about alternative process sets in the context of E-Government, we basically talk about changing the legal framework in which the processes are defined. Once one reaches that point of a discussion, an even more drastic change to the process can be considered – up to a harmonisation of vehicle registration in the European Union. A recent survey of the University of Potsdam, which analysed the registration procedure of motor vehicles in all European Union member states, came to the result that most of the member states (except France and Germany) already changed the procedure of motor vehicle registration in order to achieve E-Government and more efficient processing [ScPe02].

Within all the changed procedures and changes in the legislation the study found one comparable trend within all the recent efforts: while the storage of the data was centralised (imposing strict conditions), the actual process of registering and distributing licence plates was outsourced by the public administration to private actors. The only option to apply at the authorities directly is mostly via the internet. In Sweden and the Netherlands, for example, these private actors are limited to car traders and insurance companies. Hence, the process could be reduced by one more step: the checking on the state of the car, or the insurance cover respectively. Not even the usual online registration/authentication procedure is needed, since a user can be identified via its contract/insurance number.

In the Netherlands, the process is completely regulated by the centre for vehicle technology and information (Rijksdienst voor Wegverkeer, RDW). The RDW (www.rdw.nl) is – since 1996 - an autonomous institution within the Dutch government which is responsible for all issues relating to the regulation of the traffic in the Netherlands. Their tasks include the surveillance of the motor vehicles (technical and registration issues) and the drivers (driver's licenses, etc.). Hence, the complete data set needed for a vehicle registration (i.e. insurance, personal data and technical surveillance data) is held on the RDWs' servers. If now a car is registered in one of the registration offices (Post office or car dealer), one has to access the RDWs' online service. Figure 3 illustrates this procedure.

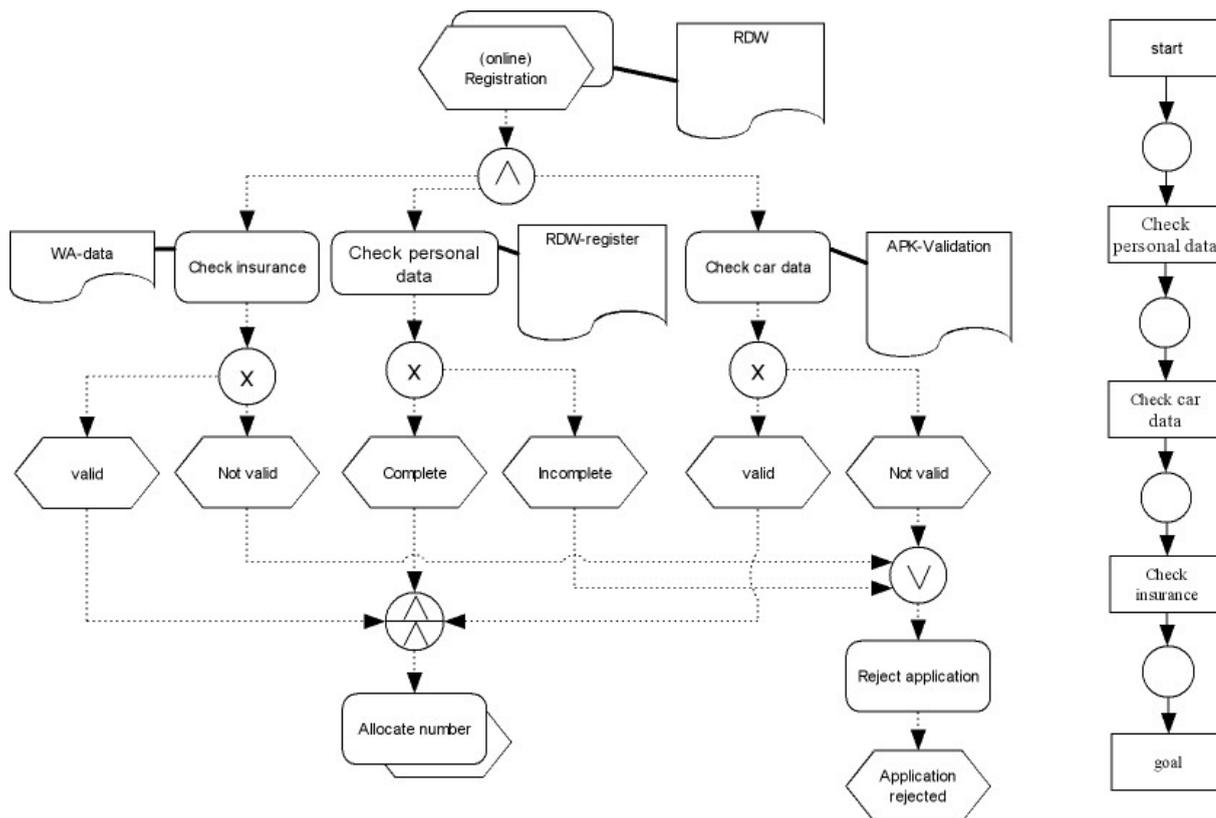


Figure 3: Dutch process of vehicle registration via the RDW

Of course, everything depends on the correctness of the central data on the RDWs' servers. Consequently, one is able to hand in new data (insurance = WA, technical data = APK) during the registration process. During the process, the personal data is cross checked with the data in the registration office.

Although the Dutch procedure is considered quite innovative, and indeed reduces the overall process time (compared to the German procedure), two shortcomings can be identified: first, the RDW holds redundant data which have to be updated frequently - e.g. if a technical surveillance deadline is reached. Second, the holding of too many personal data results in concerns about private data protection. On top of that, from a BPM point of view, one could ask why RDW is executing the registration task if private actors come into play; could those private actors not access the data directly?

In order to reform the Austrian motor vehicle registration process in late 1999 neither a new institution that holds the complete set of data was founded nor was any other organisational change made within the governmental organisation. The only legal change made was to

outsource the registration process to private actors. Unlike in the Netherlands, there were no specific actors picked, but several conditions on who and how to execute the registration task were created (*Zulassungsstellenverordnung – ZustV* [Zust98]). As demonstrated in figure 4, any private actor who fulfils the requirements of the *ZustV* can technically register cars in Austria.

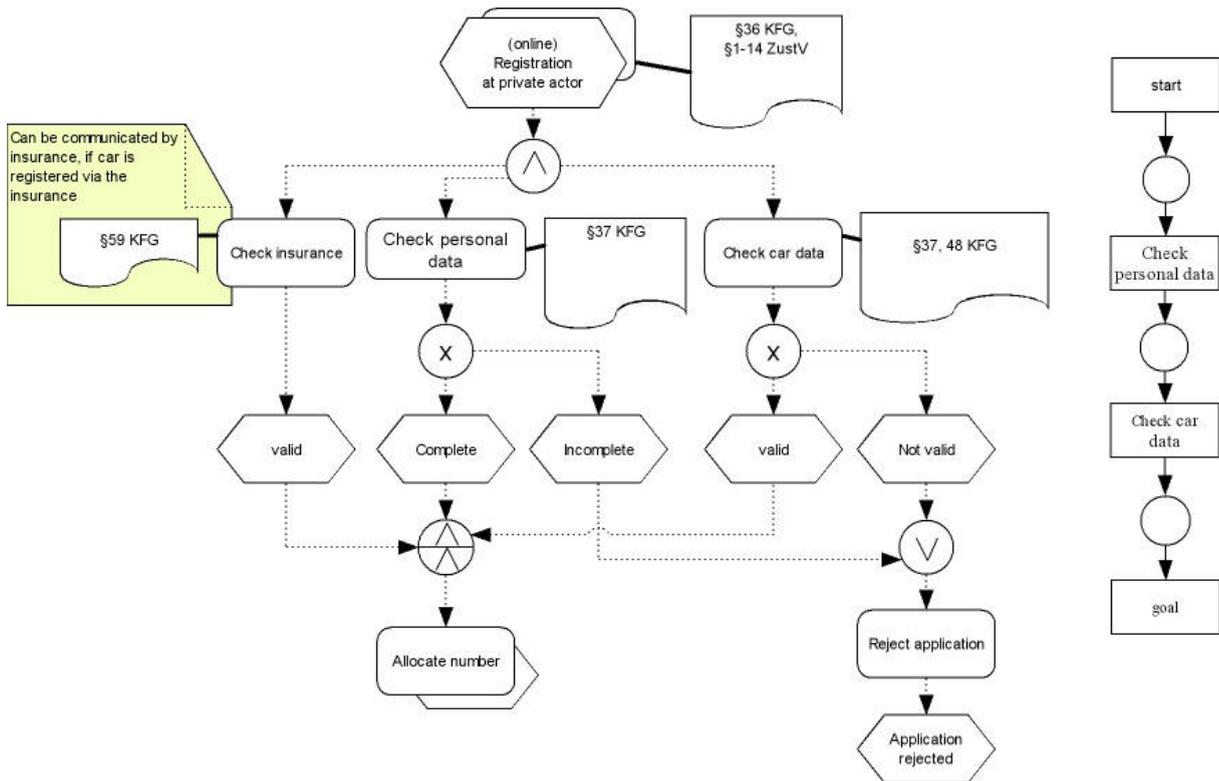


Figure4: Motor vehicle registration process in Austria

The private actors who execute the registration process in Austria turn out to be mostly insurances who offer the registration as a service to their clients. We can assume the insurance will only register cars that are properly insured with them. Hence, the checking of the insurance coverage can be ignored in the Austrian procedure. Since an insurance will check on the personal data and the correctness of the vehicle data anyway, the collection of information (and hence, the overall process time) is further reduced.

2.5 Harmonising the process across the European Union

Without the integration of legal constraints into the eEPC, the Module Nets of the modified German (figure 2), the Dutch (figure 3) and the Austrian (figure 4) motor vehicle registration

process would be quite similar. In fact the process steps (checking personal data, vehicle data and insurance coverage) must always be the same. Giving the power of attorney to a third party – as modelled in the German common-practise model in figure 2 - actually equals a lot the fully outsourced process as it is done in many European countries.

Nevertheless, the common German process of figure 2 is a legal exemption. The current legislation still describes a far more complex process as it is depicted in figure 1. In order to simplify the procedure in Germany, the (optional) checking must be abandoned. While all the countries that already reformed the registration of motor vehicles towards E-Government implemented centralised data storage and outsourced the front office of the registration process, the required data in Germany are still distributed asymmetrically; partly centralised and partly in local administration. Having built formal models, we can join the diagrams and see how a common process without any change of legislation (figure1, 3 and 4) would look like.

The formal process specification allows the verification of implemented process sets against this specification given that both are described by Module nets. For this, we have built the intersection of specification and implementation nets which is achieved by joining equally interpreted transitions, i.e. transitions that represent the same kind of actions. Afterwards, the join of the nets can be calculated automatically. Figure 5 shows the resulting net. Due to its complexity – resulting from Germany’s optional regulations – the resulting net obviously does not represent a desired outcome.

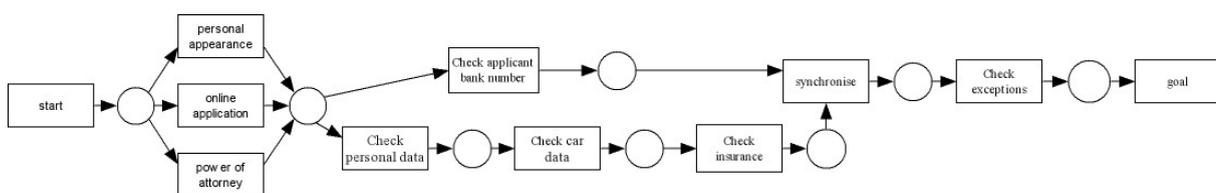


Figure5: Module Net of joined processes

In Germany, a first step towards harmonisation will be taken in 2008. By then, the Federal Bureau of Motor Vehicles and Drivers will store all personal and vehicle data needed for a registraion. The registration offices will dynamically access the data in order to issue a new registration. By then, the common process of car registration in Germany will already look more like the alternative process set of figure 2. Additionally, assuming the common practice of

third party representation (online applications, third party brokers, etc.), we join the nets of figure 2, 3 and 4 to get figure 6 as the resulting net.

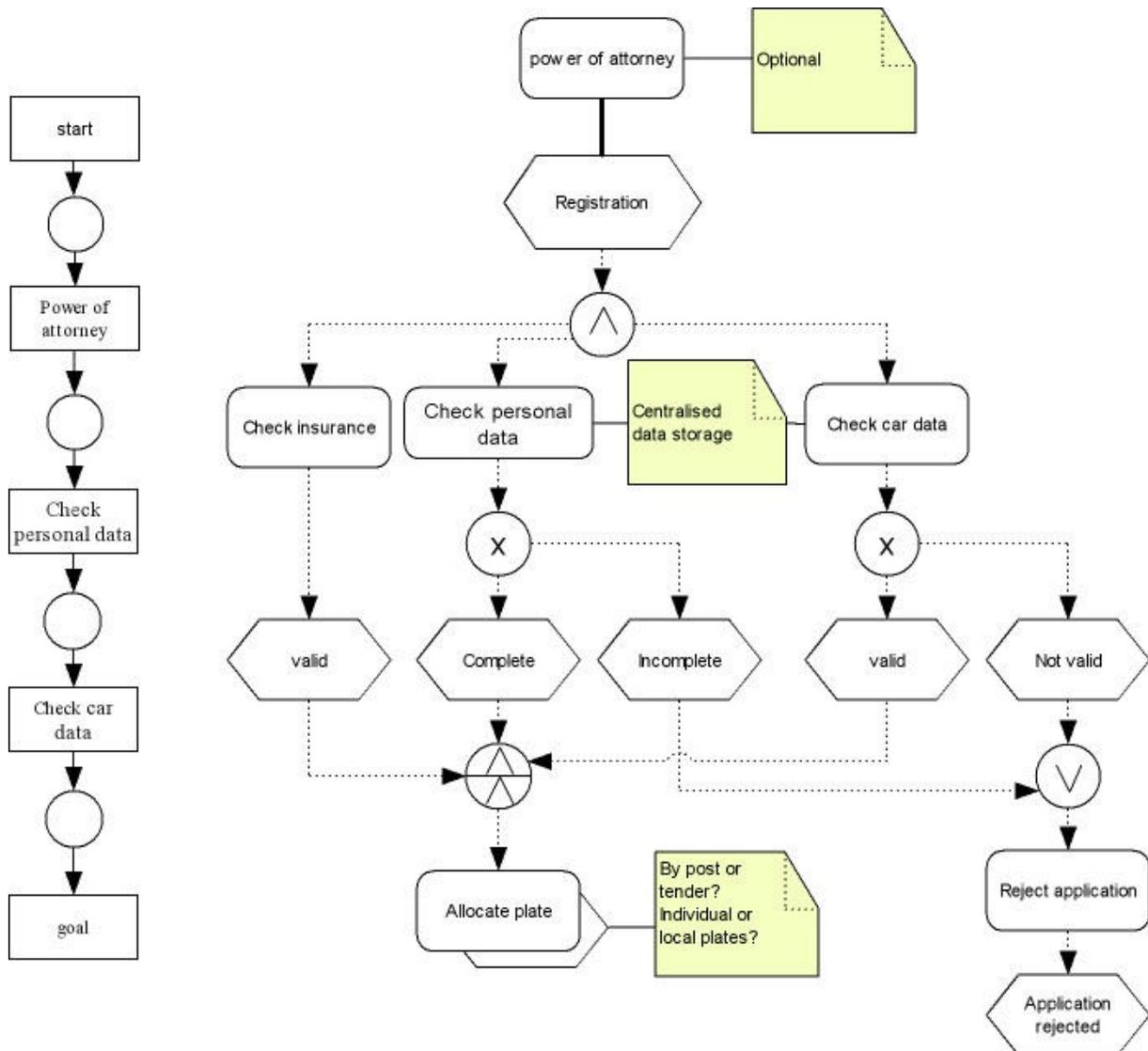


Figure6: Joined process set of modules 2, 3 and 4

As we can observe in the joint graph of figure 6, the registration is now harmonised within the three examined countries. To make the join work, however, two central assumptions have been made: first, there must be centralised data storage in order to enable a check on the data independently of the location of the front office. Second, as long as the registration process in Germany (and France) remains with the local authorities, the power of attorney must still be given in order to outsource the process to a third party.

Taking further EU member states into the account, it is questionable whether a complete harmonisation will be legally possible. In Sweden, for example, license plates are issued centrally and sent to the applicant by post. In all the other observed countries, licence plates must be received in person. Therefore, Sweden is so far the only country that executes the process completely online.

3 Conclusion

Businesses and private citizens are not only constrained by economic conditions but also by legal regulations formulated at different administrative levels – from cities, states and nations up to international agreements such as in the European Union. Although the legal regulations are of significant importance to the structure of the society, little research on the integration of processes exists.

In this paper, we have showed the influence of legal regulations on the definition of an E-Government process. As a demonstration example, we analysed the process of registering a motor vehicle. We compared the resulting process sets with workflow descriptions of equal tasks in different European countries. Our semi-formal process diagrams already point out that the core processes do not differ much. In order to analyse the complete harmonisation of the process in the European Union, a formal representation is required. Translating the resulted joined Module Net back into an eEPC, one can observe that actually little change in the process and the legal framework is required to achieve harmonisation. In contrast to previous work, which took the position that standard E-Government processes should respect local variety [OISi06], here it was shown that it is possible to push the German registration process towards a centralised E-Government strategy without losing functionality.

Using two different modelling approaches helps to address two different issues: eEPC, as one possible option to model business processes in a semi-formal way, allow us to explicitly include the legal framework and the organisational perspective into our considerations. Open workflow nets, as a representation of formal Petri-Nets, permit a mathematical verification of the joined processes. Additionally, inputs for workflow management systems and concepts of automation – such as web services [SFO06] - can be derived.

Our example also shows the paradigmatic approach of specifying a legal framework with a process model. The different processes described above are quite similar – only when adding legal requirements the different constraints become visible. By considering both, the formal and the semi-formal notation including the legal constraints, the verification of the actual workflow implementation against these models becomes possible. The current trend in (German) public administrations to integrate information systems and to restructure administration around government processes [Rein95] increases the necessity of such process modelling approaches.

4 Outlook

Applying Module Nets as a formal description of public workflows in addition to a semi-formal notation that represents laws as information objects appears to be a useful improvement for several ongoing projects. One of these is the E-Justice concept of the European Union which is part of the 6th Research Framework Program funded by the action plan for eEurope 2005 by the European Commission [EuCo05]. Next, we plan to enhance the user-interface management for public services [FrZa06] by our method of integrating the legal constraints. The goal is to improve European markets by deriving process models automatically that already include the legal constraints of each country. This could bring us closer to unified and transparent process sets within the European Union.

Literature

- [Aals98] van der Aalst W. M. P.: The Application of Petri Nets to Workflow Management. In: The Journal of Circuits, Systems and Computers, 1998.
- [ALOI05] Alpar P.; Olbrich S.: Legal Requirements and Modelling of Processes in e-Government. In: Electronic Journal of e-Government Volume 3 Issue 3 2005, www.EJEG.com (2006-07-15), Reading, GB, S.107-116.
- [BAD03] Becker, J.; Algermissen, L.; Niehaves, B.: Prozessmodellierung als Grundlage des E-Government – Ein Vorgehensmodell zur prozessorientierten Organisationsgestaltung am Beispiel des kommunalen Baugenehmigungsverfahrens. In: W. Uhr, W. Esswein and E. Schoop (Hrsg.): Wirtschaftsinformatik 2003 / Band II. Heidelberg 2003, S. 859-878
- [BuOn05] BMI. Umsetzungsplan für die eGovernment-Initiative Bund Online 2005. Bundesministerium des Inneren, Stabsstelle Moderner Staat - Moderne Verwaltung www.bundonline.de (24-03-2006).
- [Dehn03] Dehnert J.: A Methodology for Workflow Modeling - From business process modeling towards sound workflow specification. PhD thesis, TU Berlin 2003.
- [eÖst03] Posch, R. (Ed.): E-Government in Österreich: -Informationen für Wirtschaft und Verwaltung, Bundeskanzleramt Wien, 2003, <http://www.cio.gv.at> (07-04-2004).
- [EuCo05] European Commission: eEurope 2005 - An information society for all, http://europa.eu.int/information_society/eeurope/2005/index_en.htm (06-07-06).
- [FRV87] Fahrzeugregisterverordnung vom 20. Oktober 1987 (BGBl. I S. 2305).
- [FrZa06] Freiheit J.; Zangl A.: Model-based User-Interface Management for Public Services. In: D. Remenyi (editor): Conference proceedings of the 6th European Conference on Electronic Government (ECEG), Reading, GB 2006, S. 141-151.

- [KBK06] Knackstedt, R.; Brelage, C.; Kaufmann, N.: Entwicklung rechtsicherer Web-Anwendungen – Strukturierungsansatz, State-of-the-Art und ausgewählte Aspekte der fachkonzeptionellen Modellierung. In: WI 48 (2006) 1, S.27-35.
- [KraftStG02] Kraftfahrzeugsteuergesetz i.d.F. vom 26. September 2002 (BGBl. I S. 3818), geändert durch Artikel 30 des Gesetzes vom 21. Juni 2005 (BGBl. I S. 1818).
- [LeTr99] Lenk, K.; Traummüller, R.: Öffentliche Verwaltung und Informationstechnik – Perspektiven einer radikalen Neugestaltung der öffentlichen Verwaltung mit Informationstechnik, Verwaltungsinformatik Nr. 20, Decker Verlag, 1999.
- [OlSi06] Olbrich S.; Simon C.: Process engineering towards E-Government – modelling process variety and best practice. In: D. Remenyi (editor): Conference proceedings of the 6th European Conference on Electronic Government (ECEG), Reading, GB 2006, S. 339-349.
- [Rein95] Reinermann H.: Perspektiven einer Verwaltungsreform mittels Informationstechnik. In: R. Traummüller (Hrsg.), Geschäftsprozesse in öffentlichen Verwaltungen - Neugestaltung mit Informationstechnik, Heidelberg 1995, S. 53-69.
- [Sche94] Scheer A. W.: Business Process Engineering, ARIS-Navigator for Reference Models for Industrial Enterprises. Springer, Heidelberg 1994.
- [ScPe02] Schuppan, T.; Penning-Poggenbeck, J.: eGovernment im Kfz-Zulassungswesen Konzeption zur Umsetzung, Kommunalwissenschaftliches Institut (KWI) der Universität Potsdam, KWI-Projektberichte 2, Potsdam 2002, <http://www.uni-potsdam.de/u/kwi/publ/kwi-pb.htm> (01-06-2006).
- [ScPe03] Schuppan, T.; Penning-Poggenbeck, J.: Kundenanalyse und Evaluierung zu eGovernment im Kfz-Zulassungswesen - Zwei empirische Studien. In: Prof. Dr. Michael Nierhaus (Hrsg.), Kommunalwissenschaftliches Institut (KWI) der Universität Potsdam, KWI-Projektberichte 7, Potsdam 2003.
- [SFO06] Jörn Freiherr, Carlo Simon, Sebastian Olbrich: Usability of Event-driven Process Chains, EPK-Kongress der GI, Wien 2006 (accepted).

- [Simo05] Simon C.: Incremental Development of Business Process Models. In: U. Frank, J. Desel (Hrsg.) EMISA 2005 - Development Methods for Information Systems and their Application, Gesellschaft für Informatik, Lecture Notes in Informatics P-56, Klagenfurt 2005, S. 95-106.
- [SKH03] Scheer, A.-W.; Kruppke, H.; Heib, R.: E-Government – Prozessoptimierung in der öffentlichen Verwaltung, Springer Verlag, Berlin Heidelberg, 2003.
- [SnZu97] Snellen, I.; Zuurmond, A.: From Bureacracy to Infocracy: Management through Information Architecture, in: Tyler, Snellen, Zuurmond (Ed.), Beyond BPR in Public Administration, Amsterdam, IOS Press, 1997, S. 205-224.
- [StVG03] Straßenverkehrsgesetz in der Fassung der Bekanntmachung vom 5. März 2003 (BGBl. I S. 310, 919), zuletzt geändert durch Artikel 2 des Gesetzes vom 14. August 2006 (BGBl. I S. 1958).
- [Voss05] Vossen G.: Was Informatiker und Wirtschaftsinformatiker zu Prozessen beitragen. In: HMD - Praxis der Wirtschaftsinformatik (241) 2005, S. 5-6.
- [VBPO05] Kompetenzzentrum Vorgangsbearbeitung und Organisation: Softwareprodukte zur Geschäftsprozessanalyse und –Optimierung - Bedeutung ihres Einsatzes und eine Übersicht zu ausgewählten Produkten. In: E-Government-Handbuch des Bundesamts für Sicherheit in der Informationstechnik (BSI), Kapitel 6, (online 2005-12-17): http://www.bsi.bund.de/fachthem/egov/download/6_GPA.pdf.
- [WiTr03] Wimmer, M.; Traummüller, R.: Geschäftsprozessmodellierung in E-Government: eine Zwischenbilanz, eGov days 2003, www.egov.ocg.at (06-02-2004).
- [WoKr05] Wolf, P.; Krcmar, H.: Prozessorientierte Wirtschaftlichkeitsuntersuchung für E-Government. In: Wirtschaftsinformatik 47 (2005) 1, S.337-346.
- [Zust98] Verordnung des österreichischen Bundesministers für Wissenschaft und Verkehr, mit den Bestimmungen über die Einrichtung von Zulassungsstellen (Zulassungsstellenverordnung - ZustV) StF: BGBl. II Nr. 464/1998.

Prozessorientierte Gestaltung von Behördenkontakten

Untersuchung zu eGovernment-Anforderungen aus Unternehmenssicht

Petra Wolf, Helmut Krcmar

Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
85748 Garching b. München
{petra.wolf;krcmar}@in.tum.de

Abstract: Im vorliegenden Artikel wird ein Ansatz zur prozessorientierten Ermittlung von eGovernment-Anforderungen aus Unternehmenssicht entwickelt. Als Ergebnis einer explorativen Untersuchung können Merkmale von Unternehmenslagen identifiziert werden, die ausschlaggebend sind für technische und organisatorische Gestaltungsaspekte unternehmerischer Behördenkontakte. Aus den ermittelten Anforderungen wird das Konzept einer eGovernment-Kooperationsplattform abgeleitet.

1 eGovernment-Anforderungen aus Unternehmenssicht

1.1 Hintergrund

eGovernment doesn't matter! So könnte die Quintessenz zahlreicher Gespräche mit Unternehmensvertretern lauten. eGovernment, d.h. die elektronische Abwicklung von Behördenkontakten, spielt aus unterschiedlichen Gründen für viele Unternehmen heute noch keine oder nur eine untergeordnete Rolle. Während vor einigen Jahren in erster Linie die mangelnde Bedarfsorientierung der Online-Angebote der öffentlichen Verwaltung als Ursache für geringe Akzeptanz und Nutzung ins Feld geführt wurde [InDD02], wird heute spezifischer die überfällige Umstellung auf das in der Wirtschaft akzeptierte Paradigma der Prozessorientierung angemahnt [o.A.06a]. Erkenntnisse zum Unterstützungsbedarf der Wirtschaft bei Behördenkontakten beschränken sich bislang auf sehr allgemeine Aussagen, die kaum für die Gestaltung konkreter Angebote herangezogen werden können [vgl. BeBu03; KnIn04; vgl. o.A.02]. Gefordert ist daher eine systematische Erhebung dieses Bedarfs orientiert an den Geschäftsprozessen, die un-

ternehmerischen Behördenkontakten zugrunde liegen. Mit Unterstützung der SAP Deutschland AG wurde eine solche Untersuchung zu unternehmerischen Behördenkontakten durchgeführt. Ziel der Untersuchung war die Identifikation charakteristischer Merkmale von B2G-Prozessen, um auf dieser Basis Anforderungen aus Unternehmensperspektive an die Gestaltung von eGovernment formulieren zu können. Der vorliegende Artikel stellt Vorgehensweise und Ergebnisse der Untersuchung dar.

1.2 Prozessorientiert im eGovernment

Aus Unternehmenssicht wird der elektronisch unterstützte Behördenkontakt als eine weitere Ausprägung von eBusiness verstanden – in diesem Fall mit dem Kooperationspartner Behörde [vgl. ScHä01]. In diesem Sinne geht es bei der Gestaltung von eGovernment darum, die Abwicklung des Behördenkontaktes möglichst nahtlos in die eBusiness-Infrastruktur des Unternehmens einzugliedern. Informationen, die für die Erstellung von Anträgen erforderlich sind, werden in der Regel bereits elektronisch erfasst und in vielen Fällen auch für weitere Zwecke im Unternehmen benötigt. Für den Behördenkontakt werden diese Informationen zusammengetragen, in die vorgegebene Formularstruktur gebracht und an die zuständigen Ämter übergeben, wo sie von verschiedenen Fachbereichen begutachtet werden.

Eine prozessorientierte Betrachtung von eGovernment fokussiert auf die Unterstützung der Zusammenarbeit der Partner auf Unternehmens- und Behördenseite, im Gegensatz zu einer verkürzten Deutung des Begriffs eGovernment, der häufig lediglich die Unterstützung der Übergabe von Informationen an der Schnittstelle zwischen Verwaltungskunde und Behörde meint.

Die prozessorientierten Unterstützung und Koordination der Zusammenarbeit von Unternehmen und Behörde mit Hilfe von IT verspricht Nutzenpotenzial für alle Beteiligten [o.A.06b; WoKr05, S. 344]. Die Grundvoraussetzung dafür, die Neuausrichtung von Betrachtungsperspektive und -fokus sowie die Erweiterung des Konzepts der Unternehmenslage, werden im Folgenden erläutert.

2 Untersuchungsansatz

2.1 Unternehmenslagen

Um die Orientierung im umfangreichen Angebot von Verwaltungsleistungen zu erleichtern, werden diese in Bürgerportalen nach dem Prinzip der Lebenslagen gruppiert [LeTr01, S. 17f]. Analog findet mittlerweile auf zahlreichen eGovernment-Portalen eine Gruppierung nach Unternehmenslagen statt.

Im Sinne eines prozessorientierten eGovernment sind diese Unternehmenslagen um behördenähnliche Kontakte (bspw. Wirtschaftsprüfer, Zertifizierung) zu ergänzen, die aus Sicht von Unternehmen im jeweiligen Kontext ebenfalls erledigt werden. Dies gilt nicht nur für die integrierte Darstellung von Informationen zu Ansprechpartnern und Ablauf, sondern auch für die Unterstützung einer möglichst effizienten Abwicklung.

Wichtig für die Analyse und Unterstützung von eGovernment-Prozessen ist neben deren vollständiger Betrachtung auch die Bewertung und Typisierung aus der Sicht von Unternehmen. Wie in anderen Bereichen ist die Wahrnehmung eines Ablaufs und seiner charakteristischen Merkmale und Ziele durch die Akteure entscheidend dafür, ob eine elektronische Unterstützung für sinnvoll oder nützlich erachtet wird. Die Identifizierung dieser Merkmale ist daher Voraussetzung für eine bedarfsorientierte eGovernment-Entwicklung.

2.2 Betrachtungsperspektiven

2.2.1 Perspektivwechsel

Der Wechsel der Betrachtungsperspektive von der bislang vorherrschenden verwaltungszentrierten [vgl. BrGi02, S. 8f] hin zur Kundensicht gehört zu den Grundprinzipien des hier dargestellten Verständnisses von prozessorientiertem eGovernment, d.h. nicht die Analyse von Produktkatalogen der Verwaltung ist Ausgangspunkt der Untersuchung, sondern die Auslöser und Anknüpfungspunkte von Behördenkontakten in Unternehmen.

Die Analyse von Unterstützungsbedarf auf Seiten der Verwaltungskunden rückt in den Vordergrund, da die Akzeptanz von eGovernment ausschlaggebend ist für den Erfolg. Gerade im Hinblick auf die Relation von Kosten und Nutzen hat die Verwaltung mit Unternehmen besonders kritische Kunden, die eGovernment-Angebote nur dann nutzen, wenn diese im Vergleich zum herkömmlichen Verfahren weniger aufwändig [Lee05] und ihre Vorteile klar ersichtlich sind.

2.2.2 Fokuswechsel

Eine prozessorientierte Betrachtung von eGovernment verlagert den Fokus weg von der Schnittstelle zwischen Unternehmen und Verwaltung, an der Anträge oder Informationen

übergeben werden. Ziel ist es, die Verknüpfung des Behördenkontaktes mit dem auslösenden Geschäftsprozess im Unternehmen als Ausgangspunkt zu nehmen, um dort weitere Anknüpfungspunkte zu Behördenkontakten oder behördenähnliche Kontakten zu identifizieren und in die Unterstützung zu integrieren. Ebenso gilt es auf Seiten der beteiligten Behörden, die relevanten Kooperationspartner zu identifizieren und über eine gemeinsame Prozessunterstützung miteinander zu verbinden.

2.3 Betrachtete Unternehmenslagen

Im Rahmen der Untersuchung wurden die Unternehmenslagencluster Abfall, Bauen, Umwelt, Gewerbeauflagen sowie Produktzulassung, Patentierung und -genehmigung betrachtet.

Da hier größtenteils Behörden auf kommunaler und Landesebene zuständig sind, wurden Behörden in unterschiedlichen Bundesländern (Bayern, Berlin, Brandenburg, Hamburg, Hessen, Rheinland-Pfalz, Sachsen) befragt, um regionale Unterschiede auszugleichen bzw. zu identifizieren. Abgesehen von unterschiedlichen Zuständigkeiten konnten allerdings keine bedeutsamen Unterschiede festgestellt werden. Auch hinsichtlich der betrachteten Unternehmensgrößen (s. Tab. 1, Klassifikation nach [Inst02]) konnten aufgrund der geringen Fallzahlen keine signifikanten Unterschiede abgeleitet werden. Die Unternehmen wurden bewusst ausgewählt, da die Komplexität des Themas sowie die Schwierigkeit der Verortung von eGovernment-Ansprechpartnern eine andere Vorgehensweise nicht sinnvoll erscheinen ließ.

	Unternehmen		Behörden
	mittel	groß	
Anzahl Interviews zu Abfall/Bauen/Umwelt/Gewerbeauflagen	2	5	17
Anzahl Interviews zu Produktzulassung/Patentierung	3	4	4

Tab. 1: Anzahl der geführten Interviews

Für die Erhebung wurde ein strukturierter Interviewleitfaden entwickelt, der als Grundlage für 30-60-minütige Interviews (telefonisch und persönlich) diente. Unternehmen wurden zunächst nach Einbettung der jeweiligen Verwaltungskontakte in die Unternehmensprozesse befragt und die Abwicklung in Form von Kurzepisoden aufgenommen. Im Anschluss wurden Fragen zur aktuellen Informations- und Abwicklungssituation und möglichem Unterstützungspotenzial durch IT diskutiert. Interviews mit Behördenvertretern verliefen analog.

Die Auswertung der Interviewprotokolle erfolgte in Form einer hermeneutischen Inhaltsanalyse [vgl. Tsch96, S. 38ff], die in mehreren Iterationszyklen auf die episodischen Prozessdarstellungen

gen angewendet wurde, um Merkmale von Unternehmenslagen und Unterstützungsanforderungen anzuleiten. Die dort gewonnenen Erkenntnisse wurden mit Hilfe der Ergebnisse aus den strukturierten Interviewteilen auf ihre Plausibilität hin überprüft.

3 Merkmale von Unternehmenslagen

Aus den Ergebnissen der Interviews lassen sich verschiedene Merkmale ableiten, die die Eigenschaften von B2G-Kontakten aus den Perspektiven von Unternehmen und Behörden charakterisieren. Diese Merkmale sind ausschlaggebend dafür, ob, in welchem Umfang und in welcher Form diese Austauschprozesse unterstützt werden können oder sollen.

Bei den betrachteten Unternehmenslagen lassen sich Entscheidungscharakter, Routinegrad, Komplexität und Phase als Faktoren identifizieren, die aus Sicht der beteiligten Akteure Einfluss auf die Gestaltung von eGovernment haben. Diese werden im Weiteren dargestellt.

3.1 Entscheidungscharakter

Der Prozess des Neubaus einer Fertigungsanlage hat für unternehmerische Antragsteller i.d.R. Aushandlungscharakter, da auf Seiten der genehmigenden Behörden ein gewisser Entscheidungsspielraum besteht. Bauanträge werden daher bereits im Vorfeld in persönlichen Gesprächen zwischen Antragsteller und Ansprechpartnern auf Behördenseite diskutiert, um die jeweiligen Schwerpunkte und ggf. Vorbehalte kennen zu lernen. Für die Unterstützung im Sinne von eGovernment bedeutet dies, dass diese Phase der persönlichen Abstimmung bzw. multilateralen Verhandlung nicht durch elektronische Workflows ersetzt werden kann [Lenk02].

Im Gegensatz dazu hat die Beantragung industrieller Abfälle den Charakter einer Konditionalentscheidung [vgl. Fuch88, S. 13f]: Werden die geforderten Nachweise und Informationen erbracht, kann mit einer positiven Entscheidung gerechnet werden. Das wird auch durch die bereits realisierte elektronische Unterstützung bis hin zu einer Teilautomatisierung in diesem Bereich widergespiegelt.

3.2 Routinegrad

In Bezug auf Workflow-Unterstützung werden i.d.R. strukturierte, teilstrukturierte und unstrukturierte Geschäftsprozesse unterschieden [Enge00, S. 402f]. Der Routine- oder Strukturierungsgrad von Behördeninteraktionen in Zusammenhang mit der Frequenz, in der diese anfallen

(monatlich, jährlich, seltener), ist auch ausschlaggebend für das Potenzial das einer elektronischen Unterstützung seitens der Unternehmen beigemessen wird. Stark routinisierte Prozesse wie bspw. die Lieferung statistischer Informationen oder Umsatzsteuervoranmeldungen werden als sehr geeignet für automatisierte Workflow-Unterstützungen gesehen. Häufig gehören solche Prozesse auch zu den Behördenkontakten mit wenig oder keinem Entscheidungsspielraum.

Behördenkontakte, die eher Einzelfallcharakter haben und/oder nur selten erfolgen, sind auf Seiten des Unternehmens auch nicht fest mit bestimmten Geschäftsprozessen und Organisationsstrukturen verknüpft, so dass das Potenzial einer eGovernment-Unterstützung eher im Bereich der Koordination von Kooperationen und der Informationen über Zuständigkeiten und Vorgaben auf Behördenseite sowie Fristen und Formulare gesehen wird. Darüber hinaus ist in vielen dieser Bereiche (Gewerbeaufsicht, Abwasser, Emissionen) Expertenwissen erforderlich, das im Unternehmen nicht immer aktuell vorgehalten werden kann. Automatisierte Prozesse würden hier die Gefahr bergen, den Überblick bzw. die Hoheit über das Verfahren zu verlieren. Unterstützung im Sinne von Aufbereitung von Hintergrundinformationen und Hilfestellung auf Abruf stoßen in diesen Situationen auf deutlich mehr Akzeptanz. Insbesondere die Bereitstellung von Informationen über Aktualisierungen oder Änderungen des Verfahrens oder zu Fälligkeitsdaten wird hier als hilfreich angesehen.

3.3 Komplexität

Die Komplexität eines kooperativen eGovernment-Prozesses ist Ausdruck sowohl der Anzahl der beteiligten Kooperationspartner auf Unternehmens- und Behördenseite als auch der Prozesslänge und der Zahl der Prozessvarianten [vgl. PiRe85]. Der Bauantragsprozess ist ein typisches Beispiel für einen Prozess mit zahlreichen Beteiligten und mehreren Varianten, je nachdem, um welche Art von Bau es sich handelt und welche Vorschriften und Vorgaben dadurch berührt werden. Die Komplexität eines unternehmerischen Behördenkontaktes hat ebenfalls Einfluss auf die Anforderungen und Möglichkeiten der eGovernment-Unterstützung.

eGovernment-Prozesse von geringer Komplexität mit wenigen Alternativvarianten und einer überschaubaren und v.a. stabilen Menge von Akteuren bieten sich eher für eine weitergehende Integration und eine Unterstützung im Sinne eines Workflows an. Dies wird am Beispiel des Abfallbegleitscheins anschaulich dokumentiert. Hochkomplexe Prozesse mit wechselnden Akteuren und verschiedenen Varianten stellen insbesondere aus Unternehmenssicht die Anforderung einer Koordinationsunterstützung, die flexible Integrationsmöglichkeiten bietet. Aufgrund der hohen Variantenbreite bietet sich eine Workflow-Unterstützung nicht an.

Aus Behördensicht werden die mit einer Unternehmenslage verbundenen anderweitigen Kooperationsbeziehungen von Unternehmen bspw. zu einem Zertifizierer oder einem spezialisierten Berater ausgeblendet. Eine Unterstützung, die diese Kontakte integriert, muss daher von dritter Seite angeboten werden.

3.4 Merkmale der betrachteten Unternehmenslagen

Die Einordnung der betrachteten Unternehmenslagen in das entwickelte Merkmalsraster ist in Tab. 2 dargestellt.

	Entscheidungscharakter		Routinecharakter		Komplexität	
	Konditional	Aushandlung	Einzelprojekt	Routine	wenig	hoch
Abfall	x			x	x	
Bauen		x	x			x
Umwelt		x	Antrag	Bericht		x
Gewerbeaufsicht	x		Antrag	Bericht		x
Produktzulassung	x		x	(x), branchenabh.	x	
Patentanmeldung	x		x		x	

Tab. 2: Merkmalsausprägung für die betrachteten Unternehmenslagen

Die Anforderungen, die sich aus den kombinierten Merkmalen Entscheidungs- und Routinecharakter ableiten lassen, sind in Tab. 3 zusammengefasst. Eine weitere Überlagerung der resultierenden Anforderungen mit dem Merkmal Komplexität bietet sich aus Gründen der Nachvollziehbarkeit nicht an.

	Konditionalcharakter	Aushandlung
Einzelprojekt	Expertenwissen und strukturelle IT-Unterstützung (Formulare, Fristen, Ansprechpartner) für effiziente Abwicklung erwünscht	IT-Unterstützung (Formulare, Fristen, Ansprechpartner) für bestimmte Planungsphasen erwünscht, Aushandlung persönlich
Routine	Hoher Automatisierungsgrad umsetzbar und akzeptabel; Bedürfnis der indiv. Einflussnahme gering	Vorbereitung, Einreichung, Begutachtung je nach Bereich IT-gestützt, keine Automatisierung, flexible Nutzung der Instrumente und jederzeit Medienwechsel möglich

Tab. 3: Anforderungen aus Entscheidungscharakter und Routinegrad

4 Unterstützungsansätze für prozessorientiertes eGovernment

Nicht nur aus den charakteristischen Merkmalen von Unternehmenslagen lassen sich Gestaltungsanforderungen ableiten: Auch die Unterscheidung verschiedener Phasen von B2G-Interaktionen zeigt, dass sich phasenbezogen spezifische Unterstützungsanforderungen ergeben. Diese werden im Folgenden dargestellt und anschließend Überlegungen zum Design für prozessorientiertes eGovernment vorgestellt.

4.1 Phasen

Der Prozess einer Antragstellung oder eines beliebigen B2G-Kontaktes lässt sich in der Regel in drei grundlegende Phasen untergliedern (vgl. Abb. 1): 1. Vorbereitung von Unterlagen für einen Antrag oder einen Bericht, 2. Übergabe der offiziellen und fixierten Dokumente an die Behörde und 3. Bearbeitung bzw. Begutachtung der eingereichten Dokumente durch die Behörde.

Im Rahmen der **Vorbereitung und Koordination** eines Antrags oder eines Berichts hat der Antragsteller bzw. Berichtsverfasser die Hoheit über das Verfahren, d.h. das betreffende Unternehmen entscheidet, welche internen und ggf. externen Kooperationspartner in welcher Form zu beteiligen sind. Es gibt je nach Behördenkontakt zwar amtliche Vorgaben und Rahmenbedingungen für die Abwicklung des Verfahrens, die bereits zu diesem Zeitpunkt prägend sind, aber große Teile der Ausgestaltung der Vorbereitung liegen in der Hand des Unternehmens.

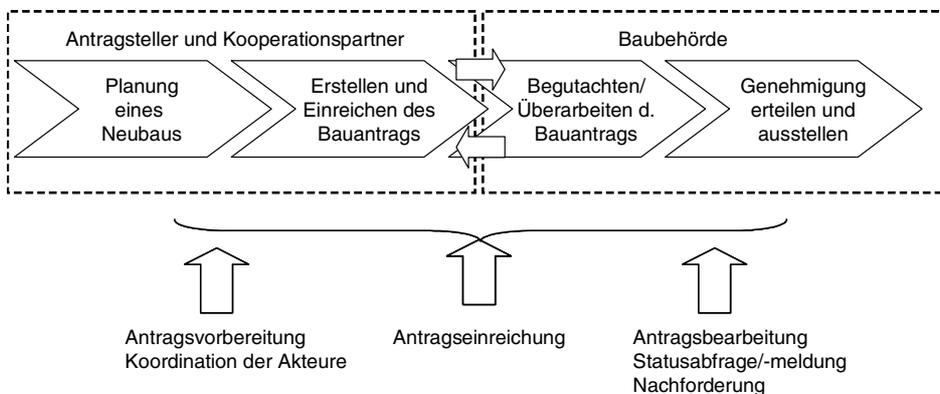


Abb. 1: Phasen des Behördenkontakts

Zu den Anforderungen in dieser Phase gehört die Koordination der Aufgaben und Zusammenarbeit der verschiedenen Kooperationspartner. Diese sind u.U. verteilt an verschiedenen Standorten und kommen aus verschiedenen Fachbereichen. Informationen werden gesammelt und weiterverarbeitet. Die gesammelten Informationen werden ggf. für unterschiedliche Adressaten in verschiedenen Formaten aufbereitet. Aufwand entsteht auch dadurch, dass Informationen aus verschiedenen Quellsystemen extrahiert werden und die erforderlichen Berichts- oder Antrags-

formate übertragen werden müssen. Nur selten gibt es Exportschnittstellen, die diesen Arbeitsschritt unterstützen.

In dieser Phase werden Vorversionen des abschließenden Antrags oder Berichts hergestellt, die noch inoffiziellen Status innehaben. Sie werden von den einzelnen Kooperationspartnern bearbeitet und schließlich nach verschiedenen Qualitätskontrollen in eine Endfassung für die Abgabe überführt.

Von Seiten der Behörde werden Informationen benötigt über Abgabefristen, Format- und Formularvorgaben, Verfahrensschritte, Ansprechpartner, zu beachtende Rahmenbedingungen etc. Diese Informationen sollten kontextbasiert zur Verfügung stehen und sind im Fall von häufig oder regelmäßig auftretenden Routineprozessen auch im Push-Format sinnvoll. Gleichzeitig besteht von Seiten der Unternehmen die Anforderung, diese Phase in Eigenregie gestalten und so möglichst gut mit den bestehenden Unternehmensprozessen und der genutzten IT-Unterstützung abstimmen zu können. Die mit den Kooperationspartnern in dieser Phase ausgetauschten Informationen und Vorfassungen sind nicht für Außenstehende, auch auf Seiten der Behörde, bestimmt und sollten eindeutig vor deren Zugriff geschützt werden können.

In der Phase der offiziellen Übergabe bzw. **Einreichung von Antragsunterlagen** oder Berichten an eine Behörde ist insbesondere die rechtlich verbindliche Dokumentation dieser Transaktion von Bedeutung. Daneben legen die beiden beteiligten Kooperationspartner auch großen Wert auf eine eindeutige Authentifizierung und den Schutz der ausgetauschten Dokumente gegen unberechtigten Zugriff.

Aus Sicht des Unternehmens sind weitere wichtige Parameter eines Behördenkontakts eine eindeutige Eingangsbestätigung falls erforderlich mit Zeitstempel, die ggf. auch eine Referenznummer o.ä. für weitere Rückfragen liefert. Je nach Formatanforderungen wird bei einer elektronischen Einreichung an dieser Stelle auch die Transformation der Unterlagen in ein behördenkonformes Format, [vgl. Dokumentation Abfa06] relevant. Hierfür sind entsprechende Upload-Mechanismen oder eine geeignete Softwareunterstützung erforderlich. Im Fall komplizierter Antragsformulare können an dieser Stelle auch noch einmal Plausibilitätschecks eine letzte Qualitätskontrolle unterstützen.

Sind die Unterlagen im physischen oder virtuellen Posteingang der Behörden angelangt, beginnt die Phase der **Vorgangsbearbeitung** in der Sphäre der Behörde. Die Hoheit über diesen Abschnitt des Verfahrens liegt klar auf Seiten der Behörde. Im Fall mehrerer beteiligter Ämter oder Fachbereiche liegt die Koordinationsrolle bei einer federführenden Behörde, die die Zu-

weisung von Aufgaben und die Einhaltung von Fristen koordiniert und kontrolliert. Für die meisten der betrachteten Unternehmenskontakte gibt es auf Seiten der Verwaltung spezialisierte Informationssysteme, in denen Daten zu Antragstellern verwaltet oder auch Workflows unterstützt werden. Je nach Format des eingehenden Berichts oder Antrags besteht daher der erste Schritt darin, die eingegangenen Daten in das entsprechende System zu transferieren. Bei diesem Schritt wird i.d.R. eine erste Qualitätskontrolle der Unterlagen hinsichtlich Vollständigkeit und anderer formaler Kriterien durchgeführt. Bei einer elektronischen Unterstützung des Übergabeschritts können viele dieser Tätigkeiten automatisiert oder zumindest unterstützt werden. Dasselbe gilt für die Erstellung einer Eingangsbestätigung zusammen mit dem Ergebnis der formalen Prüfung und der Vergabe einer Vorgangsnummer.

Während der Vorgangsbearbeitung durch die Behörde sind je nach Vorgang verschiedene Statusmeldungen für den Antragsteller vorgesehen. Diese dienen einerseits dazu, eine gewisse zeitliche Berechenbarkeit für alle Beteiligten herzustellen und andererseits auch feste Punkte für eine Interaktion zwischen den Beteiligten zu haben bspw. zur Nachforderung von Unterlagen.

Aus Sicht der Behörde steht eine zügige und effiziente, dabei aber auch fachgerechte und rechtskonforme Abwicklung des Vorgangs im Vordergrund. Interaktionen mit Antragstellern werden daher auf festgelegte Punkte im Verfahren beschränkt, um den Durchlauf des Verfahrens effizient koordinieren zu können. Ebenso gibt es für die Interaktion mit anderen beteiligten Behörden Fristen und Regelungen, um Reibungsverluste und Koordinationsaufwand in Grenzen zu halten.

Für Antragsteller auf Unternehmensseite sind je nach Routinegrad und Vorabstimmung des Verfahrens Informationen zu aktuellem Status, voraussichtlicher Dauer, Ansprechpartner und ggf. erforderlichen Nachlieferungen wichtig. Eine Nachlieferung im Fall eines unvollständigen oder formal nicht ganz korrekten Antrags verzögert in der Regel das Verfahren beträchtlich, so dass hier eine möglichst frühzeitige Benachrichtigung oder auch u.U. Unterstützung zur Vermeidung bedeutsam ist.

4.2 Prozessorientiertes Design für B2G

Aus den Erkenntnissen zu den verschiedenen Unternehmenslagen und ihren Merkmalen und Phasen ergeben sich verschiedenen Konsequenzen und Gestaltungsanforderungen an prozessorientiertes eGovernment.

Der Entscheidungscharakter eines Behördenkontaktes hat Einfluss darauf, in welchem Umfang und in welchem Automatisierungsgrad elektronische Unterstützung für sinnvoll erachtet wird.

Aus dem Routinegrad lässt sich ableiten, ob eine Automatisierung als sinnvoll erachtet wird, allerdings spielt hierfür auch der Entscheidungscharakter eine Rolle: Routineprozesse auf der Basis von Konditionalentscheidungen eignen sich für Workflows, Aushandlungsprozesse nicht. Die Komplexität einer Unternehmenslage wirkt sich dagegen darauf aus, wie spezialisiert die IT-Unterstützung auf die Bedürfnisse eines speziellen Behördenkontaktes sein sollte bzw. in welchem Umfang auch behördenähnliche Kontakte eine Rolle spielen und deren Abwicklung ggf. berücksichtigt werden sollte.

Die Unterscheidung der Phasen von Behördeninteraktionen zeigt, dass es bestimmt durch den zeitlich-logischen Ablauf aber auch durch die Hoheitsbereiche der beteiligten Partner verschiedene Koordinationssphären zu unterscheiden gibt.

4.3 Sphären der Behördeninteraktion

Unabhängig von der IT-Unterstützung ist die Abwicklung eines kooperativen B2G-Prozesses davon gekennzeichnet, dass es verschiedene Typen von Materialien oder Dokumenten gibt, die die Grundlage für die gemeinsame Arbeit darstellen.

Von Seiten der Behörden werden in der Regel ausführliche Informationsunterlagen bereitgestellt, die die Verfahrensschritte und die Abwicklung sowie die gesetzlichen Grundlagen einer Verwaltungsleistung dokumentieren. Diese Verfahrensregeln liefern im Sinne eines Modells zur **Koordination** Strukturen, Formate und Regeln für alle Phasen des Prozesses und reichen in unterschiedlichem Ausmaß in alle Koordinationssphären hinein.

Je nach Umfang der Ausgestaltung des Verfahrens werden bspw. im Fall von Umweltberichten auch bestimmte Teile dieser Regeln und Strukturen im gemeinsamen Diskurs entwickelt und ergänzt. In den meisten Fällen gibt es allerdings bereits vollständig ausgestaltete Verfahrensvorschriften, die von Seiten der Behörde als verbindlich vorgegeben werden, so dass der Kontext der gemeinsamen Koordination in erster Linie davon abhängt, ob es sich um ein Standard- bzw. Routineverfahren handelt oder um eine Einzelfallabwicklung.

Im ersten Fall werden Koordinations- und Verfahrensvorgaben allgemein verbindlich vorgegeben und nicht auf der Basis von Einzelfällen ausgehandelt. Die Standardisierung eines Verfahrens, die häufig auch mit einer gewissen Einschränkung des Entscheidungsspielraums einhergeht, ermöglicht in diesen Fällen auch eine workflowartige Unterstützung. Erfolgt die Abwicklung eines Verfahrens einzelfallbasiert, so kann das Koordinationsmodell lediglich einen Rahmen vorgeben, der jeweils an die konkreten Bedingungen angepasst wird.

Entlang der durch das Koordinationsmodell vorgegebenen Leitlinien wird bei der **Abwicklung** eines konkreten Behördenkontakts gemeinsames Material in dem Sinne bearbeitet, dass in einem wechselseitigen Austausch- und Feedback-Prozesses Antragsunterlagen und eine entsprechende Bewilligung hergestellt werden. Die Materialien und Unterlagen, die in diesem Verlauf genutzt und erstellt werden, sind in unterschiedlichem Umfang vertraulich zu behandeln bzw. dürfen in unterschiedlichem Umfang gelesen, bearbeitet und weitergegeben werden.

Systematisiert man diese beiden Merkmale für Kooperationsmaterialien, Zugänglichkeit und Manipulierbarkeit, so ergeben sich vier Klassen [Zerb00, S. 239f]:

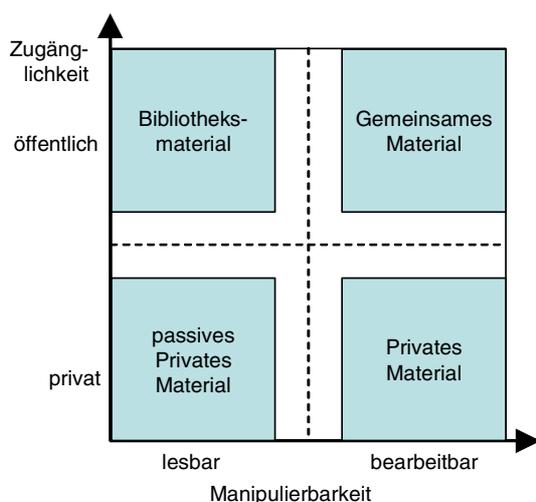


Abb. 2: Klassifikation von Kooperationsmaterialien; [Zerb00, S. 240]

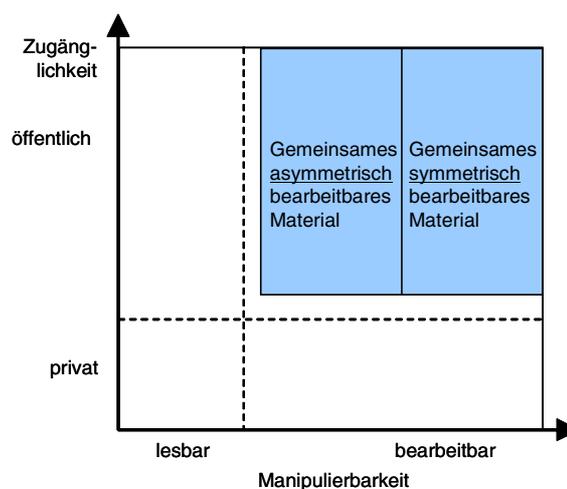


Abb. 3: Erweiterte Klassifikation von Kooperationsmaterialien

Passives privates Material: Dieser Materialtyp steht nur für den einzelnen Kooperationspartner zur Verfügung und kann nicht bearbeitet, sondern nur gelesen werden. Hierzu gehören Bücher, CD-ROMs, aber u.U. auch Daten, die als Ausgangsmaterial bspw. für einen Bericht genutzt, aber nicht verändert werden können/dürfen.

Privates Material: Privates Material kann im Gegensatz zu passivem privatem Material auch verändert werden, d.h. hierunter fallen Dokumente und Vorfassungen, die im Rahmen einer Antragsvorbereitung erstellt, aber nicht nach außen gegeben werden. Ebenso fallen hierunter Gutachten und Stellungnahmen, die nur behördenintern weitergegeben und für die Entscheidungsfindung genutzt werden.

Gemeinsames Material: Unter die Klasse gemeinsames Material fallen Dokumente und Unterlagen, die von allen Kooperationspartnern gelesen und manipuliert werden dürfen. Diese Klasse bezeichnet den Kooperationsbereich zwischen Unternehmen und Behörden, in dem Korrespon-

denz ausgetauscht, Nachforderungen gestellt und erläutert und Auskünfte erteilt werden. Auch bearbeitbare Formulare gehören zu dieser Klasse.

Bibliotheksmaterial: In diese Klasse fallen Unterlagen und Materialien, die öffentlich für den Nur-Lese-Zugriff bereitgestellt werden. Darunter fallen bspw. gesetzliche Grundlagen, Verfahrensbeschreibungen oder Informationen zu Ansprechpartnern.

Wichtig ist hier, dass die Zuordnung der Kooperationspartner zu den Gruppen ‚Privat‘ oder ‚Alle‘ bzw. ‚Öffentlich‘ auf zwei Ebenen betrachtet werden kann:

Im ersten Schritt trennt diese Klassifikation zwischen dem Unternehmen inklusive seiner Partner und Berater auf der einen und den betreffenden Behörden auf der anderen Seite. Dies ist eine Kooperationsbeziehung, die von Seiten der Behörden parallel zu unterschiedlichen Antragstellern gepflegt wird. Sie ist daher in Abb. 4 als 1:n Beziehung dargestellt.

Auf der nächst-detaillierteren Ebene kann auch jeweils innerhalb der beiden Parteien differenziert werden zwischen Privat und Alle. Vor allem auf Unternehmensseite ist diese Unterscheidung zwischen internen und externen Kooperationspartnern hinsichtlich der Vertraulichkeit von Informationen von Bedeutung.

Für die Behördeninteraktion und die dabei immer wieder als wichtig hervorgehobenen Statusinformationen wird eine Erweiterung dieser Klassifikation erforderlich: Im Rahmen der Bearbeitung des Vorgangs durch die Behörde werden Informationen generiert, die für die Behörde les- und bearbeitbar sind, auf die das betreffende Unternehmen jedoch Nur-Lese-Zugriff hat. Die Dimension der beiden beteiligten Partner muss hier also ergänzt werden, was zu einer Aufspaltung des gemeinsamen Materials in gemeinsames symmetrisch bearbeitbares Material und gemeinsames asymmetrisch bearbeitbares Material führt (vgl. Abb. 3).

4.4 Kooperationsinfrastruktur für B2G-Kontakte

Aus der Integration der dargestellten Merkmale und Sphären von Behördeninteraktionen lassen sich Elemente einer Kooperationsinfrastruktur für B2G-Kontakte ableiten. Die Darstellung bezieht sich zunächst auf einen generischen Behördenkontakt, der nicht weiter hinsichtlich Entscheidungscharakter oder Routinegrad spezifiziert ist. Diese Merkmale werden im Anschluss berücksichtigt.

Im Rahmen der Wertschöpfungsprozesse eines Unternehmens werden Daten und Informationen generiert, die zunächst privates Material der Abteilung oder des Bereichs darstellen, der für ihre Administration und Verarbeitung verantwortlich ist. Im Fall der Vorbereitung eines Behördenkontaktes werden diese privaten Informationen ggf. von der für die Berichterstattung oder An-

tragstellung zuständigen Abteilung angefordert und stellen dort je nach Art der Informationen privates oder passives privates Material dar.

Für die Vorbereitung eines Antrags werden meist aber auch öffentlich bereitgestellte Informationen genutzt, die Auskunft über Ansprechpartner und Abläufe gehen. In vielen Fällen werden auch Formulare oder spezielle (Software-)Werkzeuge für die Vorbereitung von Anträgen angeboten, die als gemeinsames Material zu bezeichnen sind.

Für die Gestaltung einer entsprechenden Kooperationsinfrastruktur ergibt sich daraus zum einen die Anforderung, die Phasen eines B2G-Kontaktes abzubilden, dabei aber andererseits die Abgrenzung der verschiedenen Koordinationssphären zu berücksichtigen. Um die Anforderungen an eine vertrauliche Behandlung von Informationen entsprechend der jeweiligen Phase zu erfüllen, ist darüber hinaus auch das Konzept der erweiterten Klassifikation von Kooperationsmaterialien anzuwenden.

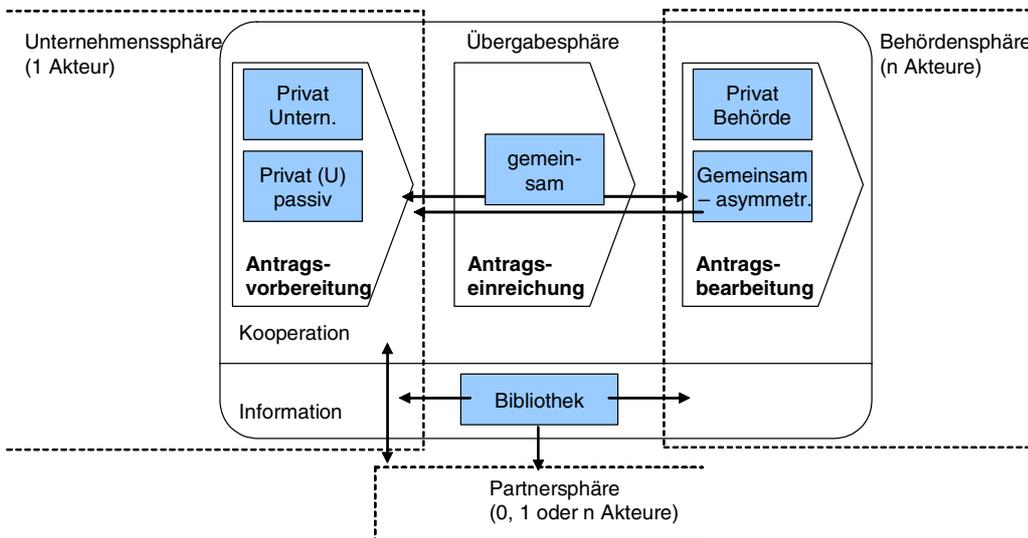


Abb. 4: Kooperationsinfrastruktur für B2G-Kontakte

Die resultierende Kooperationsinfrastruktur (Abb. 4) verbindet diese Konzepte in einem Raummodell, das sich im Wesentlichen in drei Sphären untergliedert, die Unternehmenssphäre (links), die Behördensphäre (rechts) sowie eine gemeinsame Kooperations-sphäre (mitte). Je nach Ausprägung der Komplexität der abzubildenden Unternehmenslage ist darüber hinaus auch eine Partnersphäre zu berücksichtigen. Eine weitere grundsätzliche Unterscheidung trennt einen reinen Informationsraum von einem Kooperationsraum. Im Informationsraum werden im Sinne des Bibliothek-Prinzips öffentliche Dokumente bereitgestellt, die zu Ablauf, gesetzlichen Grundlagen, Ansprechpartnern aber auch zu Aspekten einer erweiterten Unternehmenslage Auskunft geben.

Innerhalb der Koordinationssphäre des Unternehmens bzw. der Behörde(n) herrschen private Materialien vor. Aus dem Bereich der gemeinsamen Materialien werden in der Phase der Antragsvorbereitung bspw. Formulare importiert und in einem geschützten Raum von internen oder auch externen Partnern bearbeitet. Diese Fassungen der Antragsunterlagen sind vertrauliche Vorfassungen, die erst nach einer abschließenden Qualitätskontrolle in eine endgültige Version überführt werden. Die Kooperationsplattform stellt in dieser Phase Formulare und Antragsunterlagen mit umfangreichen Importschnittstellen zur Verfügung. Darüber hinaus sind die verschiedenen Elemente der Antragsunterlagen intelligent miteinander verlinkt, so dass neben Fristerinnerungen auch ein Änderungs- und Versionsmanagement möglich ist. Wird ein Vorgang von einem Unternehmen bereits zum wiederholten Mal durchgeführt, kann es auf eine private Bibliothek von vorangegangenen Antragsunterlagen zugreifen und wieder verwendbare Elemente importieren.

Im Moment der Antragseinreichung werden die Antragsunterlagen in gemeinsames Material überführt, das nun von Behördenseite gelesen und bearbeitet werden kann. Diese Übergabe wird durch entsprechende Transaktionsprotokolle (Zeitstempel, Vorgangsnummer) dokumentiert. Für die Übergabe stellt die Kooperationsplattform Upload-Mechanismen, Konvertierungswerkzeuge und Plausibilitätskontrollen zur Verfügung. Rückmeldungen von Seiten der Behörde bezüglich der Vollständigkeit und Eignung der eingereichten Unterlagen oder auch Nachforderungen werden hier ebenso eingestellt und veranlassen das Unternehmen ggf. dazu, noch einmal Unterlagen vorzubereiten und einzureichen.

Für die Begutachtung der Antragsunterlagen importiert die Behörde die Dokumente in interne Vorgangsbearbeitungssysteme und damit in ihre Koordinationssphäre. Hier werden Arbeitspakete koordiniert und die Bearbeitung gemäß betreffender Verfahrensvorschriften durchgeführt. An festgelegten Meilensteinen werden Statusinformationen generiert, die für das Unternehmen im Nur-Lese-Zugriff zu Verfügung stehen. Werden Mängel an den Antragsunterlagen festgestellt, die durch eine Nachbesserung seitens des Unternehmens behoben werden können, erhält das Unternehmen eine entsprechende Meldung in die gemeinsame Kooperationsphäre.

Im Fall routinemäßiger Konditionalentscheidungen können zahlreiche Teile des Verfahrens automatisiert unterstützt werden. Auf Seiten des Unternehmens ist es u.U. lohnenswert Standardschnittstellen zu entwickeln, die die erforderlichen Daten automatisch in die betreffenden Formulare importieren. Nur eine letzte Qualitätskontrolle und ggf. ergänzende Kommentare müssen nun noch händisch hinzugefügt werden oder können evt. auch aus archivierten erfolgreichen

Anträgen oder Berichten wieder verwendet werden. Nach Freigabe und Unterschrift können die Unterlagen eingereicht werden. Auch auf Behördenseite können in diesen Fällen einige Schritte automatisch durch Plausibilitätskontrollen und Checklisten unterstützt werden.

Anders sieht die Unterstützungssituation im Fall von einzelprojektartigen Behördenkontakten aus. Hier ist für Unternehmen in erster Linie die Unterstützung durch Formulare und Ablaufdokumentation wichtig sowie Illustration von Lösungswegen anhand von Beispieldokumenten und Ausfüllhilfen. Die flexible Integration von internen und externen Projektmitgliedern muss ebenso gewährleistet sein wie ein verlässliches Versionsmanagement und eine Koordinationsunterstützung. In diesen Fällen wird die elektronische Kooperation in großem Umfang unterstützt durch persönliche Interaktion von Ansprechpartnern auf Unternehmens- und Behörden-seite. Dafür sind Kontaktdaten zu Ansprechpartnern wichtig und ggf. auch die Kontaktvermittlung zu spezialisierten Dienstleistern dieses Themenbereichs.

5 Zusammenfassung

Das Konzept des prozessorientierten eGovernment hat sich als sehr geeignet erwiesen, den Unterstützungsbedarf von Unternehmen in verschiedenen Unternehmenslagen zu untersuchen. Mit Merkmalen wie Entscheidungscharakter oder Routinegrad sind Unternehmenslagen weiter zu präzisieren, um den verschiedenen Gegebenheiten von Behördenkontakten gerecht zu werden.

Die Zurückhaltung vieler Unternehmen gegenüber eGovernment zeigt sich augenblicklich in der eher vorsichtigen Akzeptanz bestehender B2G-Angebote. Angesichts des hohen Verbreitungsgrades von eBusiness-Lösungen für die unterschiedlichsten Geschäftsbereiche und Branchen überrascht dieser Befund zunächst. Bislang werden allerdings die verschiedenen Phasen und Koordinationssphären im Laufe einer Behördeninteraktion nicht oder nur unzureichend in eGovernment-Lösungen berücksichtigt. Dieser Mangel führt dazu, dass vorhandene Lösungen sich entweder auf die Phase der Übergabe von Antragsdokumenten beschränken und der erzielte Effekt überschaubar bleibt, oder dass vorschnelle Automatisierungsversuche auf Unternehmensseite Befürchtungen hinsichtlich der Vertraulichkeit und Beherrschbarkeit der eigenen Daten hervorrufen.

Die dargestellten Merkmale von Unternehmenslagen sind das Ergebnis einer explorativen Untersuchung und bieten eine Hilfestellung zur Beschreibung des eGovernment-Unterstützungsbedarfs von Unternehmen. Die Konkretisierung der Merkmale für verschiedene Unternehmens-

typen (Branche, Größe) sowie die Klassifikation von Verwaltungsleistungen entsprechend des resultierenden Schemas sind Aufgaben zukünftiger Forschung. Auf der Basis dieser Ergebnisse kann das entwickelte Schema einer Kooperationsplattform weiter ausgearbeitet und schließlich als Grundlage für prozessorientiertes eGovernment-Design genutzt werden.

Literaturverzeichnis

- [Abfa06] Abfallmanagement AG: ZEDAL - Leistungsumfang.
<http://www.zedal.de/index.php?id=60>, Abruf am: 06.11.2006.
- [BeBu03] BearingPoint; Bundesministerium des Innern (2003). Elektronische Abwicklung von Geschäftsprozessen der Verwaltung mit der Wirtschaft. o.O.
- [BrGi02] Brücher, H.; Gisler, M.: E-Government – von den Grundlagen zur Anwendung. In: HMD (2002) 226, S. 5-19.
- [Enge00] Engel, A.: Einführung der IT-gestützten Vorgangsbearbeitung - Strategien für die öffentliche Verwaltung. In: Reiner mann, H. (Hrsg.): Regieren und Verwalten im Informationszeitalter. R.v.Decker, Heidelberg 2000, S. 391-413.
- [Fuch88] Fuchs, E.: Entscheidungen in großen Verwaltungen bei unvollständiger Information. In: Grimmer, K. (Hrsg.): Informatisierung wenig strukturierter Verwaltungsaufgaben. Gesamthochschul-Bibliothek, Kassel 1988, S. 13-22.
- [InDD02] Institute of Electronic Business; DIHK; DE-CODA (Hrsg.): E-Government B2G: Anforderungen der deutschen Wirtschaft.
http://www.dihk.de/inhalt/themen/branchen/information_kommunikation/egovernment/egovernmentstudie.pdf, Abruf am: 13.08.
- [Inst02] Institut für Mittelstandsforschung: Mittelstand - Definition und Schlüsselzahlen.
<http://www.ifm-bonn.org/index.htm?/dienste/definition.htm>, Abruf am: 07.11.2006.
- [KnIn04] KnowLogy Solutions AG; Institute of Electronic Business (2004). Modellprojekt „Wirtschaftsorientierte Verwaltungsservices“. Berlin.

- [Lee05] Lee, D.H.-D.: Contextual IT Business Value and Barriers: an E-Government and E-Business Perspective. In: Sprague, R.H. (Hrsg.): 38th Hawaii International Conference on Systems Sciences 2005. Big Island, Hawaii.
- [Lenk02] Lenk, K.: Notwendige Revisionen des Geschäftsprozessdenkens. In: Wimmer, M. (Hrsg.): eGov-Day 2002. Wien.
- [LeTr01] Lenk, K.; Traummüller, R.: Electronic Government - ein Wegweiser. In: Computer kommunikativ 4 (2001), S. 15-18.
- [o.A.02] o.A. (2002). Survey on eGovernment Services to Enterprises: European Commission.
- [o.A.06a] o.A.: eBusiness gewinnt für Unternehmen an Bedeutung. <http://www.pressebox.de/pressemeldungen/bitkom-ev/boxid-66797.html>, Abruf am: 12.07.2006.
- [o.A.06b] o.A.: eGovernment kann Standortfaktor werden. <http://www.netzeitung.de/internet/416159.html>, Abruf am: 12.07.2006.
- [PiRe85] Picot, A.; Reichwald, R.: Bürokommunikation - Leitsätze für den Anwender. 2. Aufl., CW-Publ., München 1985.
- [SchHä01] Schubert, P.; Häusler, U.: E-Government meets e-business: a portal site for startup companies in switzerland. In: (Hrsg.): Hawai'i International Conference On System Sciences, HICSS 2001. Maui. Hawaii.
- [Tsch96] Tschamler, H.: Wissenschaftstheorie: Eine Einführung für Pädagogen. Klinkhardt, Bad Heilbrunn 1996.
- [WoKr05] Wolf, P.; Krcmar, H.: Prozessorientierte Wirtschaftlichkeitsuntersuchung für E-Government. In: Wirtschaftsinformatik 47 (2005) 5, S. 337-346.
- [Zerb00] Zerbe, S.: Globale Teams. Gabler, Wiesbaden 2000.

Zahlungsbereitschaft für elektronische Signaturen

Heiko Roßnagel, Oliver Hinz

Fachbereich Wirtschaftswissenschaften
Johann Wolfgang Goethe-Universität Frankfurt
60054 Frankfurt am Main
heiko.rossnagel@m-lehrstuhl.de
oliver.hinz@wiwi.uni-frankfurt.de

Abstract

Eine zentrale Voraussetzung für E-Government ist die Möglichkeit, rechtssichere Willenserklärungen auch elektronisch abzugeben. Daher hat der Gesetzgeber mit dem Signaturgesetz und der Signaturverordnung bereits vor Jahren einen entsprechenden Rechtsrahmen geschaffen. Dennoch ist es bisher nicht gelungen, einen funktionierenden Markt für Signaturanwendungen bzw. Zertifizierungsdienstleistungen zu schaffen. Ein Grund für die schleppende Verbreitung von elektronischen Signaturen und somit ein zentrales Hemmnis der Verwaltungsmodernisierung dürfte in den hohen Preisen für Zertifizierungsdienstleistungen zu finden sein. Mit diesem Beitrag versuchen wir auf Basis einer empirischen Studie, die Zahlungsbereitschaft für elektronische Signaturen zu ermitteln und sie den aktuellen Marktpreisen gegenüberzustellen.

1 Einleitung

Mit dem Übergang von der papierbasierten Akte hin zur elektronischen Akte ergeben sich im E-Government enorme finanzielle Einsparpotentiale [KPБу01]. Traditionelle Vorgangs- und Dokumentenbearbeitungen werden durch die Integration von Workflow- und Dokumentenmanagementsystemen zur elektronischen Abwicklung von Verwaltungsprozessen verdrängt. Um aber das volle Einsparpotential von E-Government abzurufen, ist es unumgänglich, dass Willens- und Wissenserklärungen rechtssicher auf elektronischem Weg abgegeben und dokumentiert werden können [Roßn03a]. Hierfür hat der Gesetzgeber mit Signaturgesetz [Deut01a] und Signaturverordnung [Deut01b] einen geeigneten Rechtsrahmen

geschaffen. Mit Hilfe qualifizierter elektronischer Signaturen ist es möglich, elektronische Willenserklärungen abzugeben, die in ihrer Rechtswirkung handschriftlich unterschriebenen Erklärungen gleichgestellt sind [Roßn03a].

Seit Jahren sind nun die technischen und rechtlichen Rahmenbedingungen für qualifizierte elektronische Signaturen vorhanden. Dennoch ist es bisher nicht gelungen, einen funktionierenden Markt für Signaturanwendungen bzw. Zertifizierungsdienstleistungen zu schaffen [FrRo05]. Bisher wurden in Deutschland gerade einmal etwa 30.000 Zertifikate für qualifizierte elektronische Signaturen ausgestellt [Siet04]. Da die qualifizierte elektronische Signatur Netzeffekten unterliegt, steigt ihr Nutzen überproportional mit der Anzahl der Nutzer. Die qualifizierte elektronische Signatur wird sich nur dann durchsetzen, wenn eine hinreichende Anzahl der Nutzer sie regelmäßig verwendet [KPBU01]. Bleibt dies aus, so entsteht ein ungünstiges Kosten-/Nutzenverhältnis [LiRo05], und es entsteht ein Verhinderungskreislauf [Lamb03]. Um die möglichen Einsparpotentiale im E-Government zu realisieren, sind aber ein hoher Verbreitungsgrad von Signaturerstellungseinheiten und eine Partizipation der einzelnen Bürger notwendig [KPBU01]. Daher wäre es wünschenswert, dieses Hemmnis der Verwaltungsmodernisierung zu beseitigen.

Als ein Grund für die bisher geringe Verbreitung von qualifizierten elektronischen Signaturen werden häufig die hohen Preise für Zertifizierungsdienstleistungen genannt [LiRo05] [FrRo05] [Schr05] [HoRo04] [Büge06] [Leis06] [BüEK04]. Jedoch sind uns keine Untersuchungen zur Zahlungsbereitschaft von potenziellen Nutzern der elektronischen Signatur bekannt. Diese Lücke möchten wir mit diesem Beitrag schließen.

Hierfür werden zunächst in Abschnitt 2 die Marktteilnehmer und ihre Bepreisungsmodelle beschrieben. In Abschnitt 3 wird dann die Methodik unserer empirischen Studie dargestellt. Die Ergebnisse der Studie werden dann in Abschnitt 4 vorgestellt und in Abschnitt 5 ausführlich diskutiert.

2 Aktuelle Marktsituation

2.1 Marktteilnehmer

Derzeit gibt es in Deutschland 24 akkreditierte Zertifizierungsdiensteanbieter, allerdings bieten viele nur branchenspezifische Lösungen an. Unter den Anbietern sind neun

Steuerberaterkammern, sechs Rechtsanwaltskammern, die Bundesnotarkammer und die Wirtschaftsprüfkammer. Bei der folgenden Analyse liegt der Fokus auf den großen Zertifizierungsdiensteanbietern. Daher werden in Tabelle 1 nur die nicht auf spezielle Berufsgruppen spezialisierten Zertifizierungsdienstleister berücksichtigt

	Privatanwender	Unternehmen / Behörden
AuthentiDate International AG	0	✓
Deutsche Post Signtrust GmbH	✓	✓
D-Trust GmbH	✓	✓
TC TrustCenter AG	0	✓
T-TeleSec	✓	✓

Tabelle 1: Das Angebot an personalisierten Chipkarten auf dem Markt

Es werden sowohl für Privatanwender als auch für Unternehmen und Behörden personalisierte Chipkarten angeboten. Die AuthentiDate International AG hat sich allerdings mit ihren Komplettlösungen auf Geschäftskunden spezialisiert. Die TC TrustCenter AG bietet für Privatanwender nur ein Demo-Zertifikat an. Der Fokus der Unternehmen liegt auf der Gewinnung von Geschäftskunden [LiRo05].

2.2 Preisstrategie

Tabelle 2 zeigt die aktuellen Preise für das Ausstellen einer signaturgesetzkonformen personalisierten Chipkarte zur Erzeugung qualifizierter elektronischer Signaturen.

	Ausstellung des Zertifikates	Jährliche Grundgebühr	Summe für 2-jährige Nutzung
D-Trust GmbH	41 €	29 €	99 €
Deutsche Post Signtrust GmbH	0 €	39 €	78 €
TC TrustCenter AG	8 €	62 €	132 €¹
T-TeleSec	23,57 €	42,95 €	109,47 €

Tabelle 2: Preise für eine personalisierte Chipkarte²

Bei allen Unternehmen müssen die Kunden eine jährliche Grundgebühr unabhängig von der Nutzungsintensität der Karte entrichten. Ein zusätzliches Entgelt für eine Zertifikatsüberprüfung oder ähnliches fällt nicht an [LiRo05].

Entscheidet sich ein Kunde für die Anschaffung einer personalisierten Chipkarte, kann er sich nicht zwischen verschiedenen Tarifen entscheiden. Dabei gibt es unterschiedliche

¹ Dieses Angebot richtet sich ausschließlich an Geschäftskunden.

² Alle Preise verstehen sich zuzüglich MwSt.

Marktsegmente, die ganz unterschiedliche Verwendungsprofile haben. Dies wurde bei der Preisfestsetzung überhaupt nicht berücksichtigt. Vor allem fehlt ein attraktives Einstiegsangebot, um ganz gezielt neue Privatkunden zu gewinnen [Roßn06]. In [LiRo05] wurde daher ein alternatives Preissystem vorgestellt, in dem eine Aufteilung der Kosten zwischen der Signaturerstellung und der Signaturprüfung über die Erhebung einer geringeren Grundgebühr für den Signierenden und über die transaktionsbasierte Abrechnung der Signaturprüfung stattfindet. Bis heute hat aber keiner der Zertifizierungsdienstleister sein Angebot angepasst.

3 Methodik der Untersuchung

3.1 Ermittlung der Zahlungsbereitschaften

Für die optimale Preispolitik eines Unternehmens ist es unerlässlich, möglichst genau die Zahlungsbereitschaften von Konsumenten für Produkte zu kennen. Die Zahlungsbereitschaft ist dabei als der Preis definiert, den ein Konsument maximal zu zahlen bereit ist [SkRe99]. Dabei gilt als Annahme, dass die Zahlungsbereitschaft für ein Produkt positiv mit der Präferenz des Konsumenten für dieses Produkt korreliert ist. Zur Ermittlung von Präferenzen gilt die Conjoint-Analyse als das bekannteste Verfahren [BaBr04]. Allen Varianten der Conjoint-Analyse liegt die Annahme zugrunde, dass das Produkt, das der Konsument am meisten bevorzugt, ihm auch gleichzeitig beim Konsum den größten Nutzen stiftet. Zusätzlich wird angenommen, dass der Gesamtnutzen eines Produkts sich aus der Summe der Teilnutzen der einzelnen Produkteigenschaften ergibt.

Mittlerweile gibt es eine Vielzahl von Conjoint-Varianten und –Modifikationen wie z. B. die Adaptive Conjoint- [John87], die Limit-Conjoint- [VoHa98] und die Choice-Based Conjoint (CBC)-Analyse [LoWo83].

In der traditionellen Conjoint-Analyse bringen die Konsumenten unterschiedliche Produkte in eine Rangfolge oder bewerten diese anhand einer Skala. Allen Conjoint-Varianten ist gemeinsam, dass die Bewertungen der Befragten sich auf das ganzheitliche Produkt beziehen. Durch diese Vorgehensweise entspricht die Conjoint- Analyse in hohem Maße dem tatsächlichen Bewertungsprozess einer realen Kaufsituation, in der der Konsument ebenfalls mit ganzheitlichen Produkten konfrontiert ist. Bei der Auswertung werden diese Bewertungen

allerdings auf die Eigenschaften und deren Ausprägungen umgerechnet. Dieser dekompositionelle Ansatz beruht auf der Grundannahme, dass sich der Gesamtnutzen eines Produkts linear additiv aus den Nutzenbeiträgen der einzelnen Eigenschaftsausprägungen ergibt. Der resultierende Gesamtnutzen spiegelt so die unterstellten Bewertungen der Konsumenten intervallskaliert wider und so können Nutzenunterschiede zwischen verschiedenen Eigenschaftsausprägungen und Produkten ermittelt werden. Es werden Nutzen für Eigenschaftsausprägungen geschätzt, die dann zu Gesamtnutzenwerten aggregiert werden. Die Teilnutzenwerte werden so geschätzt, dass Gesamtnutzenwerte die tatsächliche Präferenzrangfolge oder Auswahlentscheidung bestmöglich abbilden. [Srin82] zeigte bereits, dass diese Nutzenunterschiede durch eine Normierung der Abstände in Geldeinheiten umgerechnet werden können, so dass dadurch die Zahlungsbereitschaft für die Veränderung einer Eigenschaftsausprägung errechnet werden kann. Allerdings ist es nicht möglich, die Zahlungsbereitschaft für ein Produkt zu ermitteln, da die Nutzenfunktion intervallskaliert ist und über keinen Nullpunkt verfügt. [Srin82] zeigte also eine Möglichkeit auf, wie Veränderungen des Produktdesigns auf die Zahlungsbereitschaft in monetären Einheiten wirkt, aber letztlich konnte keine Aussage darüber getroffen werden, ob ein Konsument ein vorgegebenes Produkt kauft oder nicht.

[GeST06] greifen die Idee von [Srin82] wieder auf und benutzen die Choice-Based Conjoint-Analyse, um diese Unzulänglichkeit zu heilen, da bei der CBC die Bestimmung eines Nullpunkts möglich ist. CBC gilt mittlerweile als die am häufigsten eingesetzte Variante der Conjoint-Analyse [HaSa02]. Bei der CBC werden den Konsumenten eine Menge an Produkten – so genannte Choice Sets – vorgelegt, und diese müssen sich dann für eines der Produkte entscheiden. Optional können die Choice-Sets auch eine Nichtkauf-Option enthalten, die der Konsument wählen kann, wenn ihm keines der Produkte zusagt. Auf diese Weise ist es möglich, eine Aussage darüber zu machen, welche Produkte der Konsument kaufen würde und welche nicht. Allerdings enthalten diese Auswahlentscheidungen weniger Informationen als die Bewertungen anhand einer Skala. Daher konnten die Nutzenfunktionen der CBC-Analyse lange Zeit nicht konsumentenindividuell geschätzt werden, sondern mussten aggregiert geschätzt werden [LoWo83]. Erst mit dem Latent Class-Verfahren war es möglich, die Heterogenität der Präferenzen der Konsumenten abzubilden und damit Schätzungen auf Segment-Ebene durchzuführen.

Dabei wird zur Schätzung der individuellen Nutzenfunktion die Informationen, die die Wahl der Nichtkauf-Option ermöglichen, zur Bestimmung des Nullpunkts in der intervallskalierten Nutzenfunktion des Konsumenten herangezogen, so dass dadurch ratioskalierte Zahlungsbereitschaften errechnet werden können. Als weiterer Vorteil der CBC gilt, dass die Auswahl-situation im Vergleich zur traditionellen Conjoint-Analyse realistischer dargestellt wird und so eine höhere externe Validität erreicht wird [LoWo83].

Daher halten wir das Verfahren, Zahlungsbereitschaften mithilfe der Choice-Based Conjoint zu bestimmen, für eine geeignete Methode, um die Zahlungsbereitschaften für elektronische Signaturen zu ermitteln. Genaue Informationen zur Methode finden sich in [GeST06]. Wir wenden dieses Verfahren auf unsere Fragestellung entsprechend in einer empirischen Studie an, die im folgenden Kapitel beschrieben wird.

3.2 Aufbau der empirischen Studie

Um bestimmte Segmente und deren Zahlungsbereitschaft zu ermitteln, entwickelten wir einen Fragebogen, der neben demographischen und psychographischen Daten insgesamt 12 Choice-Sets plus ein Holdout-Set enthielt. Jedes Choice-Set enthielt drei Wahlmöglichkeiten und eine Nichtkauf-Option. Als Eigenschaften wurden das Signaturniveau, die Form des Gerätes, auf der die Signatur gespeichert werden kann, und die Einsatzmöglichkeit der Signatur gewählt. Mit der ersten Eigenschaft, dem Signaturniveau, sollte überprüft werden, ob für die unterschiedlichen Sicherheitsstufen und die damit verknüpften unterschiedlichen Rechtsfolgen [Roßn02] auch eine unterschiedliche Zahlungsbereitschaft vorhanden ist. Insbesondere da in Verbindung mit qualifizierten und akkreditierten Signaturen häufig von einer Überregulierung gesprochen wird [Lamb03], soll diese Untersuchung zeigen, ob potenzielle Kunden bereit sind, für die gegenüber der fortgeschrittenen Signatur [Roßn03b] höheren Anforderungen und Rechtsfolgen auch höhere Preise zu bezahlen.

Die zweite Eigenschaft beschreibt die Form des Gerätes. Auch hier standen drei unterschiedliche Geräteformen zur Wahl. Zum einen die bisher erhältliche Lösung mit einem an den PC anzuschließenden Chipkartenleser und der dazugehörigen Signaturkarte, die das geringste Maß an Mobilität ermöglicht. Auf der anderen Seite wurde eine Geräteform angeboten, bei der die Signaturerstellung mit Hilfe des eigenen Mobiltelefons erfolgen kann [Roßn04] und somit das höchste Maß an Mobilität erlaubt. Dazwischen befindet sich die dritte Geräteform in Form eines USB Sticks. Mit der Eigenschaft der Geräteform sollte überprüft

werden, wie wichtig den potenziellen Kunden die mobilen Einsatzmöglichkeiten sind und wie diese bepreist werden können.

Da bisher kaum Anwendungen für die elektronische Signatur außerhalb des E-Government vorhanden sind, dient die dritte Eigenschaft dazu, zu ermitteln, inwieweit sich dieses begrenzte Einsatzgebiet negativ auf die Zahlungsbereitschaft der potenziellen Kunden auswirkt.

Tabelle 3 enthält eine detaillierte Auflistung der Eigenschaften und der Eigenschaftsausprägung. Aus diesen Vorgaben ermittelten wir mit Sawtooth ein effizientes Design, das wir schließlich in unser Online-Fragebogen-Tool integrierten. Bei Sawtooth handelt es sich um eines der meist genutzten Tools zur Erstellung von optimalen Conjoint Designs (siehe [Sawt2006] für weitere Informationen).

Eigenschaft	Eigenschaftsausprägung	Beschreibung der Eigenschaft
Signaturniveau	<ul style="list-style-type: none"> • Fortgeschrittene Signatur • Qualifizierte Signatur • Akkreditierte Signatur 	Sicherheitsstufe der eingesetzten Signatur
Geräteform	<ul style="list-style-type: none"> • Mobiltelefon mit Signaturfunktionalität • PC mit Chipkartenleser • Kartenleser als USB Stick 	Bevorzugte Geräteform auf der sich die elektronische Signatur befinden soll
Einsatzgebiet	<ul style="list-style-type: none"> • Nur E-Government • E-Government und E-Commerce • E-Government, E-Commerce und private Kommunikation 	Gültiges Einsatzgebiet
Preis	<ul style="list-style-type: none"> • 5 EUR pro Jahr • 10 EUR pro Jahr • 15 EUR pro Jahr • 25 EUR pro Jahr • 50 EUR pro Jahr 	Preis für die elektronische Signatur

Tabelle 3: Eigenschaften und Eigenschaftsausprägungen der Choice-Sets

Abb. 1 zeigt eine typische Auswahlentscheidung in Form eines Choice-Sets, vor die die Teilnehmer der Studie gestellt wurden. Der Fragebogen war vom 14.02.2006 bis zum 02.05.2006 im Internet zugänglich und wurde über verschiedene Kanäle beworben. Die Teilnehmer der Studie hatten Aussicht auf den Gewinn eines DVD-Players und mehrerer Amazon-Gutscheine.

Bitte wählen Sie die von Ihnen bevorzugte Kombination				
Signaturniveau	Fortgeschrittene Signatur	Qualifizierte Signatur	Akkreditierte Signatur	Ich würde keine dieser Signaturen wählen
Geräteform	Kartenleser als USB-Stick	Mobiltelefon mit Signaturfunktionalität	PC mit Chipkartenleser	
Einsatzgebiet	E-Government und E-Commerce	Nur E-Government	E-Government, E-Commerce und private Kommunikation	
Preis	25 Euro pro Jahr	50 Euro pro Jahr	5 Euro pro Jahr	
	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abb. 1: Auswahlentscheidung im Fragebogen

Mithilfe der Daten konnten wir sowohl verschiedene Segmente ermitteln als auch eine Schätzung der Zahlungsbereitschaften durchführen. Diese Ergebnisse sind im folgenden Kapitel wiedergegeben. Ferner ist es möglich, anhand der demographischen und psychographischen Daten Aussagen darüber zu machen, welche Charaktereigenschaften besonders auf die Zahlungsbereitschaft wirken, und so die Segmente genauer zu beschreiben. Diese Ergebnisse sind allerdings nicht Teil dieser Arbeit.

4 Ergebnisse

4.1 Deskriptive Ergebnisse

Durch die Umfrage erhielten wir 238 vollständige Antworten. Auf eine Auswertung der psychographischen Daten, die solche Faktoren wie Vertrauen, Fortschrittlichkeit und Sicherheitsbedenken beinhalten, sei an dieser Stelle verzichtet, da sie sich vom Thema dieser Arbeit zu weit entfernen. Insgesamt war der Großteil der Teilnehmer überdurchschnittlich gebildet (91,9% mit Abitur oder Hochschulabschluss) und männlich (74,4%).

Das Durchschnittsalter lag bei 28,5 Jahren, und ca. 47,9% der Teilnehmer waren Studenten, so dass mit ziemlicher Sicherheit gesagt werden kann, dass das vorliegende Sample nicht repräsentativ ist. Allerdings halten wir die Ergebnisse dennoch für wertvoll, da gerade die in unseren Daten vorherrschende Gruppe durchaus als Interessent für elektronische Signaturen gilt

[KPBU01]. Hochgerechnet auf die Gesamtbevölkerung sollte das Interesse deutlich niedriger liegen als in unserem Sample.

Insgesamt waren die Teilnehmer auch recht aktiv beim E-Mail-Verkehr. So gaben 65,1% an, dass sie mehr als 5 E-Mails am Tag versenden, so dass wir davon ausgehen können, dass die Teilnehmer zumindest potenzielles Interesse an einer elektronischen Signatur haben könnten. Allerdings wiegen Konsumenten den Nutzen und die Kosten von Produkten bzw. Dienstleistungen gegeneinander ab. Dieser Überlegung wird besonders bei der Choice-Based Conjoint Rechnung getragen, da dem Konsumenten drei Auswahlentscheidungen mit den jeweiligen Kosten präsentiert werden und zusätzlich die Möglichkeit eines Nichtkaufs besteht. Mit Hilfe des Verfahrens von [GeST06] kann dann auf Segmente und deren Zahlungsbereitschaft geschlossen werden. Zusätzlich ist es möglich, optimale Produkte für die Segmente zu bestimmen und auf diese Art und Weise die Konsumentenrente besser abzuschöpfen bzw. die Allokationseffizienz des Marktes zu erhöhen.

Bei der vorliegenden Umfrage wählten 24 Teilnehmer bei allen vorgelegten Choice-Sets die Nichtkaufoption, so dass wir nur feststellen können, dass ihre Zahlungsbereitschaft unter 5 EUR liegt, wenn nicht sogar gegen 0 EUR tendiert. Auf der anderen Seite des Spektrums gibt es 55 Teilnehmer, die immer eine Wahl getroffen haben, so dass wir über deren maximale Zahlungsbereitschaft keine Aussage treffen können. Die genaue Zahlungsbereitschaft dieser beiden Segmente kann nach der Methode von [GeST06] nicht bestimmt werden, und daher müssen diese Teilnehmer aus dem Sample eliminiert werden, bevor wir die Segmentierung und die Bestimmung der Zahlungsbereitschaft durchführen können. Dadurch wurden auch weitgehend die Teilnehmer eliminiert, die lediglich durch die Gewinnchance zur Teilnahme bewogen wurden und den Fragebogen ohne weiteres Nachdenken beantwortet hatten.

4.2 Schätzung der Zahlungsbereitschaft

Die Schätzung der Parameter erfolgt mit der Latent Class Choice-Based Conjoint-Analyse, die Lösungen für zwei bis zehn Segmente betrachtet, um die optimale Anzahl der Segmente zu bestimmen. Als Entscheidungskriterien werden das Consistent Akaike Informationskriterium (CAIC) und die Entropy (Trennschärfe) der Segmente herangezogen. Darauf aufbauend werden die individuellen Parameter geschätzt.

Die Analyse ergibt eine optimale Anzahl von drei zusätzlichen Segmenten, die sich sowohl in ihrer Zahlungsbereitschaft als auch in ihrer Präferenz für die Eigenschaftsausprägungen unterscheiden. Die Prognosevalidität beträgt 67,9% und kann somit als gut eingestuft werden,

da sich bei einer rein zufälligen Auswahl bei drei Choice-Sets plus Nichtwahl-Option nur eine Wahrscheinlichkeit von 25% ergäbe [Hens00].

Die durchschnittliche Zahlungsbereitschaft aller Segmente für eine akkreditierte Signatur, die über einen PC-Kartenleser verfügbar gemacht wird und sowohl für E-Commerce als auch für E-Government und private Kommunikation genutzt werden kann, liegt bei 21,57 EUR pro Jahr. Damit liegt sie bei etwa der Hälfte des aktuell günstigsten angebotenen Preises.

Allen Segmenten ist gemein, dass sie eine Signatur bevorzugen, die für alle drei Bereiche, Electronic Government, Electronic Commerce als auch private Kommunikation, geeignet ist. Dieses Ergebnis war zu erwarten und kann als erstes – wenn auch schwaches – Gütekriterium der Schätzung angesehen werden. Auch der Preis hat für alle Segmente erwartungsgemäß einen negativen Nutzen, allerdings ist Segment 1 nicht so preissensitiv wie die beiden anderen. In den weiteren Eigenschaftsausprägungen unterscheiden sich die Segmente deutlich. Tabelle 4 zeigt die Nutzen für die einzelnen Segmente.

	akkreditiert	qualifiziert	fortgeschritten	Chipkartenleser	USB-Stick	Mobiltelefon	E-Gov	E-Gov & E-Com	E-Gov & E-Com & privat	Preis
Segment1	0,11	0,07	-0,17	0,36	0,42	-0,78	-0,53	-0,05	0,58	-0,04
Segment2	0,96	0,23	-1,19	-0,03	-0,52	0,55	-1,59	0,27	1,32	-0,10
Segment3	0,09	0,26	-0,34	0,04	-0,21	0,17	-0,67	-0,12	0,79	-0,28

Tabelle 4: Nutzen für die einzelnen Segmente

Aus dieser Tabelle kann für jedes Segment ein optimales Produkt hergeleitet werden. Segment 1 und Segment 2 bevorzugen eine akkreditierte Signatur, während Segment 3 eine qualifizierte Signatur bevorzugt. Segment 2 und Segment 3 haben den höchsten Nutzen für eine Signatur, die ein Mobilfunktelefon als Signaturerstellungseinheit verwendet, während Segment 1 einen USB-Stick als Trägermedium bevorzugt. Kann allerdings nur ein Produkt zur Verfügung gestellt werden, so sollte die elektronische Signatur mit einer Chipkarte als Trägermedium verwendet werden, da diese Eigenschaftsausprägung von allen Segmenten weitgehend akzeptiert wird.

Für alle Segmente ist es äußerst wichtig, dass die elektronische Signatur für alle Bereiche eingesetzt werden kann, wie an den erheblichen Nutzenunterschieden deutlich wird.

Aus den Nutzenwerten kann über den Nullpunkt in Form der Nichtwahl-Option die Zahlungsbereitschaft in monetären Werten errechnet werden. Tabelle 5 gibt für jedes Segment

die optimale Produktausgestaltung an und die durchschnittliche Zahlungsbereitschaft des Segments für dieses Produkt. Ebenso zeigt diese Tabelle auch, dass die Segmente ungefähr gleich groß sind.

	Optimales Produkt	Durchschnittliche Zahlungsbereitschaft	Segmentgröße
Segment 1	Akkreditiert, USB-Stick, universell einsetzbar	28,27 EUR / Jahr	55
Segment 2	Akkreditiert, Mobilfunktelefon, universell einsetzbar	33,62 EUR / Jahr	53
Segment 3	Qualifiziert, Mobilfunktelefon, universell einsetzbar	8,63 EUR / Jahr	51

Tabelle 5: Segment-optimale Produkte und die entsprechende Zahlungsbereitschaft der Segmente

Deutlich wird, dass es offenbar ein sehr preissensitives Segment gibt, dem eine qualifizierte Signatur genügt, aber auch nur eine durchschnittliche Zahlungsbereitschaft von 8,63 EUR pro Jahr hat. Die beiden anderen Segmente 1 und 2 unterscheiden sich lediglich in der bevorzugten Form des Speichermediums.

Zurzeit gibt es allerdings kaum Anwendungen für elektronische Signaturen außerhalb des Electronic Government. Untersucht man demnach die Zahlungsbereitschaften für elektronische Signaturen, die ausschließlich für das E-Government geeignet sind, so wird schnell deutlich, dass die durchschnittliche Zahlungsbereitschaft für alle drei Segmente unter 5 EUR pro Jahr liegt. Daher folgt, dass E-Government als Signaturanwendung alleine wohl kaum zu einem flächendeckenden Einsatz der elektronischen Signatur führen wird.

5 Diskussion

In unserer Studie haben wir gezeigt, dass für akkreditierte und qualifizierte elektronische Signaturen durchaus eine nennenswerte Zahlungsbereitschaft vorhanden ist. Allerdings besteht eine Diskrepanz zwischen den in Kapitel 2.2 vorgestellten Preisen und der ermittelten Zahlungsbereitschaft. Die Preise der Zertifizierungsdienstleister orientieren sich aber bisher ausschließlich an den eigenen Kosten [Leis06].

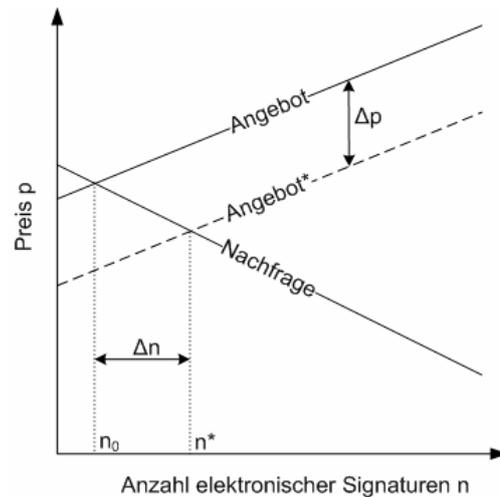


Abb. 2: Angebots- und Nachfragekurve

Wie Abbildung 2 zeigt, treffen sich Angebots- und Nachfragekurve zurzeit bei einem sehr niedrigen n_0 . Eine Parallelverschiebung der Angebotskurve nach *Angebot** durch eine Kostensenkung führt zu einer Erhöhung der Anzahl von Signaturerstellungseinheiten um $\Delta n = n^* - n_0$.

Der Argumentation von [SMRG05] folgend ergebe sich durch diese Kostenreduktion und die damit verbundene Nachfragesteigerung durch Netzeffekte auf der Anwenderseite auch gleichzeitig ein größerer Gesamtnutzen, der wiederum zu niedrigeren Kosten führen kann (siehe Abb. 3). Wir halten es daher speziell für das E-Government für unabdingbar, dass eine Preissenkung notwendig ist und eine Subventionierungsstrategie erforderlich ist, um dieses Hemmnis der Verwaltungsmodernisierung zu beseitigen. Diese Subventionierung könnte auf lange Sicht zu enormen Einsparpotentialen im E-Government führen [KPBU01] und letztendlich auch einen wirtschaftlichen Betrieb der Zertifizierungsdienstleister ermöglichen [LiRo05]. Häufig wird gefordert, dass daher der Staat in Vorleistung gehen soll, um die kritische Masse an Benutzern zu erreichen [VonB06] [Leis06]. Auf der anderen Seite kann man auch aus betriebswissenschaftlicher Sicht argumentieren, dass die Unternehmen selbst höchstes Interesse daran haben sollten, aus eigener Kraft profitabel zu werden [LiRo05].

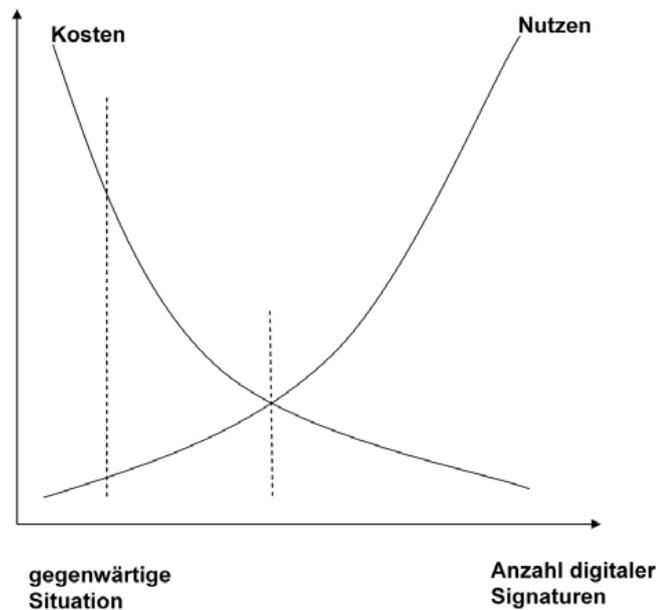


Abb. 3: Kosten-Nutzen-Relation elektronischer Signaturen in Abhängigkeit der Verbreitung [SMRG05]

Wie unsere Nutzenanalyse in Tabelle 4 gezeigt hat, kann E-Government alleine kein ausreichend starker Treiber für die Verbreitung der elektronischen Signatur sein. Da die Zahlungsbereitschaft für reine E-Governmentdienstleistungen wesentlich geringer ist, bedarf es offensichtlich weiterer Anwendungen aus dem E-Commerce und dem privaten Umfeld, wie beispielsweise [MuRo06].

Des Weiteren zeigt unsere Analyse, dass Zertifizierungsdienstleister eine Produktdifferenzierungsstrategie verfolgen sollten. Es zeigt sich zum Beispiel, dass mit Segment 3 ein preissensitives Segment vorhanden ist, für das eine qualifizierte Signatur den größten Nutzen aufweist. Die anderen beiden Segmente sind durchaus bereit, einen höheren Betrag zu zahlen, und unterscheiden sich dabei nur in dem bevorzugten Trägermedium. Während das Segment 1 den USB-Stick bevorzugt, wünscht sich Segment 2 das Mobiltelefon als Trägermedium. Interessanterweise ist im Durchschnitt allerdings der Chipkartenleser mit Signaturkarte das am meisten akzeptierte Medium und sollte als Lösung angeboten werden, wenn eine Produktdifferenzierung nicht möglich ist.

Ein weiteres interessantes Ergebnis ist, dass die Nutzer eine wesentlich höhere Zahlungsbereitschaft für akkreditierte und qualifizierte Signaturen im Vergleich zu fortgeschrittenen Signaturen aufweisen. Dies ist insbesondere für das E-Government interessant, da viele E-Government-Anwendungen rechtlich bindende Willenserklärungen erfordern [Kuns04].

6 Zusammenfassung

Die geringe Verbreitung von elektronischen Signaturen ist ein zentrales Hemmnis für die Verwaltungsmodernisierung. Ein Grund für die schleppende Verbreitung könnte in den aktuellen Marktpreisen zu finden sein. Daher haben wir in diesem Beitrag mit Hilfe einer empirischen Studie die Zahlungsbereitschaft von potenziellen Kunden für die elektronische Signatur ermittelt. Unsere Ergebnisse zeigen, dass eine Zahlungsbereitschaft durchaus vorhanden ist, aber deutlich unter den aktuellen Marktpreisen liegt. Weitere Ergebnisse sind, dass die Teilnehmer der Studie qualifizierte und akkreditierte Signaturen gegenüber fortgeschrittenen Signaturen bevorzugen und dass mit einer Produktdifferenzierung einzelne Kundensegmente gezielt angesprochen werden können. Weiterhin bleibt festzuhalten, dass die Zahlungsbereitschaft deutlich sinkt, wenn die Signatur ausschließlich für E-Government-Anwendungen einsetzbar ist.

Einschränkend bleibt festzuhalten, dass unsere Studie nicht repräsentativ ist. Allerdings halten wir die Ergebnisse dennoch für wertvoll, da gerade die in unseren Daten vorherrschende Gruppe durchaus als Interessent für elektronische Signaturen gilt [KPBU01]. Hochgerechnet auf die Gesamtbevölkerung sollte das Interesse deutlich niedriger liegen als in unserem Sample.

Literaturverzeichnis

- [BaBr04] Backhaus, K.; Brzoka, L.: Conjointanalytische Präferenzmessungen zur Prognose von Preisreaktionen. In: Die Betriebswirtschaft, 64 (2004), S. 39-57.
- [BüEK04] Bürger, M., Esslinger, B., Koy, H.: Das deutsche Signaturlbündnis: Ein pragmatischer Weg zum Aufbau einer interoperablen Sicherheitsinfrastruktur und Applikationslandschaft. In: Datenschutz und Datensicherheit 28, (2004) 3, S. 133-140.
- [Büge06] Bürger, M.: Fünf Thesen zur Regulierung. In M. Fluhr (Hrsg.), Kontaktbehaftet oder kontaktlos: auf jeden Fall Sicherheit. 2006, S. 424-427.
- [Deut01a] Deutscher Bundestag: Gesetz über Rahmenbedingungen für elektronische Signaturen (SigG). 2001. (a)

- [Deut01b] Deutscher Bundestag: Verordnung zur elektronische Signatur (SigV). 2001. (b)
- [FrRo05] Fritsch, L., Roßnagel, H.: Die Krise des Signaturmarktes: Lösungsansätze aus betriebswirtschaftlicher Sicht. In H. Ferderrath (Hrsg.), Sicherheit 2005. Köllen Druck+Verlag GmbH, 2005, S. 315-327.
- [GeST06] Gensler, S.; Skiera, B.; Theysohn, S.: Schätzung von Zahlungsbereitschaften mit der Conjoint-Analyse, Arbeitspapier, Goethe-Universität Frankfurt 2006.
- [HaSa02] Hartmann, A.; Sattler, H.: Commercial Use of Conjoint Analysis in Germany, Austria, and Switzerland, Arbeitspapier, Universität Hamburg 2002.
- [Hens00] Hensel-Börner, S.: Validität computergestützter hybrider Conjoint-Analyse, Wiesbaden 2000.
- [HoRo04] Hornung, G., Roßnagel, A.: Die JobCard: 'Killer-Applikation' für die elektronische Signatur? In: Kommunikation und Recht 7, (2004) 6, S. 263-269.
- [John87] Johnson, R. M.: Adaptive Conjoint Analysis Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing. Sawtooth Software Inc., 1987, S. 253–265.
- [KPBu01] KPMG, Bundesministerium für Wirtschaft und Technologie: Einsatzmöglichkeiten der Elektronischen Signatur in öffentlicher Verwaltung und Wirtschaft: Abschlussbericht, Berlin, 2001.
- [Kuns04] Kunstein, F.: Die elektronische Signatur als Baustein der elektronischen Verwaltung. Tenea Verlag, Bristol, Berlin 2004.
- [Lamb03] Lamberti, H. J.: Neuansatz für Digitale Signatur: Mit jeder Chipkarte jede Internetanwendung nutzen, In: Die Bank (2003) 3, S. 188-191.
- [Leis06] Leistenschneider, M.: Behördliche Realität degradiert das Signaturgesetz zur Farce. In M. Fluhr (Hrsg.), Kontaktbehaftet oder kontaktlos: auf jeden Fall Sicherheit. 2006, S. 428-431.

- [LiRo05] Lippmann, S., Roßnagel, H.: Geschäftsmodelle für signaturgesetzkonforme Trust Center. In O. K. Ferstl, E. J. Sinz, S. Eckert and T. Isselhorst (Hrsg.), *Wirtschaftsinformatik 2005*. Physica-Verlag, 2005, S. 1167-1187.
- [LoWo83] Louviere, J. J.; Woodworth, G.: Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregated Data. In: *Journal of Marketing Research*, 20 (1983), S. 350-367.
- [MuRo06] Muntermann, J., Roßnagel, H.: Security Issues and Capabilities of Mobile Brokerage Services and Infrastructures. In: *Journal of Information System Security* 2, (2006) 1, S. 26-41.
- [Roßn02] Roßnagel, A.: Rechtliche Unterschiede von Signaturverfahren. In: *Multimedia und Recht* (2002) S. 215-222.
- [Roßn03a] Roßnagel, A.: Eine konzertierte Aktion für die elektronische Signatur. In: *Multimedia und Recht* (2003) 1, S. 1-2. (a)
- [Roßn03b] Roßnagel, A.: Die fortgeschrittene elektronische Signatur. In: *Multimedia und Recht* (2003) S. 164-170. (b)
- [Roßn04] Roßnagel, H.: Mobile Signatures and Certification on Demand. In S. K. Katsikas, S. Gritzalis and J. Lopez (Hrsg.), *Public Key Infrastructures*. Springer, 2004, S. 274-286.
- [Roßn06] Roßnagel, H.: On Diffusion and Confusion: Why Electronic Signatures Have Failed, in S. Fischer-Hübner, S. Furnell and C. Lambrinoudakis (Hrsg.), *Proceedings of the 3rd International Conference on Trust, Privacy, and Security in Digital Business (TrustBus 06)*, Springer, 2006, S. 71-80
- [Sawt06] Sawtooth Software: Software for Online Surveys, Conjoint Analysis, and Quantitative Marketing Research. <http://www.sawtoothsoftware.com/>, Abruf am 10.11.2006.
- [Schr05] Schreiber, S.: Die qualifizierte elektronische Signatur: Vertrauensbonus vom Gesetzgeber, Schaffung von Vertrauen bei den Bürgern durch das deutsche

- Signaturbündnis?. In O. K. Ferstl, E. J. Sinz, S. Eckert and T. Isselhorst (Hrsg.), Wirtschaftsinformatik 2005. Physica-Verlag, 2005, S. 1187-1207.
- [Siet04] Sietmann, R. Chipkarten im Aufwind, c't 2004, S.26-28.
- [SkRe99] Skiera, B.; Revenstorff, I.: Auktionen als Instrument zur Erhebung von Zahlungsbereitschaften. In: Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung, 51 (1999), S. 224-242.
- [Srin82] Srinivasan, V.: Comments of the Role of Price in Individual Utility Judgement. In: Sheth, Jadesh (Hrsg.): Choice Models for Buying Behavior, Research in Marketing Supplement 1, Greenwich (Conn.) 1982, S. 81-90.
- [SMRG05] Strasser, M., Müller, G., Roßnagel, A., Gitter, R.: Kosten und Umsetzungsmodelle. In H. Reichl, A. Roßnagel, G. Müller (Hrsg.), Digitaler Personalausweis. Deutscher Universitäts-Verlag, 2005, S. 243-280.
- [VoHa98] Voeth, M.; Hahn, C.: Limit Conjoint-Analyse. In: Marketing – Zeitschrift für Forschung und Praxis, 20 (1998), S. 119-132.
- [VonB06] von Berg, A.: Digitale Signatur: Ist die Regulierung gescheitert? In M. Fluhr (Hrsg.), Kontaktbehaftet oder kontaktlos: auf jeden Fall Sicherheit. 2006.

Cross-organizational processes in Public Administrations: Conceptual modeling and implementation with Web Service Protocols

Jörg Ziemann, Timo Kahl, Thomas Matheis

Institut für Wirtschaftsinformatik (IWi)
im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI)
66123 Saarbrücken
{joerg.ziemann, timo.kahl, thomas.matheis}@iwi.dfki.de

Abstract

An increasing level of cooperation between public administrations on regional and national levels requires methods to develop interoperable E-Government systems. It leads to the necessity of an efficient modeling of cross-organizational business processes and their subsequent implementation. This is a complex task, since it includes modeling of processes from various perspectives - modeling of internal and cross-organizational processes - and on various technical levels to enable both conceptual modeling and the execution of processes that comply with the conceptual models. In this paper a methodology is described for creating models of cross-organizational business processes based on Event-driven Process Chains (EPC). Building on this, it is described how these conceptual models can stepwise be transformed to technical process models in the form of Web Service protocols for implementing the cross-organizational processes. The methodology is motivated and explained on the basis of an E-Government reference model for the German Plan Approval Procedure.

1 Introduction

Due to the increasing heterogeneity and dynamics of the European Union, more and more public administrations within Europe are challenged to work together and to adapt continuously to rapid technological changes. New legal settings, strategic commitments, higher expectations for improved quality of service as well as rapid technological advances create a new dynamic

and complex administration environment, which requires transparency, flexibility and mobility from European public administrations. In order to improve the transparency of collaborative processes in public administrations, enterprise modeling is a promising approach. Enterprise models are easy to understand for various stakeholders of an organization and thus can help to get a joint understanding of processes, organizational structures and underlying IT-systems over organizational borders. After such cross-organizational process (CBPs) has been developed, the challenge is to ensure an implementation complying with this model. Thus, in this paper, a concept for collaborative modeling and a corresponding protocol based implementation will be introduced, which help organizations to define interactions on design-level and to transform them into a machine-readable protocol.

Web Service protocols enable public administrations to customize and implement CBPs in a decentralized and vendor-independent way. Due to the heterogeneity of IT-systems in public administrations and a relatively high resistance to changing underlying application systems, especially the last point recommends the use of the open Web Service standards in the E-Government. Interactions of public administrations comprise various groups, e.g. not only Government-to-Government (G2G) but also Government-to-Business (G2B) and Government-to-Citizen (G2C). Since citizens are not expected to explicitly model their business processes for preparing an electronic exchange of messages, this paper focuses on G2G and G2B scenarios.

After describing related research in section 2, in section 3 a method is presented that allows the creation of CBPs on a conceptual level. Here, the requirements and single steps of the collaborative modeling method are illustrated on the E-Government case study. In section 4 we change from design time to run time and describe how to implement the modeling approach presented before with Web Service protocols. Section 5 concludes with a summary and an outlook on future research.

2 Related Research

Various standards for promoting interactions have been developed especially for the public sector. The so called SAGA has been developed by the German government within the initiative „BundOnline 2005“ [KBSt05]. SAGA defines standards and architectures for E-Government applications. This includes, for example, standards for data descriptions based on

XML, middleware technologies and protocols (e.g. J2EE, .Net, SOAP), or security standards like ISIS-MTT. The Online Services Computer Interface (OSCI) represents a protocol standard for the local authorities of Germany. It consists of a set of protocols like the OSCI transport protocol and the OSCI XÖV protocols. SAGA and OSCI allow public administrations to realize E-Government interactions, but do not support efficient modeling and implementation of cross-organizational business processes, nor do they comprise the implementation of CBPs with Web Service protocols as provided by the Business Process Execution Language for Web Services (BPEL, [ACDG03]).

With the rise of SOA and an increasing focus on coarse grained business functions than on fine grained technical modules, protocol standards - like BPEL - moved closer to conceptual business processes [ACKM04]. In the industrial field business protocols provide a widely-used approach for the message exchange between the involved partners of a collaboration as well as the design of application systems. The definition from [LeRo04] of a protocol focuses on the viewpoint of one partner: “a business protocol specifies the potential sequencing of messages exchanged by one particular partner with its other partners to achieve a business goal. I.e. a business protocol defines the ordering in which a particular partner sends messages to and expects messages from its partners based on actual business context“. Other authors use the term conversation protocol, which they define as “... a specification of a set containing all correct and acknowledged conversations”, where a conversation represents “sequences of operations (i.e., message exchanges) that could occur between a client and a service as part of the invocation of a Web service” [ACKM04]. The term protocol is closely related to the concept of Web Service Choreography, defined by as “... a multi-party contract that describes from a global view point the external observable behavior across multiple clients (which are generally Web Services but not exclusively so) in which external observable behavior is defined as the presence or absence of messages that are exchanged between a Web Service and its clients” [ABPR04]. In this paper the term Web Service protocol is used for technically detailed descriptions of allowed interactions between Web Services, as provided for example by BPEL or the Web Service Choreography Description Language (WS-CDL, [KaBR04]).

Literature on protocols and business processes usually focuses on modeling the exchange of fine grained, technical objects, like [MASN06], where state charts are used to display protocols, and disregards interaction descriptions stemming from the business level. This goes also for working on protocol engineering from agent related research, e.g. [ViHu03], [AFLS05].

Another example is [DeSi04], who claim to take into account business process models for development of protocols, but only provide models on a technical level represented with state charts, disregarding the organizational perspective in CBP development.

Recently the development of CBP on the business level for preparation of a CBP implementation was investigated, e.g. in [GLKZ06]. Also in that context, [ZDHB06] described a graphical language to describe service interactions, concentrating on global models. However, this language focuses on ensuring technically correct models, and the global models displayed by it are inadequate for illustrating complex CBPs to non-technicians.

The approach described in this paper focuses on transforming conceptual process models in form of EPC into executable protocols (abstract BPEL processes), and thus is comparable to OMGs MDA [JiJo03] approach. But in difference to that approach, our approach is based on a notation developed for business analysts and provides a specific methodology of how to derive the necessary models in different organizational contexts.

Extending former ATHENA results regarding the conceptual modeling of CBPs, we describe a way of how to prepare and develop Web Service protocols in an eGovernment environment. Thus, following the aim of the R4eGov project, the problem field of interoperability in the E-Government is tackled.

3 Conceptual Modeling of Cross-organizational Business Processes

In order to model the entire enterprise and its interfaces, different modeling dimension are necessary. Modeling frameworks which previously have been applied for enterprise modeling include, for example, the “Framework for Information Systems Architecture” (Zachman Framework) [Zach87] and the “Architecture of Integrated Information Systems” (ARIS) [Sche99]. Both frameworks offer modeling support for various dimensions of an enterprise. Although both frameworks combine different user perspectives and allow modeling on different levels of abstraction, they lack methods which allow modeling of cross-organizational collaborations, as a creation of an external view on the organization is not supported.

The modeling methodology described in this paper aims at developing easy-to-understand business process models that are transformable to Web Service protocols. For modeling conceptual CBPs the EPC was chosen. The EPC has been established for business process management for more then one decade and represents business processes by sequences of

events and functions put in logical and timely order; a further description can be found in [KeNS92]. The concepts for collaborative modeling and protocol based implementation will be described and evaluated on the basis of a scenario introduced in the following chapter.

3.1 A Cross-organizational Process in Public Administration: The Plan Approval Procedure

A Plan Approval Procedure (PAP) is a special, formal administrative procedure, whose single regulations are described in detail in the §§ 72 et seqq. of the Administrative Procedures Law (VwVfG). The aim of the PAP is to achieve the obligatory official approval of a plan (OAP) by a so called Plan Approval Decision. An OAP enables the matching of a number of public and private interests, which are addressed by a plan. Thus, it represents a substitution of necessary official decisions like for example administrative decisions or permissions by only one administrative act and at the same time eliminates possible further claims for omission [Beck05]. An OAP also determines exactly where the construction will be located later on. The regulations of VwVfG apply only if there is no prescription in a more specific law. These specific laws also determine for which procedures a PAP is necessary. An OAP is, for example, necessary for the construction of streets (according to the Federal Street Law (FStrG)) or the construction of airports (according to the Air-Traffic Law (LuftVG)) [HoSB01]. A PAP can roughly be divided into two main steps: a “hearing procedure“ and a “decision procedure”. Because a PAP is required for most of the construction projects and involves various partners like public administrations, businesses as well as citizens, it is a suitable procedure to illustrate the conceptual modeling and implementation of a cross-organizational business process in the public sector.

Fig. 1 gives a general overview about the PAP. At the beginning of the process the project agency (e.g. a manufacturer) submits the plan for the construction project to the hearing agency. This plan is then checked and published by the hearing agency. After the plan has been published, the parties involved (e.g. the citizens and the retailer) have the right to have a look at the plan and, if necessary, raise objections. The hearing agency checks these objections and decides, in consultation with the parties involved, whether the plan has to be revised or not. In case of a necessary revision, the plan has to be revised by the project agency (e.g. the manufacturer). Afterwards, the revised plan is submitted to the hearing agency again, which

then forwards the plan to the approval agency. The Plan approval agency has the assignment to check the procedure according to the rules, to accept the plan and to announce it.

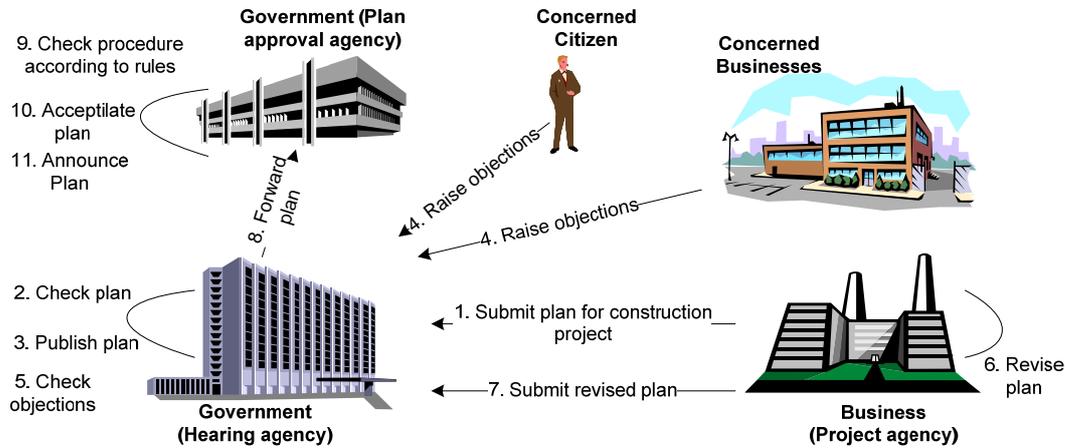


Figure 1: Overview of Plan Approval Procedure

3.2 Modeling Private, View and Cross-organizational Processes with EPCs

In a first step the scenario was analyzed concerning the requirements of the different partners. The analysis resulted in the following requirements for CBP models: First, the internal business processes of each partner have to be linked into a CBP without revealing confidential information. Second, the CBP models should only capture collaboration essentials, e.g. the collaborating parties only have to be interested in that part of a process which determines the interaction behavior; additional information could be confusing. Third, a simplified process adoption has to be achieved, e.g. a Hearing Agency interacting with different citizens and business partners should not require different private processes for each collaboration it is involved in. Fourth, internal process should still be flexible, e.g. although the interaction is fixed, the private processes behind the view processes have to remain flexible.

In order to realize the scenario requirements, it was necessary to introduce another abstraction layer which allows the externalization of information, which is necessary for the interaction with external partners. In the following, internal processes are defined as Private Processes (PP) which are only known to their owning organization and not exposed to the outside world. The abstraction of information is achieved by the introduction of Process Views as an additional abstraction layer between the PPs and the CBP model as proposed by Schulz [Schu02], [ScOr04]. Process views provide a process-oriented interface towards business partners. Process views are an abstraction of the private processes, containing information that needs to be

published for the purpose of an interaction. Several tasks of a private process can be combined to one view task. This leads to the following definition: A View Process (VP) abstracts information from one or more PPs and thus enables companies to hide critical information from unauthorized partners. It is an interface to the outside world which extracts only information, which is necessary for the interaction with one or more potential partners. Thus, a VP can be seen as a general interaction description of one or more PPs from the perspective of one partner.

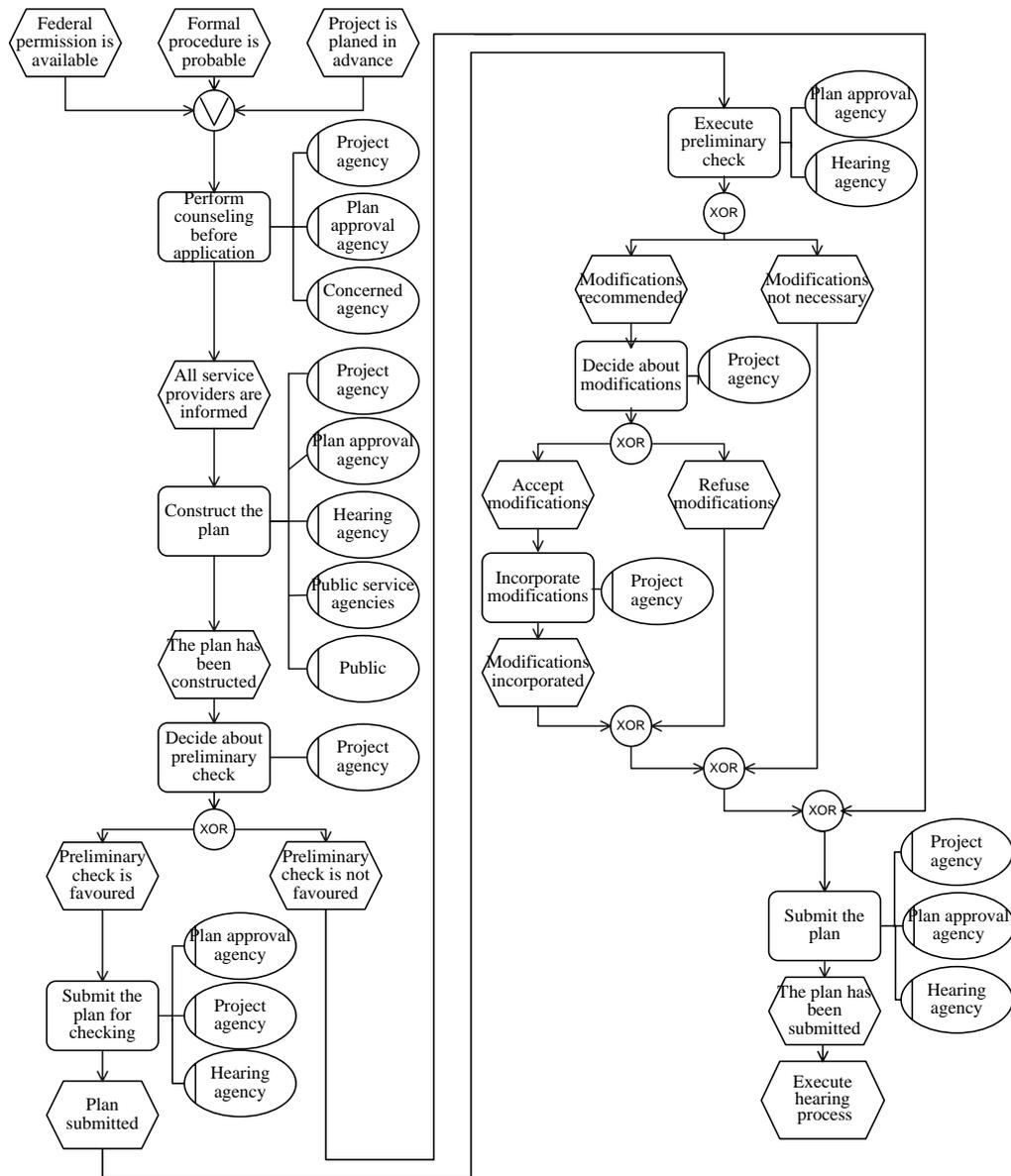


Figure 2: First part of the Plan approval Procedure – Global process

While a VP describes valid interactions from the perspective of one partner, a CBP describes these interactions from a neutral perspective, capturing all allowed interactions between all partners. Since a VP can contain interactions with different partners, in special cases a VP can

capture all information contained in a CBP even in multi-party collaborations, when the collaboration activities consist only of those interactions between the partners and an organization that are described in the VP, offered by this organization.

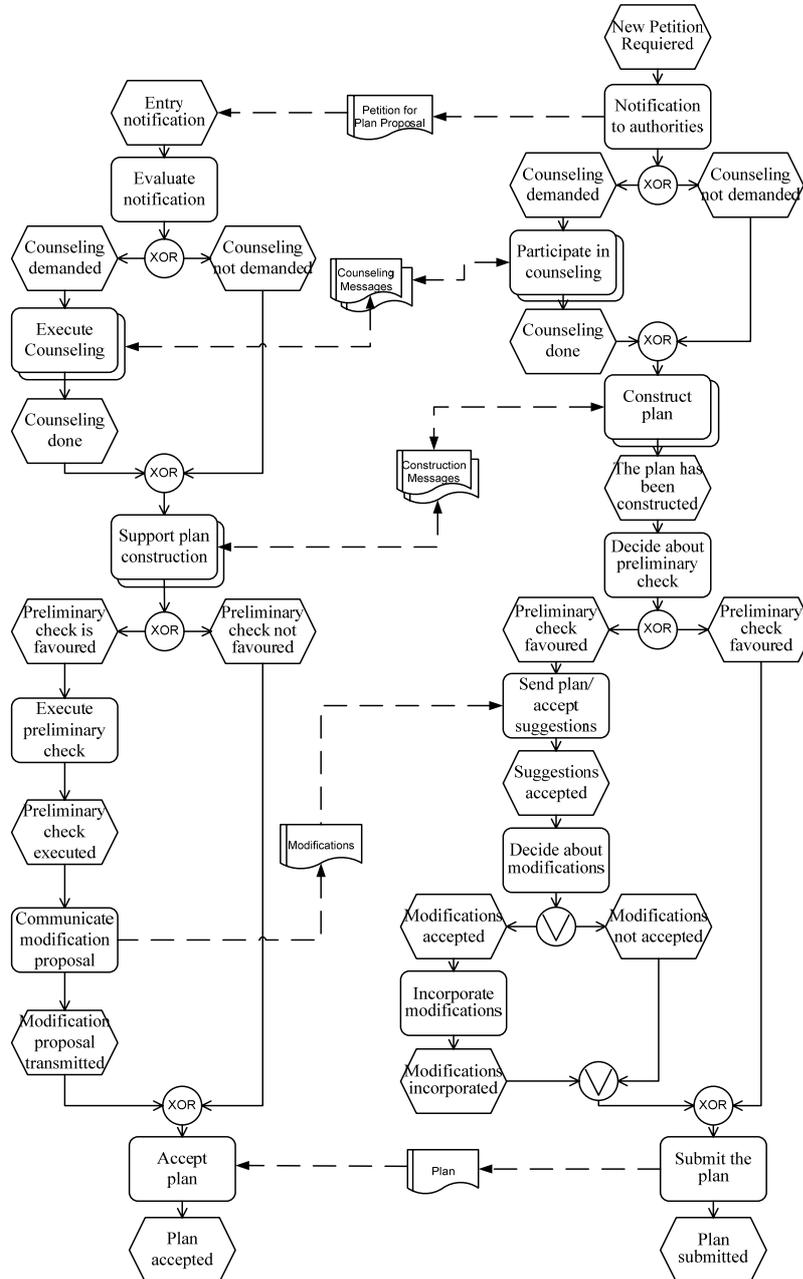


Figure 3: First part of the Plan Approval Procedure - View Processes of Government and Project Agency

While more technical definitions of view processes reduce them to descriptions of digital message exchanges (cp. [Buss02]), on the conceptual level also partner interactions regarding money (“Payment received”) or material (e.g. “Deliver Container”) can be described in a view process. In the following, the CBP-concept will be shown on the basis of the scenario. In fig. 2

the overall CBP is illustrated as a global process, fig. 3 shows two corresponding view processes.

The disassembly of the global process to view processes was guided by the following principles: First, the view process has to show the other party in which sequence which messages are exchanged. Nonetheless, since the global process is public anyway, no information hiding must be involved in developing the view processes, e.g. all information shown in the global process can be also shown in the view processes. For the sake of reducing complexity, these functions might be left out of the view process. On the other hand, even if functions are not directly involved in interactions, they might provide the collaboration partner with a better understanding of the process and also allow their own party to adapt its private processes better to the view process. The function “Execute preliminary check” of the government view process is an example of such a function.

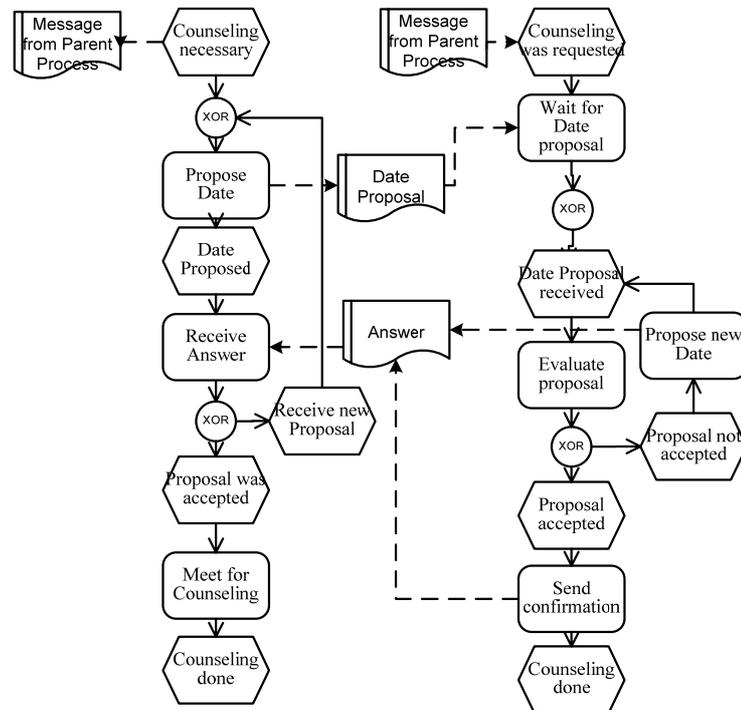


Figure 4: Sub View Processes to agree on a Meeting Date between Government and Project Agency

Second, to constrain complexity, the view processes show only interactions on the level of the global process. Interactions happening below that level are displayed in sub-protocols. For example, the function of “Counseling” of the government contains a communication with the function “participate in counseling” of the project agency. The symbol of the messages lying on each other indicates a finer grained exchange of messages. The corresponding sub protocols are displayed in sub view processes. Thus fig. 4 illustrates the sub view process representing the

interactions between the functions “execute counseling” and “participate in counseling” of the view process shown in fig. 3. Note, that view processes on this level of granularity often display standard situations which are applicable in various contexts and thus can be seen as re-usable interaction patterns.

Third, since the global process indicates the organizational units responsible for functions, the view process functions should be derived accordingly: Each function of a global process that is annotated with organization A must appear in the view process of organization A. If more parties have an organizational unit attached to this function, it has to be decided how this function has to be split up. Normally, an exchange of messages between the parties must take place. An example is splitting of the function “Perform counseling before Application” from the global process to the two functions “Execute Counseling” and “Participate in Counseling” from the view processes of government and project agency.

3.3 Modeling Procedure

In section 3.2 we introduced concepts which support users in designing collaborative processes in public administration. But, providing methods to externalize internal information is just one challenge in supporting interoperable organizations. At least as important as the appropriate method is a procedure model which supports organizations in planning and designing CBPs. Thus, a procedure is required that describes in which order models have to be created to make best use of the approach. Concerning the creation of views and CBPs, three possible procedures can be identified. The procedure that was used in chapter 3.2 was a top-down procedure. In a top-down approach, the partners start identifying a common picture of the collaborative scenario in terms of a CBP model. This contains the definition of the whole interaction of all partners. Afterwards, each partner has to create his views according to the process steps that he will be executing. This also might need iterations for redefining the CBP (solid arrows 1 and 2 in Fig. 3). As a last step the partners have to define their private processes (solid arrows 3). Within a bottom-up approach each organization starts with the specification and modeling of their private processes. The next step contains the creation of process views which represent the necessary interaction of the PP (cp. the dashed arrows 1 in Fig. 3). The views are then combined into CBPs (dashed arrows 2). Depending on how well the process views of the process partner fit, variations on the own view might be necessary to finalize the modeling activities (dashed

arrows 3). The third scenario is a middle-out procedure. It starts with one partner specifying and modeling its private processes and offering a process view to its partners. The partners can use this process based interface to link it to their internal processes via process views. This would conform to a bottom-up approach for one partner and a top-down for the others. Which procedure is suitable depends on existing partner processes and the relationship between the different organizations.

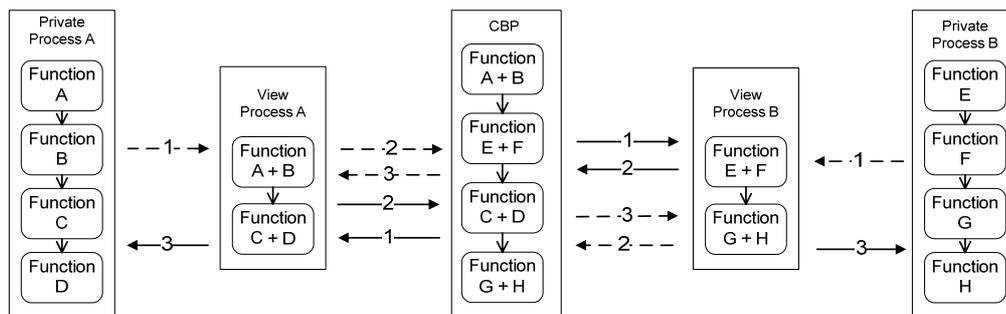


Figure 5: Modeling procedures: Top-down vs. bottom-up

4 Deriving Business Protocols from View Processes

4.1 Business Protocols based on Web Services

Various standards exist to describe protocols, including the Business Process Specification Schema (BPSS) of ebXML [CCKH01], the Partner Interface Processes (PIPs) of RosettaNet (<http://rosettanet.org>), WS-CDL and the abstract processes of BPEL, which are also called BPEL protocols [ACDG03]. To establish an E-Business conversation, several components are necessary: interfaces published in a network, interaction descriptions and partner roles, a standard vocabulary and an environment of security and trust [Masu03]. RosettaNet, being a prominent example for E-Business protocols, contains all of the components listed above, but comprises only a predefined list of interaction patterns (called PIP) described with UML activity diagrams, text tables and XML documents. To provide more specific and technical process descriptions, [Masu03] and [Khal05] propose to transform PIPs, which represent proven well established reference models for CBPs, to BPEL processes. Though the methodology described in this paper also aims on generating BPEL processes, it does not focus on pre-established interaction patterns (like PIPs do), but allows the development of individual CBPs. Since Web Service standards provide interface descriptions (WSDL) and interaction

descriptions (e.g. abstract BPEL processes), they can be seen as complementary to established E-Business protocol standards like RosettaNet.

Similar to the different ways of describing conceptual CBPs, different methods exist to specify protocols: the first describes the interactions of all partners with the help of global models; the second only describes interactions of one partner with a so called abstract process (also called process skeleton, process stub or public process). Abstract Processes describe interactions from the viewpoint of one partner, thus they can only describe the interactions between this partner and one or more of its partners but not the interactions between his partners, where this partner is not directly involved. In comparison to abstract processes, global models allow for better use of model checking techniques (cp. [FuBJ03]). WS-CDL seems to be the only Web Service based standard for describing global processes, but only a few prototypical tools are supporting this standard. While global models are valuable in the design and analysis of CBPs, for implementing them in general, abstract processes are to be preferred: They comply with organizations demands of a de-centralized process execution (e.g. no central engine is required) and show as little process information as possible and only to immediate collaboration partners. BPEL abstract processes can either be used during design-time to ensure that private BPEL processes of the collaborating partners are complementary, they also might be used on run-time as input to an protocol engines as described in [MASN06].

4.2 Relating Conceptual CBPs to Business Protocols

In order to execute collaborations between two or more partners apart from a modeling method for CBPs a method is needed which allows companies to execute CBPs. Therefore, the concepts of Private Process, View Process and CBP described above for the aim of modeling CBPs on a conceptual level, can be matched to the more technical, Web Service and protocol related terms of abstract process and global model to enable their execution by IT systems. The interactions between various parties (e.g. CBPs respectively global models) can be described with WS-CDL, interactions of one service with its partner services (e.g. view processes respectively abstract processes) can be described by BPEL. How private EPC processes can be transformed to executable BPEL processes is described in (cp. [ZiMe05]).

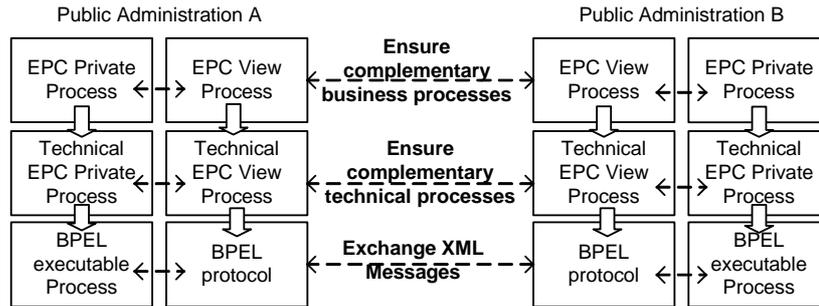


Figure 6: Process types involved in protocol development

Since the interactions realized by protocols should be controlled by business analysis, we propose an EPC based design of the (BPEL) protocols. As fig. 6 illustrates, we propose a two step transformation from the EPC level to the Web Service level. On the upper level, private and view processes are modeled by business analysts as described in section 3. Afterwards, these EPCs are enriched with Web Service specific information and shaped according to conventions for compliance with BPEL processes. These models contain the Web Service invocations (also describing the message exchanged between collaborating administrations) and all control flow information relevant to specify the sequence of interactions. Note, that this type of EPC processes can also be used as a visualization of BPEL processes. In the second step these technical EPC processes are further enriched by BPEL programmers, e.g. with variables used in a BPEL process to realize the control flow specified by the technical EPC process, and annotated in XML.

In section 3 it was described how conceptual models are derived in a coherent way, ensuring that the view processes of collaborating parties are complementary. Following this model driven approach ensures that the implementation of the CBP is compliant to the conceptual model. Nonetheless, since the technical model contains further information (e.g. the name of Web Services to be invoked), the technical view process models have to be synchronized also.

4.3 Deriving Web Service Protocols for the Plan Approval Procedure

Based on the sub view processes shown in fig. 4, now technical EPC models are derived, that contain the BPEL syntax necessary to specify the interactions between both parties. Since BPEL can represent both a graph based and a block-oriented control flow (e.g. a containing a While Loop), all EPC control flow elements can be transformed to BPEL (cp. [ZiMe05]). If the EPC functions represent interactions (e.g. “receive message”, “send message” etc.), they can be

transformed to corresponding BPEL orders, if an EPC function represents activities not captured by the BPEL syntax (like “Evaluate proposal” in fig. 4) an individual Web Service has to be created that will be invoked by the BPEL process. In [ZiMe05] further details regarding the transformation of EPCs to BPEL are described. The result of the transformation from the EPCs in fig. 4 to the BPEL aligned EPCs is illustrated in fig. 7.

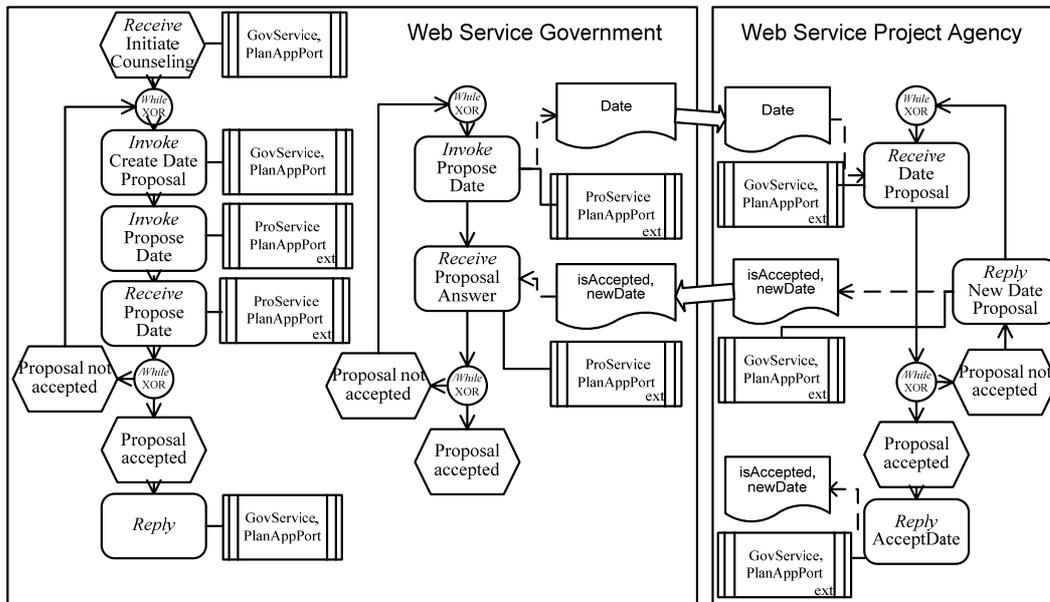


Figure 7: Technical EPC processes: Private, View Process of Government and View Process of Enterprise

The whole interaction takes place between two Web Services, one offered by the Government (“GovService”) and one by the Project Agency (“ProService”). The source and destination services are attached to the interaction activities, e.g. “Invoke propose Date” sends a message called “Date” to the Enterprise Service. The notation distinguishes between internal Web Services (e.g. “Receive initiate counseling” is invoked by a Web Service inside the public administration) and external Web Services, marked with the letters “ext”. The right side shows the abstract process of the Project Agency, which interacts with the abstract processes of the government. On the right of the government view process, the corresponding private process is shown, also. The abstract processes only contain the information necessary to describe the interaction with the other Web Service. Thus, the function “Evaluate Proposal” does not appear in the abstract process, though for illustrative reasons it was part of the conceptual view process. As the synchronous invocation of the “Create Date Proposal” function shows, these functions might be added again in the private processes. The private process also contains functions to interact with the top level process: It is called by its parent process and after the meeting is fixed, it replies the result to the parent process.

5 Summary and Outlook

We described how CBPs can be described by a combination of private, view and collaborative processes with EPCs. This concept was motivated and evaluated on the basis of an eGovernment reference model. In order to automate this collaborative process, we proposed a transformation to business protocols and described some forms of such protocols. Due to their close relationship to executable models of private process and the possibility to describe them in a machine readable format, the abstract processes of BPEL were chosen as the target protocol. Accordingly, we described a transformation of the EPC based concepts to model CBPs view processes to BPEL abstract processes and illustrated this transformation by the use case introduced before. For example, we described under what criteria conceptual view processes can be derived from conceptual global processes and showed how the use of sub protocols on the conceptual and technical level can reduce the complexity of CBPs in the E-Government.

One aspect that requires further research is the use of supporting tools that ease the task of exchanging process models between different enterprises and to distinguish between the different model types used in CBPs. This tool could support actively business process management by checking, verifying or even automatically negotiating consistency of models. As fig. 6 shows, only taking in regard view process and private processes, numerous interdependencies exist between the different model types. This paper focused on a Top-Down approach; in future we will investigate further procedures to synchronize those models, e.g. taking private processes as the starting point of the CBP development.

Acknowledgements

The work published in this paper is (partly) funded by the E.C. through the ATHENA IP and the R4eGov project. It does not represent the view of E.C. or the ATHENA or the R4eGov consortium, and authors are solely responsible for the paper's content.

References

- [ABPR04] Austin, D.; Barbir, A.; Peters, E.; Ross-Talbot, S.: Web Services Choreography Requirements-W3C Working Draft 11 March 2004. <http://www.w3.org/TR/2004/WD-ws-chor-reqs-20040311/>, online: 2004-05-30.
- [ACDG03] Andrews, T.; Curbera, F.; Dholakia, H.; Golan, Y.; Klein, J.; Leymann, F.; Liu, K.; Roller, D.; Smith, D.; Thatte, S.; Trickovic, I.; Weerawarana, S.: Business Process Execution Language for Web Services – Version 1.1. 2003. <http://dev2dev.bea.com/techtrack/BPEL4WS.jsp>, online: 2004-01-10.
- [ACKM04] Alonso, G.; Casati, F.; Kuno, H.; Machiraju, V.: Web Services – Concepts, Architectures and Applications. Springer, Berlin 2004.
- [AFLS05] Alonso, F.; Frutos, S.; Lopez, G.; Soriano, J.: A Formal Framework for Interaction Protocol Engineering.. In: Pechoucek, M.; Petta, P.; Varga, L.Z. (Eds.): CEEMAS 2005, LNAI 3690, Springer-Verlag Berlin 2005, pp. 21–30.
- [Beck05] Becker, L.; Fügemann, M.; Gerlach, A.; Theis, F.-J.: Praxis Ausbildung – Verwaltungsverfahren und Verwaltungsprozess. Deutscher Anwaltverlag, Bonn 2005.
- [Buss02] Bussler, C.: Public Process Inheritance for Business-to-Business Integration. In: Buchmann, A. P.; Casati, F.; Fiege, L.; Shan, M. C.: Technologies for E-Services – Third International Workshop TES 2002 Hong Kong. Lecture Notes in Computer Science 2444, Springer 2002, pp. 19-28.
- [CCKH01] Clark, J.; Casanave, C.; Kanaskie, K.; Harvey, B.; Clark, J.; Smith, N.; Yunker, J.; Riemer, K.: ebXML Business Process Specification Schema Version 1.01. UN/CEFACT and OASIS 2001.
- [DeSi04] Desai, N.; Singh, M. P.: IEEE International Conference on Web Services: Protocol-Based Business Process Modeling and Enactment. San Diego 2004.

- [FuBJ03] Fu, X.; Bultan, T.; Jianwen, S.: A top-down approach to modeling global behaviours of Web Services. Workshop on Requirements Engineering and Open Systems (REOS), Monterey, CA 2003.
- [GLKZ06] Greiner, U.; Lippe, S.; Kahl, T.; Ziemann, J.; Jaekel, F. W.: A Multi-level Modeling Framework for Designing and Implementing Cross-Organizational Business Processes. In: Sadiq, S.; Reichert, M.; Schulz, K. (Ed.): Proceedings of the 1st International Workshop on Technologies for Collaborative Business Process Management (TCoB), ICEIS 2006, Paphos/Cyprus 2006, pp. 13-23.
- [HoSB01] Hoppe, W.; Schlarman, H.; Buchner, R.: Rechtsschutz bei der Planung von Straßen und anderen Verkehrsanlagen. 3. edition, Beck, München 2001.
- [JiJo03] Mukerji, J., Miller, J.: MDA Guide Version 1.0.1, 2003, <http://www.omg.org/docs/omg/03-06-01.pdf>, online: 2006-07-20.
- [KaBR04] Kavantzias, N.; Burdett, D.; Ritzinger, G.: Web Services Choreography Description Language Version 1.0 - W3C Working Draft 27 April 2004, <http://www.w3.org/TR/2004/WD-ws-cdl-10-20040427/>, online: 2004-06-05.
- [KBS05] KBSt: Standards und Architekturen für E-Government-Anwendungen (SAGA). Version 2.1, Schriftenreihe der KBSt 82 (2005). <http://www.kbst.bund.de/saga>, online: 2006-07-20.
- [KeNS92] Keller, G.; Nüttgens, M.; Scheer, A. W.: Semantische Prozeßmodellierung auf der Grundlage "Ereignisgesteuerter Prozeßketten (EPK)". In: Scheer, A.-W. (eds.): Veröffentlichungen des Instituts für Wirtschaftsinformatik 89, Saarbrücken 1992.
- [Khal05] Khalaf, R.: From RosettaNet PIPs to BPEL Processes: A Three Level Approach for Business Protocols. 3rd International Conference, BPM 2005, Nancy, 2005.
- [LeRo04] Leymann, F.; Roller, D.: Modeling Business Processes with BPEL4WS. In: Nüttgens, M.; Mendling, J.: XML4BPM 2004 - XML Interchange Formats for Business Process Management. 1st Workshop of German Informatics Society

e.V. (GI) in conjunction with the 7th GI Conference “Modelierung 2004“, Marburg 2004, pp. 7-24.

- [MASN06] McNeile; Ashley; Simons; Nicholas: “Protocoll modeling: A modeling approach that supports reusable behavioural abstractions”. In: *Software&Systems Modeling* 5 (2006) 1, pp. 91-107.
- [Masu03] Masud, S.: Use RosettaNet-based Web Services, Part 1: BPEL4WS and RosettaNet, 2003. <http://www-128.ibm.com/developerworks/webservices/library/ws-rose1/>, online: 2006-04-28.
- [Sche99] Scheer, A. W.: *ARIS – Business Process Modeling*, 2. ed., Springer, Berlin 1999.
- [Schu02] Schulz, K. A.: *Modeling and Architecting of Cross-Organizational Workflows*. PhD thesis, The School of Information Technology and Electrical Engineering, The University of Queensland, Australia 2002.
- [ScOr04] Schulz, K. A.; Orlowska, M. E.: Facilitating cross-organizational workflows with a workflow view approach. *Data&Knowledge Engineering* 51 (2004) 1, pp. 109-147.
- [ViHu03] Vitteau, B.; Huget, M.: Modularity in Interaction Protocols. In: Digum, F.: *ACL 2003*, LNAI 2922, pp. 291-309.
- [Zach87] Zachman, J. A.: A Framework for Information Systems Architecture. *IBM Systems Journal* 26 (1987) 3, pp. 276-292.
- [ZiMe05] Ziemann, J.; Mendling, J.: Transformation of EPCs to BPEL – A pragmatic approach. 7th International Conference on the Modern Information Technology in the Innovation Processes of the industrial enterprises, Genoa 2005.
- [ZDHB06] Zaha, J.M; Dumas, M; Hofstede, A ter; Barros, A.; Decker, G.: Service Interaction Modeling: Bridging Global and Local Views. In: *Proceedings 10th IEEE International EDOC Conference (EDOC 2006)*, Hong Kong 2006.

Einführung in den Track

Information Services in der Logistik

Prof. Dr. Christian Bierwirth

Universität Halle-Wittenberg

Prof. Dr. Herbert Kopfer

Universität Bremen

Prof. Dr. Kai Furmans

Universität Karlsruhe

Dr. Dieter Pütz

Deutsche Post AG

Neben den modernen Informationstechnologien hat vor allem die Leistungsfähigkeit von Transport- und Umschlagsystemen einen erheblichen Einfluss auf die zukünftige Entwicklung der Arbeitsprozesse zwischen Industrie-, Handels- und Dienstleistungsunternehmen.

Der Track widmet sich daher Strategien, Modellen, Algorithmen und Softwaresystemen, die der Verbesserung der Güterflüsse in einer Supply Chain dienen. Im Fokus stehen hierbei die Planung und Konfiguration logistischer Netzwerke, die strategische Transportplanung, die operative Transportplanung und ihre Unterstützung durch Telematik und Verkehrsinformationssysteme, das Management von Terminals und intermodalen Hubs sowie Fragen der Gestaltung kooperativer Systeme und elektronischer Marktplätze für Logistikleistungen.

Programmkomitee:

Prof. Ulrich Stache, Universität Siegen

Prof. Essig, BW-Hochschule München

Prof. Andreas Otto, Universität Regensburg

Prof. Dr. Martin Grunow, Technical University of Denmark, Copenhagen

***ComEx*: Kombinatorische Auktionen zum innerbetrieblichen Austausch von Logistikdienstleistungen**

Oleg Gujo, Michael Schwind, Jens Vykoukal, Kilian Weiß

Institut für Wirtschaftsinformatik
Johann Wolfgang Goethe Universität Frankfurt
60056 Frankfurt am Main
{gujo, schwind, jens.vykoukal, weiss}@wiwi.uni-frankfurt.de

Tim Stockheim, Oliver Wendt

Lehrstuhl für Wirtschaftsinformatik und Operation Research
TU Kaiserslautern
67653 Kaiserslautern
{stockheim,wendt}@wiwi.uni-kl.de

Abstract

Der Austausch von Ladekapazitäten in Transportbörsen ist ein etabliertes Verfahren der logistischen Praxis. Wenige der meist Web-basierten Marktplätze sind jedoch in der Lage, Synergien zu berücksichtigen, die durch geeignete Kombination von Lieferstrecken verschiedener Transportanbieter entstehen. Eine Methode hierfür ist die Verwendung kombinatorischer Auktionen, bei denen auf Bündel von Streckenabschnitten geboten werden kann. Der Beitrag beschreibt eine kombinatorische Auktion zum innerbetrieblichen Austausch von Logistikdienstleistungen. Dabei werden Lieferaufträge für ein in Profitcenterstruktur organisiertes Unternehmen der Nahrungsmittelindustrie zum Outsourcing vom jeweiligen Profitcenter freigegeben, wenn die geografische Lage des Kunden eine günstigere Belieferung durch ein anderes benachbartes Profitcenter zu günstigeren Kosten ermöglicht. Umgekehrt können Profitcenter Aufträge annehmen, wenn dadurch insgesamt niedrigere Gesamtkosten entstehen. Die Kostenberechnung wird durch ein integriertes Routingsystem vorgenommen, das In- und Outsourcing findet mittels des Auktionsmechanismus *ComEx* statt. Die Aufträge verbleiben zum Zwecke der Kundenbindung beim jeweiligen Profitcenter, der Anreiz zum Tausch entsteht durch eine entsprechende Beteiligung der Profitcenter an der Kostenersparnis. Nach der Darstellung der Web-basierten Logistikauktion zusammen mit der Tourenplanungssoftware *DynaRoute* suchen wir die Strategie für das Outsourcing der gebündelten Lieferaufträge, welche die größte Kostenreduzierung erzielt.

1 Einleitung

Seit einigen Jahren gewinnt die Anwendung von kombinatorischen Auktionen (KA) für Beschaffungs- und Ressourcenallokationsprozesse an Bedeutung. Getrieben durch die Entwicklung von Mechanismen für die Vergabe von Frequenzbändern an TK-Unternehmen in England¹, Deutschland und den USA² gelangte die KA in den Fokus des *Electronic Market Engineering* [McMi95]. Der vorliegende Beitrag beschäftigt sich mit einer elektronischen KA zum Austausch von Lieferleistungen in einem mittelständischen Unternehmen, welches in Profitcenterstruktur organisiert ist. Dabei haben die Profitcenter, denen regionale Liefergebiete zugeordnet sind, die Möglichkeit, Lieferaufträge, welche nur zu ungünstigen Kosten beliefert werden können, an benachbarte Profitcenter abzugeben. Die erzielte Kostenersparnis wird zum Zwecke der Anreizkompatibilität nach einem vorher festzulegenden Schlüssel auf die Profitcenter umgelegt. Kernpunkt dieser Arbeit ist die Frage, wie eine Bündelung auszulagernder Lieferaufträge vorgenommen werden kann, so dass das Zusammenwirken von KA und lokaler Routenoptimierung der Profitcenter unter Berücksichtigung der Lieferzeiten kostenoptimal wird.

2 Kombinatorische Auktionen in der Logistik

Für den Bereich der KA in der *Logistik* ist der Ansatz von Caplice [Cap196; CaYo03] für eine mit der Streckenplanung kombinierten Zuweisung von Transportkapazitäten in einem Logistiknetzwerk wegweisend. Dieser verbindet die Bildung der kostengünstigsten Transportkombination durch einzelne Transportanbieter mittels einer KA mit einer einfachen Tourenplanung. Einen ähnlichen Ansatz verfolgen auch Regan und Song [ReSo03]. Sie schlagen einen Spotmarkt für überschüssige bzw. kurzfristig benötigte Transportdienstleistungen vor. Kameshwaran und Narahari [KaNa01] stellen eine sequentielle KA zur Lösung eines Tourenkompositionsproblems vor. In diese Richtung geht auch der Ansatz von Pankratz [Pank03], bei dem die Anreizkompatibilität für Bieter und Nachfrager im Vordergrund steht. Einige Anbieter von Logistiksoftware und Frachtbörsenbetreiber haben bereits einfache KAs in ihre Routenplanungs- und Allokationsverfahren einbezogen³. Im Bereich der Transportdienstleistungen werden seit 1995 die Busrouten der Londoner Region von der *London Regional Transport* per KA versteigert

¹ <http://www.ofcom.org.uk/static/archive/spectrumauctions/index.htm>

² <http://www.fcc.gov>

³ <http://www.nex.com>, <http://www.i2.com>, <http://www.logistics.com>, www.benelog.com, www.combinenet.com

[CaPe05]. Elmaghraby und Keskinocak [ElKe05] dokumentieren die Durchführung einer zwei-stufigen KA zu Beschaffung von Transportkapazitäten durch die US-Baumarktkette *Home Depot* zur Versorgung von über tausend Filialen. Zusammen mit *i2-Technologies* entwickelte das Unternehmen einen flexiblen Auktionsmechanismus, der den Logistikanbietern helfen sollte, ihre individuellen Synergien zwischen Routen mit Hilfe kombinatorischer Gebote auszudrücken. Dabei werden die Bieter durch eine Software bei der Formulierung der Gebote unterstützt, denn die Formulierung der Gebotsbündel und ihre Bewertung stellen ein kombinatorisches Optimierungsproblem dar, das normalerweise nicht manuell gelöst werden kann [Schw05]. Aus diesem Grund ist das hier vorgestellte *ComEx*-System so gestaltet, dass die In- und Outsourcing-Entscheidungen zusammen mit der Gebotsbündelbildung und deren Bewertung mit Hilfe des *DynaRoute*-Tourenoptimierers von *VARLOG*⁴ automatisch durchgeführt werden.

3 Funktionsprinzip und Systembeschreibung von *ComEx*

Im Folgenden geben wir einen Überblick über das Prinzip der KA und stellen die *ComEx*-Systemarchitektur und den Ablauf der Auktions- bzw. Optimierungsprozesse dar. Abb. 1 skizziert die Vorgehensweise für den Austausch von Dienstleistungen zwischen 2 Profitcentern:

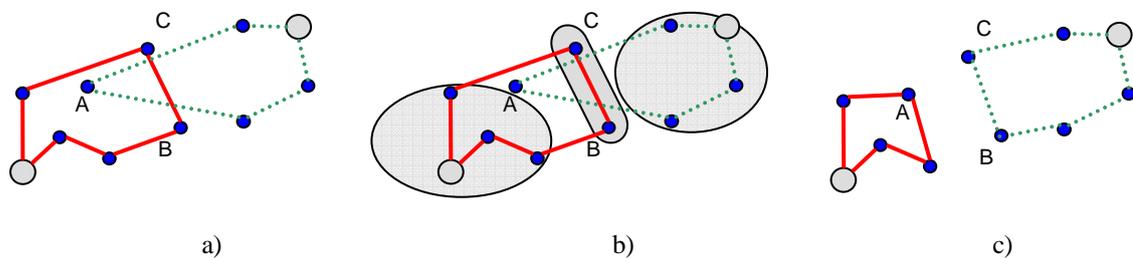


Abb. 1: Austausch von Kunden zwischen benachbarten Profitcentern

Zur Vereinfachung wird angenommen, dass jedes Profitcenter über eine Teilmenge von Kunden verfügt, die bereits einer Lieferroute zugeordnet ist (Abb. 1a). Für jedes Profitcenter wird basierend auf der geografischen Nähe zum Profitcenter entschieden, welche der Kunden weiterhin beliefert werden sollen. Diese Kunden werden dem so genannten Fixbereich zugeordnet (graue Kreise in Abb. 1b). Die übrigen Kunden werden unter Berücksichtigung der geografischen Lage und der Lieferzeiten zu Clustern zusammengeführt (Abb. 1b Punkte B und C). Für jedes Cluster werden die Kosten ermittelt, die das Profitcenter im Falle der Abgabe aller Kunden dieses Clusters an ein anderes Profitcenter einsparen würde. Es findet ein Informationsaustausch zwi-

⁴ <http://www.varlog.de>

sehen den Profitcentern statt, bei dem die Outsourcing-Kandidaten benachbarter Profitcenter als Insourcing-Kandidaten des jeweiligen zentralen Profitcenters deklariert werden. Für jedes Cluster wird sodann von den jeweiligen Profitcentern untersucht, ob die Insourcing-Kandidaten zeitlich und geografisch zu der bestehenden Kundenmenge passen. Ist dies der Fall, werden die Kosten für die Belieferung der neuen Kundenmenge ermittelt. Diese Kosten werden als Gebotspreise für die *ComEx*-Auktion gesetzt. Nachdem jedes Profitcenter seine Gebote abgegeben hat, findet eine KA statt, um eine optimale Allokation von Geboten mit minimalen Lieferkosten im Gesamtsystem zu finden. Nach der Auktion erfährt jedes Profitcenter, welche Kunden von ihm beliefert werden. Die Lieferrouten der Profitcenter werden neu ermittelt (Abb. 1c).

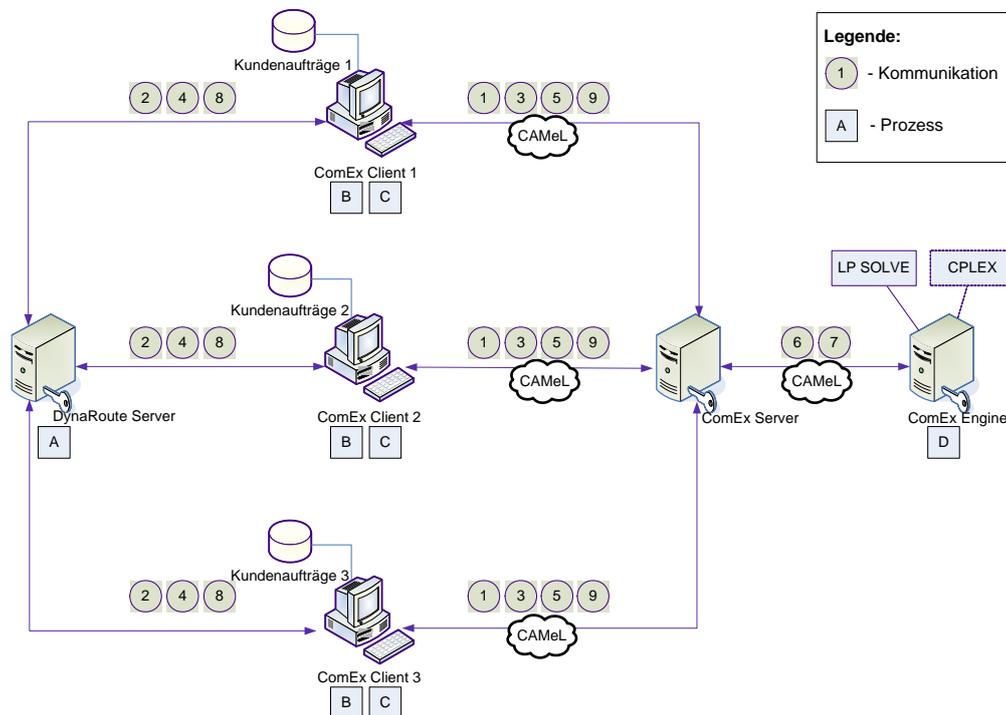


Abb. 2: Kommunikation und Prozessablauf im ComEx-System

Das *ComEx*-System besteht aus den folgenden Komponenten:

- *ComEx*-Client: Jedem Profitcenter steht ein *ComEx*-Client zur Verfügung, der alle Kunden im jeweiligen Liefergebiet verwaltet und mit Hilfe der Kostendaten vom *DynaRoute*-Server festlegt, welche Kunden in den In- und Outsourcing-Prozess einbezogen werden (in Abb. 2 sind das die *ComEx*-Clients 1 bis 3). Zudem wird die Bündelung von Kundenaufträgen zu kombinatorischen Geboten vom *ComEx*-Client durchgeführt. Das Clustering wird in Abschnitt 3.2.1 genauer erläutert.
- *DynaRoute*-Server: Er ist für die Zusammenstellung der Touren und die Schätzung der damit verbundenen Lieferkosten innerhalb eines Profitcenters zuständig.

- *ComEx*-Server: Dieser steuert den Ablauf des gesamten Auktionsprozesses. Dabei sammelt der *ComEx*-Server die In- und Outsourcing-Gebote aller Profitcenter und stößt die Berechnung der optimalen Allokation durch die *ComEx*-Engine an.
- *ComEx*-Engine: Sie ist für die Berechnung der optimalen Allokation zuständig.

Im Folgenden wird der Ablauf einer *ComEx*-Auktion im Zusammenspiel mit den beteiligten Systemkomponenten dargestellt. Insgesamt sind vier Hauptphasen zu unterscheiden⁵:

Initialisierungsphase:

- In der Phase werden Einstellungen wie das Auktionsformat, Anmeldungs- und Zulassungsdaten der Teilnehmer usw. übermittelt ^①. Dabei kommt ein spezielles XML-Format *CAMeL* für die Übertragung zum Einsatz (siehe auch Abschnitt 3.3).

Outsourcingphase:

- Für jedes Profitcenter werden die Kunden unter Berücksichtigung der zeitlichen (ob die Belieferung der Kunden nacheinander möglich ist) und räumlichen (ob die Kunden sich nah zueinander befinden) Bedingungen in Cluster zusammengeführt. Die Outsourcingkandidaten werden, basierend auf ihrer Entfernung vom Zentrum des Profitcenters, ermittelt ^{□B}. Hierauf wird im Abschnitt 3.2.1 näher eingegangen.
- Für jedes der so erzeugten Cluster wird von den Clients eine Anfrage ^② an den *DynaRoute*-Server gestellt, um die Kosten zu ermitteln ^{□A}, die sich durch die Auslagerung einer Kundengruppe einsparen lassen (siehe auch Abschnitt 3.1).
- Die als Outsourcing-Kandidaten bestimmten Cluster werden von jedem Profitcenter an den *ComEx*-Server weitergeleitet. Die Profitcenter erhalten jeweils vom *ComEx*-Server eine Liste aller durch die benachbarten Profitcenter zum Outsourcing freigegebenen Kundencluster, da diese Insourcing-Kandidaten darstellen ^③.

Insourcingphase:

- Anhand der erzeugten Kandidatenliste wird für jedes Profitcenter untersucht, ob sich die jeweiligen Kunden unter Berücksichtigung der zeitlichen und räumlichen Bedingungen zu eigenen Kunden im Outsourcingbereich hinzufügen lassen ^{□C}. Ist

⁵ Das Zusammenspiel zwischen *ComEx* und *DynaRoute* ist ein kritischer Faktor für die Gesamtperformance und sollte bei der Weiterentwicklung des Systems adressiert werden.

dies der Fall, wird eine weitere Anfrage an den *DynaRoute*-Server gestellt ^④, um herauszufinden, zu welchen Konditionen die Belieferung der neuen Kunden erfolgen kann [Ⓐ]. Diese Kosteninformation wird als Gebotspreis des Clusters gesetzt und das entstehende Insourcing-Gebot wird an den *ComEx*-Server weitergeleitet ^⑤.

- Der *ComEx*-Server sammelt die Gebote aller Profitcenter und initiiert die KA durch Aufruf der *ComEx*-Engine ^⑥. Eine optimale Allokation wird ermittelt [Ⓓ] und die Ergebnisse werden den Profitcentern mitgeteilt ^⑦.

Endbewertungsphase:

- Für alle Profitcenter werden die Routenpläne basierend auf der neuen Kundennmenge aktualisiert ^⑧ [Ⓐ], und die Ersparnis im Unternehmen wird ermittelt ^⑨.

3.1 *DynaRoute*-Server

Die Tourenberechnung bei den Clients findet mittels der Software *DynaRoute* [WeSW05] statt und löst Erweiterungen des „*Time Dependent-Stochastic Capacitated Vehicle Routing Problem with Time Windows*“ (TD-CVRP-TW). Die hohe Komplexität des Problems entsteht aufgrund der Ergänzung des bereits NP-vollständigen „*Vehicle Routing Problems*“ um zeitabhängige Fahrzeiten und der Einführung von Zeitfensterkategorien. Die Zeitfenster sind in die vier Kategorien niedrig, mittel, hoch und strikt eingeteilt, aus denen eine Bewertungsfunktion für die prognostizierte Nichteinhaltung der Zeitfenster abgeleitet wird.

Grundlage des Lösungsverfahrens ist die von [Wend95] vorgeschlagene Kombination der Metaheuristik Simulated Annealing (SA) und Genetischer Algorithmen (GA). Cooperative Simulated Annealing (COSA) übernimmt dabei die Populationsidee und die Idee des Informationsaustausches zwischen einzelnen Individuen von GAs und kombiniert diese mit der SA-Heuristik. Diese steuert lokale Suchoperatoren entsprechend stochastischer Annahmeregeln für neue Lösungen. Im Laufe der letzten Jahre wurde das Lösungsverfahren an die Anforderungen der Praxis angepasst und es wurden folgende Ergänzungen eingeführt:

- *Lösungsraumkompression*

Viele Teillösungen können mit hoher Wahrscheinlichkeit als nicht optimal identifiziert werden, und ihr Ausschluss ermöglicht eine Komplexitätsreduktion. Während entsprechende Lösungsbereiche bei bekannten Optimierungskriterien und Restriktionen häufig mit Hilfe einfacher Verfahren ausgeschlossen werden können, ist dies hier nicht

möglich. *DynaRoute* nutzt auf Ant-Systemen [BuHS99] basierende Verfahren zur Identifikation stochastischer Muster, die eine effiziente Steuerung der Suche ermöglichen. Weiterhin können im Rahmen der Restriktionsspezifikation Abbruchkriterien für die Lösungsbewertung definiert werden, die den Lösungsraum weiter einschränken.

- *Tabu-Listen*

Tabu-Listen [Glov86] erlauben die temporäre Sperrung bereits analysierter Lösungen, bzw. Lösungsregionen und damit die gezielte Diversifikation des Suchprozesses. Die Nutzung von Tabu-Listen im Rahmen der SA-Suche hat sich als ein außerordentlich effizienter Ansatz erwiesen [LiLi02].

- *Approximationsverfahren zur Lösungsbewertung*

Die exakte Bewertung einer neuen Lösung ist häufig nicht notwendig. Entsprechend helfen Approximationsverfahren, die Rechenzeit zu reduzieren, ohne die Lösungsqualität zu senken. *DynaRoute* nutzt dynamische Approximationsverfahren, die in Abhängigkeit des Suchverlaufs den Detaillierungsgrad der Berechnungen intensivieren.

Basierend auf den einzelnen Lösungsbewertungen durch die Heuristik werden Bewertungen der von *ComEx-Client* spezifizierten Teilmengen vorgenommen. Der skizzierte Lösungsansatz für das Tourenplanungsproblem bietet die Möglichkeit, nach der geringfügigen Veränderung eines Tourenplans diesen durch kurzes Reaktivieren des Berechnungsprozesses schnell zu optimieren. Ist dies geschehen, ermöglicht ein einfacher Vergleich der Kosten der Ausgangslösung und der durch Hinzufügen neuer Kunden veränderten Lösung eine Aussage über die Kosten der Integration neuer Kunden. Das Risiko einer Fehlbewertung durch den Vergleich suboptimaler Lösungen ist auf Grund des populationsbasierten Algorithmus gering, da der Vergleich über die k -besten Lösungen der beiden Populationen einzelne Fehler relativiert.

3.2 *ComEx-Client*

Jedes Profitcenter entscheidet, welche Kunden es auf jeden Fall selbst beliefern will und welche von anderen Profitcentern übernommen werden können. Im ersten Fall geht es um die Kunden, die geografisch nah am Profitcenter i liegen. Solche Kunden bilden, wie bereits oben beschrieben, den Fixbereich P_i^f . Alle anderen Kunden werden dem Outsourcingbereich P_i^o zugeordnet.

3.2.1 Clustering und Outsourcing

Das Ziel des Clusterings besteht darin, die Komplexität des vorliegenden KA-Problems zu verringern und Synergieeffekte zwischen einzelnen Lieferaufträgen zu identifizieren. Um dies zu erreichen, werden die Kunden nach bestimmten zeitlichen, räumlichen und kapazitiven Kriterien bewertet und anschließend zu *Clustern* $C_i^y \subseteq C_i$, $i, y \in \mathbb{N}$ gruppiert, wobei C_i die Menge aller Cluster des Profitcenters i darstellt. Ein Cluster wird dabei als ein ungerichteter Pfad verstanden, der ein Teilstück einer Route bilden kann, welche durch die spätere Routenplanung bestimmt wird. Das Clustering verläuft also unabhängig von der eingesetzten Routenplanung und hat die Aufgabe, geografisch nahe gelegene Kunden so zusammenzufassen, dass sie unter Einhaltung von oben erwähnten Kriterien mit einem Fahrzeug beliefert werden können.

Zu Beginn des eingesetzten Clustering-Verfahrens enthält ein Pfad jeweils einen Kunden. Um möglichst viele Pfade zu kombinieren, werden zwei verschiedene Strategien eingesetzt, die in aufeinander folgenden Phasen stattfinden. Während in der ersten Phase versucht wird, Pfade durch eine Verbindung an deren Anfang oder deren Ende mit einem anderen Pfad zu verlängern, wird in der zweiten Phase überprüft, ob sich Pfade in andere eingliedern lassen.

Sei E die Menge der Kanten, V die Menge der Knoten in einem ungerichteten Graphen G , $e_{new} = \{a_k, a_l\}$ mit $a_k, a_l \in V, k \neq l$ ein Kundenpaar mit Distanz $|e_{new}| \in \mathbb{R}^+$ zwischen den Kunden in einem Profitcenter i . Um den Algorithmus der ersten Phase zu veranschaulichen, ist in Abb. 3 dessen Pseudocode dargestellt:

ClusteringInclude

```

1    $E' := \{e_{new} = \{a_k, b_l\} \mid a_k \in C_i^a \wedge b_l \in C_i^b \wedge e_{new} \in E \wedge |e_{new}| \leq r\};$ 
2   Sortiere  $E'$  nach steigenden Distanzen  $|e_{new}|$ ;
3   for all  $\{a_k, b_l\} \in E'$  (*in sortierter Reihenfolge*) do
4       if  $((a_k \text{ Endknoten} \wedge b_l \text{ Endknoten}) \wedge (a_k \text{ und } b_l \text{ nicht im selben Cluster}))$ 
5           Verbinde Cluster  $C_i^a$  und  $C_i^b$  durch  $e_{new}$ ;
6           if (Kombiniertes Cluster valide)
7               Speichere kombiniertes Cluster  $C_i^y$ ;

```

Abb. 3: Clustering-Algorithmus zum Zusammenlegen von Streckenabschnitten (Phase 1)

Zunächst wird die Menge E' aller möglichen Verbindungen zwischen zwei Kunden a_k und a_l eines Profitcenters bestimmt. Dabei werden nur die Kunden berücksichtigt, deren Distanz zueinander kleiner ist als eine maximale Entfernung r (Zeile 1). Die Kundenpaare werden nach aufsteigender Entfernung sortiert (Zeile 2) und der Reihe nach betrachtet (Zeile 3). Falls beide

Kunden a_k und b_l Cluster-Endpunkte sind und sich zugleich nicht im selben Cluster befinden (Zeile 4), wird ein kombiniertes Cluster erzeugt, indem zwischen a_k und b_l eine Kante e_{new} eingefügt wird (Zeile 5). Anschließend wird die Validität des neu erstellten Clusters überprüft, die genau dann gegeben ist, wenn die im Cluster enthaltenen Kunden in mindestens einer der beiden Reihenfolgen unter Einhaltung der Lieferzeitfenster von einem Fahrzeug beliefert werden können. Falls Validität vorliegt, wird das Cluster C_i^y abgespeichert (Zeilen 6, 7).

Die zweite Phase zielt darauf ab, bestehende Cluster in andere einzugliedern. Dies bedeutet, dass der Pfad innerhalb eines Clusters unterbrochen wird, um die Kunden eines anderen Clusters zu beliefern. Erst wenn dies geschehen ist, wird die Abarbeitung des ursprünglichen Pfades fortgesetzt. Für diese Eingliederung eines Clusters in ein anderes sind zwei verschiedene Varianten denkbar, die in Abb. 4 dargestellt sind. Zunächst wird eine Verknüpfung zwischen einem Endknoten des ersten Clusters mit dem inneren Knoten des zweiten Clusters festgelegt (Abb. 4b). Nun wird der Pfad des zweiten Clusters zwischen a_2 und a_3 (Abb. 4a) bzw. a_1 und a_2 (Abb. 4c) unterbrochen und zwei mögliche Verbindungen werden bestimmt, die die Eingliederung des ersten Clusters ($\{b_1, b_2\}, \{b_2, b_3\}$) in das zweite ermöglichen. Analog zur ersten Phase dürfen auch hier die neu entstehenden Verbindungen die Kantenlänge r nicht übersteigen.

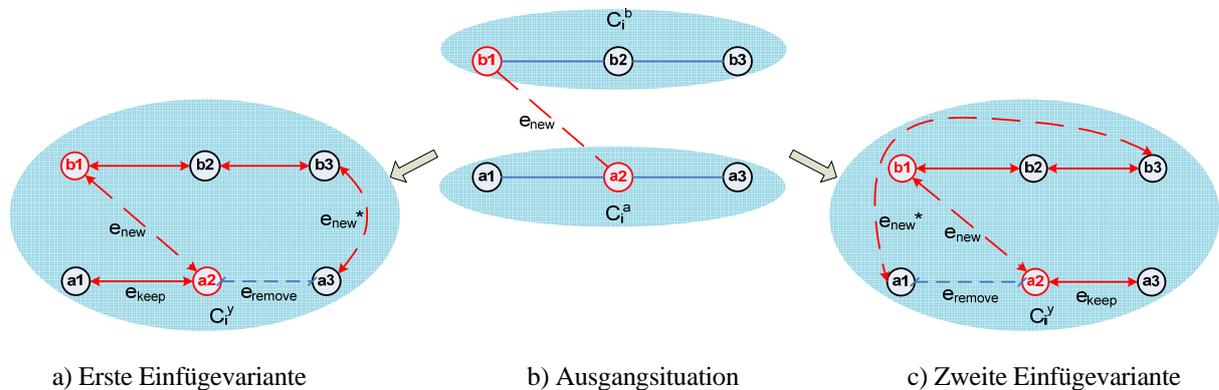


Abb. 4: Zwei Varianten für die Clusterzusammenlegung

In Abb. 5 wird der Algorithmus der zweiten Phase dargestellt. Zunächst wird eine Menge E'' generiert, die anfangs leer ist (Zeile 1). Dafür wird jedes Element der Menge E' daraufhin geprüft, ob es als erste der beiden Verbindungen zwischen zwei verschiedenen Clustern in Frage kommt (Zeilen 2 und 3). Dies ist genau dann der Fall, wenn sich die beiden Kunden a_k und b_l nicht bereits im selben Cluster befinden und wenn einer der beiden Kunden einen Cluster-Endpunkt darstellt, der andere jedoch nicht (Zeile 4). Nach der Bestimmung der beiden Varianten für die zweite Verbindung werden die Kantentupel zur Menge E'' hinzugefügt (Zeilen 5, 6).

ClusteringInclude2

```
1    $E'' := \emptyset$  ;
2    $E' := \{e_{new} = \{a_k, b_l\} \mid e_{new} \in E \wedge a_k \in C_i^a \wedge b_l \in C_i^b \wedge |e_{new}| \leq r\}$  ;
3   for all  $e_{new} = \{a_k, b_l\} \in E'$  do
4       if  $((a_k \text{ Endknoten} \wedge b_l \neg \text{Endknoten}) \wedge (a_k \text{ und } b_l \text{ nicht im selben Cluster}))$ 
5           for all  $e_{keep} = \{b_l, b_{l^*}\}$  mit  $l \neq l^* \wedge b_l \in e_{new}$  do
6               Bestimme  $e_{new}^*, e_{remove}, \Delta = |e_{new}| + |e_{new}^*| - |e_{remove}|$ ,  $E'' \leftarrow (e_{new}, e_{keep})$  ;
7   Sortiere  $(e_{new}, e_{keep}) \in E''$  nach steigendem  $\Delta$  ;
8   for all  $(e_{new}, e_{keep}) = (\{a_k, b_l\}, \{b_l, b_{l^*}\}) \in E''$  (*in sortierter Reihenfolge*) do
9       if  $(a_k, b_l, b_{l^*}$  nicht im selben Cluster)
10          Gliedere  $C_i^a$  gemäß der Kanten  $e_{new}, e_{new}^*$  und  $e_{keep}$  in  $C_i^b$  ein;
11          if (Kombiniertes Cluster valide)
12             Speichere kombiniertes Cluster  $C_i^y$  ;
```

Abb. 5 Clustering-Algorithmus zum Eingliedern von Streckenabschnitten (Phase 2)

Anschließend werden die Elemente der Menge E'' nach steigenden Distanzsummen sortiert (Zeile 7) und es wird der Reihe nach geprüft, ob die beiden repräsentierten Verbindungen zu einer validen Eingliederung eines Clusters in ein anderes Cluster führen (Zeilen 8 bis 10). Wenn die Validität des kombinierten Clusters gegeben ist, wird dieses Cluster abgespeichert (Zeilen 11 und 12).

Sei $I \in \mathbb{N}$ die Anzahl von Profitcentern im Unternehmen. Für jedes $i \in I$ werden die Kunden entsprechend ihrer Entfernung vom Zentrum des Profitcenters in zwei Gruppen eingeteilt: Ein bestimmter Prozentsatz der weiter entfernten Kunden wird dem Outsourcingbereich P_i^o zugeordnet, die übrigen näher am Zentrum liegenden Kunden bilden den Fixbereich P_i^f . Innerhalb jedes Profitcenters i wird die Menge C_i gebildet, die alle im Profitcenter erstellten kombinierten Cluster in P_i^o enthält. Des Weiteren wird C_i an *ComEx*-Server weitergeleitet und dort zu $C^* = \bigcup_{i=1}^I C_i$ gruppiert. Anschließend erhalten alle Profitcenter die Menge C^* .

3.2.2 Gebotserstellung und Insourcing

In der Literatur sind verschiedene Gebotssprachen und damit verbundene Arten der Gebotslogik beschrieben [Nisa99]. Das *ComEx*-System benutzt eine OR-of-XOR Gebotslogik. Dies bedeutet, dass die XOR-Gebotsbündel jeweils eine Auswahl eines der alternativen Gebote erlauben, während bei den OR-Gebotsbündeln mehrere Gebote gleichzeitig befriedigt werden können.

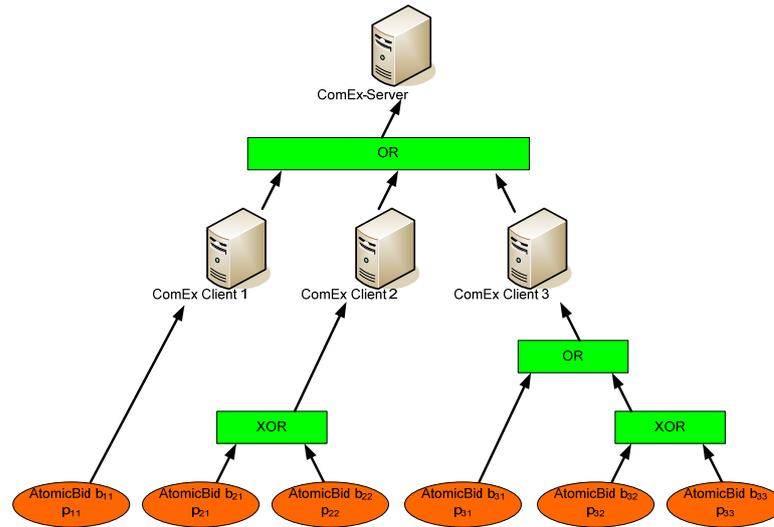


Abb. 6 Gebotslogik in ComEx

Auf Grund seiner Größe wird C^* von jedem Profitcenter anhand eines von uns entwickelten Verfahrens zu C_i^* verkleinert, indem die am weitesten entfernt liegenden kombinierten Cluster anderer Profitcenter eliminiert werden. Die Menge C_i^* wird von Profitcentern für die Erstellung der Kombinationen aus eigenen kombinierten Clustern mit kombinierten Clustern anderer Profitcenter (nach dem gleichem Verfahren wie in 3.2.1) verwendet. Anschließend werden die Gebote wie folgt gebildet: Sei $M_i \in \mathbb{N}$ die Anzahl der für Profitcenter $i \leq I$ zur Verfügung stehenden kombinierten Cluster $C_i^y \in C_i^*$, dann ist $b_{ij} = C_i^\alpha \wedge C_i^\beta \wedge \dots \wedge C_i^\delta$, $\alpha \neq \beta \neq \delta \wedge \alpha, \beta, \delta \leq M_i$ ein *AtomicBid* vom Profitcenter i . Für jedes b_{ij} wird eine Anfrage an den *DynaRoute*-Server gestellt, um die Kosten p_{ij} für die Belieferung aller Kunden in b_{ij} festzustellen. Hier kann zunächst auf die spieltheoretische Diskussion über wahrheitsgemäßes Bieten verzichtet werden, da die berechneten Lieferkosten direkt als Gebotspreise übernommen werden und durch Verwendung identischer *ComEx*-Clients bei allen Profitcentern nicht manipulierbar sind [Schw05].⁶ Sei B_i eine Zusammensetzung aller Gebote vom Profitcenter i , die mit Hilfe logischer Operationen *OR* bzw. *XOR* verknüpft sind. Verschiedene *AtomicBids* können dieselben kombinierten Cluster C_i^y enthalten, z.B. *AtomicBids* b_{21} und b_{22} in Abb. 6. In diesem Fall werden die beiden Gebote mit *XOR* verknüpft: $B_2 = b_{21} \oplus b_{22}$. Sind die Teilmengen an den in *AtomicBids* vorkommenden Clustern disjunkt, so wird *OR*-Operation angewendet, z.B. $B_3 = b_{31} \vee b_{32} \oplus b_{33}$.

⁶ Da *DynaRoute* ein stochastisches Verfahren ist, stellen die Lieferkosten eine Verteilung dar, so dass Anreizkompatibilität nur näherungsweise gewährleistet ist und eine weitere Diskussion dieses Punktes erfolgen sollte.

$B^* = B_1 \vee B_2 \vee \dots \vee B_I$ abgespeichert (Abb. 6). Alle SOAP-Nachrichten sind bei der Web-Übertragung per SSL-Technologie verschlüsselt. Nachdem die Gebote aller Profitcenter in der Datenstruktur B^* gespeichert worden sind, werden sie an die *ComEx*-Engine weitergeleitet, und die KA wird initiiert. Die optimale Allokation wird mittels der *ComEx*-Engine ermittelt und über den *ComEx*-Server an alle Clients übermittelt. Des Weiteren ist der *ComEx*-Server für die Berechnung der durch die Auktion im Gesamtsystem erzielten Ersparnisse zuständig.

3.4 *ComEx*-Engine

Wie im vorherigen Abschnitt beschrieben, werden die vom *ComEx*-Server gesammelten Gebote B_i von allen *ComEx*-Clients i in Form von *CAMEL*-Nachrichten an die *ComEx*-Engine weitergeleitet. Diese muss, um die optimale Allokation zu finden, das Combinatorial Auction Problem (CAP) [dVVo01] lösen. Seien z die Lieferkosten im Unternehmen, M die Menge aller kombinierten Cluster im System, J_i die Anzahl der Gebote vom Profitcenter i und $b_{ij}(C_i^y) = 1$ wenn das AtomicBid b_{ij} das Cluster C_i^y enthält, dann lautet das CAP für die Tourenoptimierung:

$$\min z = \sum_{i=1}^I \sum_{j=1}^{J_i} p_{ij} x_{ij} \quad (1)$$

u.d.N.
$$\sum_{i=1}^I \sum_{j=1}^{J_i} b_{ij}(C_i^y) \cdot x_{ij} = 1 \quad \forall C_i^y \in C^* \wedge x_{ij} \in \{0,1\} \quad (2)$$

$$b_{ij}(C_i^y) = \begin{cases} 1, & \text{wenn } C_i^y \text{ in } b_{ij} \text{ vorkommt} \\ 0, & \text{sonst} \end{cases} \quad \forall C_i^y \in C^* \quad (3)$$

Die Akzeptanzvariable x_{ij} zeigt an, ob ein Gebot den Zuschlag erhalten soll. Gleichzeitig definiert sie die hier notwendige Ganzzahligkeitsbedingung. Mit Gleichung (2) wird sichergestellt, dass alle kombinierten Cluster in der Lösungsallokation vorkommen. Für die Berechnung des CAP wird das Software-Paket *LP_SOLVE* 5.5⁸ verwendet. Das Paket liefert in akzeptabler Zeit eine exakte Lösung des hier beschriebenen *NP*-harten Problems bei bis zu einigen Hundert Geboten. Für Problemstellungen mit höherer Komplexität ist ein Umstieg auf Heuristiken sinnvoll. In [ScGu05] wurden daher heuristische Verfahren für das CAP [ScSR03] auf ihre Effizienz (Qualität der Lösung) und Schnelligkeit im Vergleich zu exakten Verfahren getestet.

⁸ <http://www.geocities.com/lpsolve>

4 Der Einfluss des Clustering-Prozesses auf die Gesamtlieferkosten

4.1 Ziel der Simulation und Simulationseinstellungen

Ziel des Einsatzes von *ComEx* ist die Reduktion der Gesamtlieferkosten im Unternehmen, die durch die Verkürzung der Gesamtlieferstrecke erzielt wird. Dieses Ziel kann nur erreicht werden, wenn eine geeignete Strategie zur Erstellung von Geboten für die Übernahme und Abgabe der Streckenabschnitte durch die Profitcenter angewendet wird. Im Rahmen unseres Systems kann dieses durch zwei Faktoren gesteuert werden: zum einen durch die Begrenzung des maximal zulässigen Abstandes r zwischen den Kunden innerhalb eines Clusters, zum zweiten durch die Festlegung des prozentualen Anteils a der Kunden, die dem Fixbereich P_f eines Profitcenters zugeordnet werden. Neben dem Einfluss dieser Faktoren wurde erforscht, wie sich die Änderung der Maximalzahl der zulässigen Gebote pro Profitcenter auf die Gesamtlieferkosten auswirkt.

Es wurde folgendes Testszenario gewählt: Die Simulationen basieren bezüglich geografischer Lage und Lieferzeiten auf Realdaten unseres Partnerunternehmens aus der Nahrungsmittellindustrie, wobei 3 Profitcenter mit je ca. 100 Kunden für die Auswertung zu Grunde gelegt wurden. Das CAP wurde mittels der *ComEx*-Engine basierend auf LP_SOLVE exakt gelöst. Der *DynaRoute*-Server lieferte die Routen und Kostenschätzungen basierend auf der oben erwähnten *COSA*-Heuristik⁹.

Für die Simulation wurden die relevanten Parameter innerhalb folgender Schranken variiert:

- Maximalabstand r zwischen den Kunden eines Clusters: 2–20 km (2 km Schritte)
- Anteil a der Kunden im Outsourcingbereich: 20% - 100% (20% Schritte)

Die Berechnung wurde mehrfach wiederholt, um die stochastischen Effekte zu reduzieren, die durch den Einsatz des heuristischen Verfahrens beim *DynaRoute*-Server entstanden.

4.2 Ergebnisse

Wie Abb. 8 zeigt, lässt sich bei einem Maximalabstand von 10 km zwischen den Kunden eines Clusters und bei einem Anteil der Kunden im Outsourcingbereich von 40% die Gesamtlieferstrecke am stärksten verkürzen¹⁰. Ein zu gering angesetzter Maximalabstand bei der Gebotsbil-

⁹ Es wurden folgende Einstellungen für das Cooperative Simulated Annealing Verfahren verwendet: Starttemperatur 1.000.000, 1% Abkühlrate und eine Populationsgröße von 20 Elementen.

¹⁰ Hier wird zur Vereinfachung die Lieferstrecke als Schätzer für die Lieferkosten verwendet.

ung führt zur Erstellung von Clustern, die aus nur wenigen Kunden bestehen. Dies bedeutet, dass die Abgabe bzw. Übernahme von solchen Clustern keine Synergieeffekte bewirken kann.

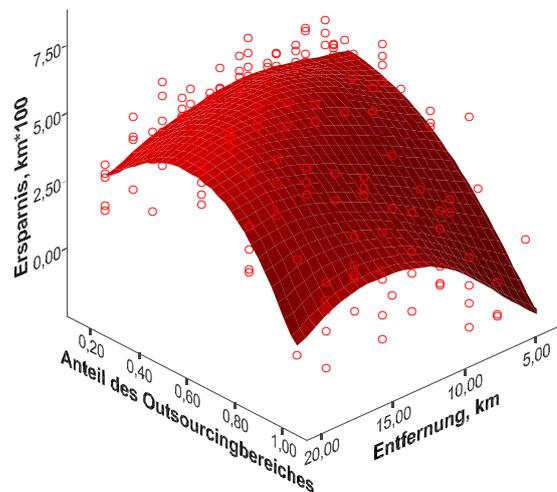


Abb. 8: Lieferkosteneinsparung durch den Einsatz von *ComEx*

Wird andererseits ein zu hoch angesetzter Maximalabstand r bei der Clustererstellung erlaubt, werden sehr komplexe Teilrouten für das Outsourcing erstellt. Solche größeren Cluster lassen sich schwer in die Touren anderer Profitcenter integrieren.

Ein zu großer Anteil von Kunden im Outsourcingbereich bei der Gebotsbildung führt zur Auslagerung vieler Kunden, die nahe zum eigenen Profitcenter liegen und eigentlich zu günstigen Konditionen von diesem beliefert werden können. Ein zu kleiner Anteil von Kunden eines Profitcenters im Outsourcingbereich verhindert die Freigabe zur Auslagerung geeigneter Kunden. In einer weiteren Simulation wurde beispielhaft untersucht, ob sich die zuvor identifizierten minimalen Lieferkosten (bei $r=10$ km, $a=40\%$) durch eine Änderung der maximal erlaubten Anzahl von Geboten weiter reduzieren lassen. Hierfür zeigt Abb. 9 die Gesamtlieferstrecke in Abhängigkeit zur maximal erlaubten Anzahl von Geboten pro Profitcenter:

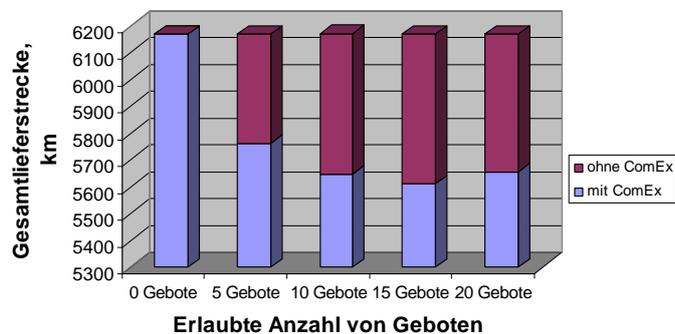


Abb. 9: Gesamtlieferstrecke in Abhängigkeit zu erlaubter Anzahl von Geboten

Die Erstellung einer höheren Anzahl von Geboten pro Profitcenter und somit die Erhöhung des Rechenaufwandes, der bei der *ComEx*-Engine anfällt, hebt das Einsparpotential im Unterneh-

men bis zur maximal zulässigen Anzahl von fünfzehn Geboten pro Profitcenter. Eine weitere Erhöhung hat jedoch einen negativen Effekt auf die Einsparungen.

5 Fazit

Das vorgestellte Softwaresystem *ComEx* wurde zum auktionsbasierten innerbetrieblichen Austausch von Transportdienstleistungen in Verbindung mit der Tourenoptimierungssoftware *DynaRoute* eingesetzt. Der Fokus der vorgestellten Untersuchungen lag auf der Ermittlung einer optimalen Strategie zur Erstellung von Geboten für die Übernahme bzw. Abgabe von Streckenabschnitten zwischen benachbarten Profitcentern, so dass für das gesamte Unternehmen Kostenreduktion entsteht. Anhand realer Lieferdaten eines Unternehmens aus der Nahrungsmittelindustrie wurden ein optimaler Maximalabstand von 10 km zwischen den Kunden und ein optimaler Anteil der Kunden im Fixbereich von 40% bei der Bildung optimaler In- bzw. Outsourcing-Gebote für die Berechnung der optimalen Allokation mit Hilfe der KA ermittelt. Eine höhere Bedeutung als im innerbetrieblichen Einsatz ist der Verwendung des *ComEx*-Verfahrens im zwischenbetrieblichen Bereich beizumessen, z.B. als allgemeine kombinatorische Transportbörse. Eine solche Realisierung eines Logistikmarktplatzes zwischen eigenständigen Unternehmen wirft jedoch in erhöhtem Maße die Frage nach anreizkompatiblen Gebotsbewertungsmechanismen, wie bspw. dem *Vickrey-Clarke-Groves*-Mechanismus, auf. Diese Problematik wurde im Zusammenhang mit dem hier diskutierten Anwendungsfall durch den Einsatz gleichartiger automatisierter Gebotsmechanismen bei allen Teilnehmern umgangen. Offen bleibt jedoch auch hier die Frage nach der gerechten und anreizkompatiblen Ersparnisverteilung zwischen Profitcentern. Diese wird zusammen mit dem Einsatz des *ComEx*-Systems als Transportbörse bei einem größeren Logistikdienstleister den nächsten Forschungsgegenstand darstellen.

Literaturverzeichnis

- [BuHS99] *Bullnheimer, B.; Hartl, R. F.; Strauss, C.:* An improved Ant System Algorithm for the Vehicle Routing Problem. In: *Annals of Operations Research* 89 (1999), S. 319-328.

- [CaPe05] *Cantillon, Estelle; Pesendorfer, Martin: Auctioning Bus Routes: The London Experience.* In: *Steinberg, R.; Shoham, Y.; Cramton, P. (Hrsg): Combinatorial Auctions.* MIT-Press: Cambridge, Massachusetts 2005. S. 573-592.
- [Capl96] *Caplice, Christopher: An Optimization Based Bidding Process: A New Framework for Shipper-Carrier Relationships.* Dissertation, Massachusetts Institute of Technology 1996.
- [CaSh03] *Caplice, Christopher; Sheffi, Yossi: Optimization-Based Procurement for Transportation Services.* In: *Journal of Business Logistics* 24 (2003) 2, S. 109-128.
- [dVVo01] *de Vries, Sven; Vohra, Rakesh: Combinatorial Auctions: A Survey.* In: *INFORMS Journal on Computing* 15 (2001) 3, S. 284-309.
- [ElKe05] *Elmaghraby, Wedad; Keskinocak, Pinar: Combinatorial Auctions in Procurement.* In: *Harrison, T. P.; Lee, H. L.; Neale, J. J. (Hrsg): The Practice of Supply Chain Management: Where Theory and Application Converge.* Springer, Berlin 2003, S. 245-258.
- [Glov86] *Glover, Fred: Future Paths for Integer Programming and Links to Artificial Intelligence.* In: *Computers and Operations Research* 13 (1986) 5, S. 533-549.
- [KaNa01] *Kameshwaran, S; Narahari, Y.: Auction Algorithms for Achieving Efficiencies in Logistics Marketplaces.* In: *Proceedings of the International Conference on Energy, Automation and Information Technology, 2001.*
- [LiLi02] *Li, Haibing; Lim, Andrew: Local search with annealing-like restarts to solve the vehicle routing problem with time windows.* ACM Press: Madrid, Spain, S. 560-565, 2002.
- [McMi95] *McMillan, John: Why Auction the Spectrum?* In: *Telecommunications Policy* 19 (1995), S. 191-199.
- [Nisa99] *Nisan, Noam: Bidding and Allocation in Combinatorial Auctions.* In: *Proceedings of the 2nd ACM Conference on Electronic Commerce (ACM-EC99), 1999.*

- [Pank03] *Pankratz, Giselher*: Zweiseitige kombinatorische Auktionen in elektronischen Transportmärkten – Potenziale und Probleme. Working Paper: FernUniversität Hagen, Fachbereich Wirtschaftswissenschaft: ftp://ftp.fernuni-hagen.de/pub/fachb/wiwi/winf/forschung/publi/gp_p6.pdf, 2003, Abruf am: 30.6.2006.
- [ReSo03] *Regan, Amelia; Song, Jiongjiong*: An Auction Based Collaborative Carrier Network. Unter Begutachtung in: Transportation Research Part E: Logistics and Transportation Review 2003.
- [Schw05] *Schwind, Michael*: Design of Combinatorial Auctions for Allocation and Procurement Processes. In: Proceedings of the 7th International IEEE Conference on E-Commerce Technology 2005, München 2005, S. 391-395.
- [ScGu05] *Schwind, Michael; Gujo, Oleg*: An Agent-based Simulation Environment for a Combinatorial Grid Scheduler. Working Paper: Institute of Information Systems, Johann Wolfgang Goethe University: Frankfurt, Germany, <http://www.wiwi.de/publikationen/AnAgentBasedSimulationEnviron1531.pdf>, 2005, Abruf am: 30.6.2006.
- [ScSR03] *Schwind, Michael; Stockheim, Tim; Rothlauf, Franz*: Optimization Heuristics for the Combinatorial Auction Problem. In: Proceedings of the Congress on Evolutionary Computation CEC 2003, Canberra 2003, S. 1588-1595.
- [ScWS04] *Schwind, Michael; Weiss, Kilian; Stockheim, Tim*: CAMEL - Eine Meta-Sprache für Kombinatorische Auktionen. Working Paper: Institut für Wirtschaftsinformatik, Johann Wolfgang Goethe Universität: Frankfurt, <http://www.wiwi.de/publikationen/CAMELEineMetaSprachefuerKombin1068.pdf>, 2004, Abruf am: 30.6.2006.
- [Wend95] *Wendt, Oliver*: Tourenplanung durch Einsatz naturanaloger Verfahren. Gabler Verlag, Wiesbaden 1995.
- [WeSW05] *Wendt, Oliver; Stockheim, Tim; Weiss, Kilian*: Intelligente Tourenplanung mit DynaRoute. In: WIRTSCHAFTSINFORMATIK 47 (2005) 2, S. 135-140.

Prozessverantwortung und Dokumentation als Determinanten der Effizienz und Qualität der Transportplanung

Ein Modell auf Basis einer empirischen Untersuchung unter 1000 deutschen Unternehmen

Volker Lanninger, Tim Stockheim, Oliver Wendt

Lehrstuhl für Wirtschaftsinformatik und Operations Research
Chair of Business Information Systems and Operations Research
Technische Universität Kaiserslautern
67663 Kaiserslautern
{lanninger, stockheim, wendt}@bisor.de

Zusammenfassung

Stetig steigende Transportkostenanteile an den Gesamtausgaben kleiner und mittlerer Unternehmen motivieren eine Untersuchung möglicher Optimierungspotenziale im Prozess der Transportplanung. Generell stellt die Schaffung geeigneter organisatorischer Rahmenbedingungen den Ausgangspunkt zur weiteren Kostensenkung und Qualitätsverbesserung dar. Der vorliegende Beitrag begründet, dass nur darauf aufbauend ein erfolgreicher Einsatz von Optimierungssoftware möglich ist.

Im Rahmen eines Forschungsprojekts zur Weiterentwicklung von Tourenplanungssoftware wird anhand zweier Regressionsmodelle untersucht, wie die Vergabe eindeutiger Prozessverantwortung sowie eine professionelle Dokumentation des Planungsprozesses auf die wahrgenommene Kosteneffizienz und Prozessqualität wirken. Zum Abschluss wird gezeigt, wie auf dieser organisatorischen Basis weitere Maßnahmen zur Verbesserung der Transportplanung ergriffen werden können.

1 Einleitung

Transportleistungen entwickeln sich für viele Unternehmen zu einem immer wichtiger werdenden Erfolgsfaktor, was beispielsweise an der Beförderungsleistung deutscher Unternehmen im

Inland von ca. 271,8 Mrd. Tonnenkilometer im Jahr 2005 auf der Straße¹ [Bund06] abzulesen ist. In vielen Branchen, wie z.B. Entsorgung, Großhandel oder Nahrungsmittelproduktion, haben die Transportkosten einen Anteil von weit über zehn Prozent an den Gesamtkosten [Stra05] und entwickeln sich zu einem wesentlichen Kostentreiber. Steigende Kraftstoffpreise und die Einführung der Autobahn-Maut² deuten zudem auf eine weitere Erhöhung der Transportkosten hin [Stat04].

Vor diesem Hintergrund hat das in diesem Beitrag beschriebene Modell das Ziel, die Wirkung von klarer Prozessverantwortung und Dokumentation auf die Kosteneffizienz und Qualität der Transportplanung aufzuzeigen. Die zur Evaluation verwendeten Daten wurden im Rahmen der Studie „Transportplanung der Zukunft“ [WKSL06] erhoben. Diese analysiert die derzeit vorliegenden Prozess- und Kostenstrukturen in der Transportplanung, identifiziert die daraus ableitbaren Optimierungspotenziale und veranschaulicht den Erfolgsbeitrag von moderner Tourenplanungssoftware.

Das Forschungsprojekt, in dessen Rahmen die Studie entstand, zielt auf die Weiterentwicklung von Tourenplanungssoftware unter Berücksichtigung stochastischer Fahrzeiten [WeSW05]. Gerade im Hinblick auf den Einsatz von Softwaresystemen werden eine gute Dokumentation der Prozesse sowie klare Prozessverantwortung als essentiell angesehen [WeSW05; Tosc03]. Das vorgestellte Modell untersucht den Einfluss von Dokumentation und Prozessverantwortung auf die Kosteneffizienz und Qualität des Planungsprozesses in der Anwendungsdomäne Transportplanung. Es wurde analog zu bestehenden Modellen anderer Domänen entwickelt (siehe dazu Abschnitt 2), auf Basis mehrerer Experteninterviews an die Transportplanung angepasst, und wird in diesem Beitrag anhand empirischer Daten evaluiert.

In vergleichbaren Studien, wie z.B. der von der Bundesvereinigung für Logistik durchgeführten Untersuchung „Trends und Strategien in der Logistik“ [Stra05], finden sich zur oben erwähnten allgemeinen Forderung nach Prozessdokumentation und eindeutiger Prozessverantwortung keine Aussagen. Diese Lücke adressiert der vorliegende Beitrag speziell für kleine und mittlere Unternehmen und leitet daraus allgemeine Empfehlungen zur Prozessoptimierung durch Tourenplanungssoftware ab.

¹ Gegenüber dem Jahr 2004 mit ca. 266,9 Mrd. Tonnenkilometer (Inland) auf der Straße bedeutet dies eine Steigerung von etwa 2 % [Bund05].

² Nach Einführung der LKW-Maut werden die Unternehmen rund neunmal höher belastet als durch die Euro-Vignette [Bund05b].

1.1 Kurzbeschreibung der zugrunde liegenden Studie

Der Fragebogen wurde an die für den Transportplanungsprozess Verantwortlichen in je 250 deutschen Unternehmen (mit Transportbedarf) der Branchen Bauwesen, Entsorgung, Nahrungsmittelindustrie und Speditionen gesendet. Kriterium bei der Auswahl der Teilnehmer war, ein ausgewogenes Bild der Transportplanung deutscher Unternehmen zu erhalten. Darüber hinaus sollte anhand der Stichproben ein Vergleich der ausgewählten Branchen untereinander durchgeführt werden.

Von den 1.000 angeschriebenen Unternehmen sandten 105 einen ausgefüllten Fragebogen zurück. Den größten Anteil bilden mit rund 80% kleine und mittlere Unternehmen mit bis zu 250 Mitarbeitern und einem durchschnittlichen Jahresumsatz von bis zu 50 Millionen Euro.

Die Analyse der empirischen Daten wurde in mehreren Stufen mit unterschiedlichen statistischen Instrumenten vorgenommen. Nach einfachen Häufigkeitsverteilungen wurden zunächst anhand von Korrelationsanalysen³ Zusammenhänge zwischen verschiedenen Einzelgrößen untersucht. Im Anschluss an die Analyse der Einzelfaktoren und ihrer bilateralen Zusammenhänge wurde ein umfassendes Modell zur Identifikation der zentralen Erfolgsfaktoren formuliert. Abschließend werden in der Studie zwei Fallstudien zum erfolgreichen Einsatz von Software in der Transportplanung vorgestellt.

1.2 Das verwendete Prozessmodell

Der Gesamtprozess der Transportplanung wurde für die Studie mittels eines fünfstufigen Modells strukturiert, welches in Abb. 1 dargestellt ist. Im Gegensatz zur Beschaffungslogistik [Kani02] und Produktionslogistik [Kern02] lässt sich für die Distributionslogistik in der Literatur kein hinreichend detailliertes und anerkanntes Referenzmodell identifizieren, was die Entwicklung eines Prozessmodells in Experteninterviews auf Basis von Standardliteratur zur Tourenplanung im Vorfeld der Studie notwendig machte. Eine anschauliche Übersicht zu den dabei berücksichtigten Aufgaben der Transportplanung liefern beispielsweise [ScDK98].

³ Zur Untersuchung der Zusammenhänge wurde zunächst der Rangkorrelationskoeffizient nach Spearman verwendet, da dieser ein parameterfreies Maß für Zusammenhänge ohne Annahmen über die Wahrscheinlichkeitsverteilung der Variablen darstellt.

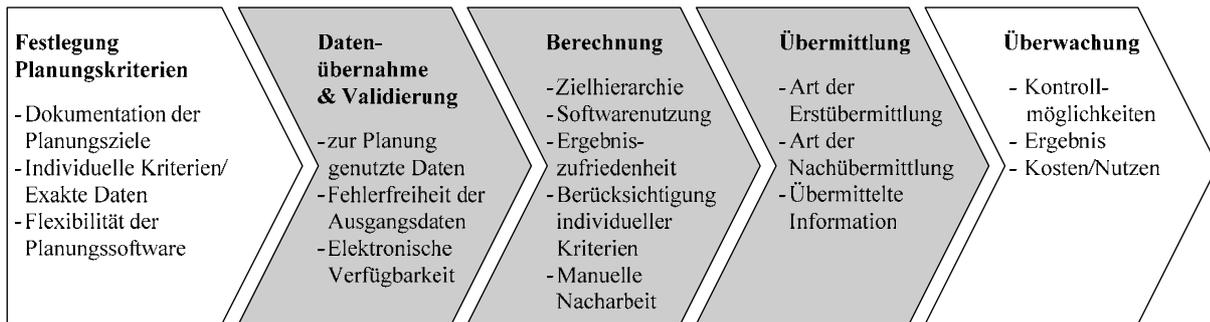


Abb. 1: Allgemeines Prozessmodell der Tourenplanung [WKSL06]

Innerhalb der Abfolge der Teilprozesse mit ihren genannten Merkmalen stellen die grau markierten Elemente den Kernprozess der Neuplanung dar. Dieser sich regelmäßig wiederholende Prozess besteht aus Datenübernahme und -validierung, Berechnung sowie Übermittlung der Pläne an die Fahrer. Dieser Kernablauf wird durch die Festlegung der Planungskriterien und die Überwachung der tatsächlich gefahrenen Touren ergänzt.

Vor einer Transportplanung müssen zunächst Planungskriterien festgelegt werden. Wichtig ist vor allem die Gewichtung verschiedener Zielkriterien, beispielsweise Pünktlichkeit versus Fahrtstrecke bzw. -dauer. Auch die Entscheidung über Kundenprioritäten muss in dieser Phase einmalig oder in Abhängigkeit von der konkreten Planungssituation getroffen werden. Neben der Erhebung, wie genau diese Zielkriterien erfasst werden, ist ein weiterer Gegenstand der Befragung, inwieweit Planungsziele dokumentiert und ob individuelle Kriterien der Planung von einzelnen Unternehmen und Softwarelösungen berücksichtigt werden.

Der Kernprozess beginnt mit der Datenübernahme der planungsrelevanten Informationen. Basierend auf diesen Planungsdaten erfolgt die Berechnung der Tourenpläne. Es wurde unter anderem erfasst, ob bestimmte Branchen unterschiedliche Rangfolgen bezüglich der Zielverfolgung aufweisen. Als weiteres Gütekriterium der Tourenplanung wird erfragt, ob eine manuelle Nachbearbeitung (oder Erstellung) der Touren notwendig ist.

Den Abschluss des Kernprozesses bildet die Übermittlung der Tourenpläne. Die Pläne können sowohl in verschiedenen Detaillierungsgraden als auch auf verschiedenen Medien übermittelt werden. Die modernste Lösung stellt ein mobiles Endgerät dar, z.B. ein PDA, auf das alle Daten der Tour inklusive einer Wegbeschreibung übertragen werden.

Die beste Planung hilft wenig, wenn die Qualität der Ausführung nicht angemessen ist. Letzter untersuchter Teilprozess der Transportplanung ist daher die Überwachung gefahrener Touren. Erhoben werden hier sowohl Informationen zur Überwachungsintensität als auch die verwendete

ten Hilfsmittel. Zum Beispiel kann der Einsatz der oben erwähnten mobilen Endgeräte zusätzlich zur Überwachung der Touren genutzt werden.

Darüber hinaus werden drei weitere Bereiche untersucht, die den Planungsprozess maßgeblich beeinflussen, ihm aber nicht direkt zuzurechnen sind. Dazu gehören das vorliegende Transport-szenario, durch das unter anderem die Anzahl der LKWs und der täglich zu beliefernden Kunden erfasst werden, allgemeine Unternehmensdaten wie Branche, Umsatz und Zahl der Mitarbeiter sowie ein Fragenblock über den Umfang von Transportleistung und -kosten.

Die Dokumentation des Planungsprozesses und die klare Benennung von Verantwortlichkeit werden in diesem Modell als grundsätzliche, prozessbegleitende Maßnahme zur Verbesserung der Transportplanung und einer damit einhergehenden Erhöhung der Zufriedenheit herausgestellt.

2 Prozessverantwortung und Dokumentation im Prozess der Tourenplanung

Abb. 2 veranschaulicht, wie die befragten Unternehmen die Aussagen „Es existiert ein Verantwortlicher für den gesamten Tourenplanungsprozess“ und „Unser Tourenplanungsprozess ist genau dokumentiert“ bewertet haben. In über 84% der befragten Unternehmen gibt es einen Verantwortlichen für den gesamten Tourenplanungsprozess (Antworten „trifft voll zu“ und „trifft eher zu“), aber nur gut die Hälfte der Unternehmen (58,8%) betrachtet ihren Planungsprozess als genau dokumentiert⁴.

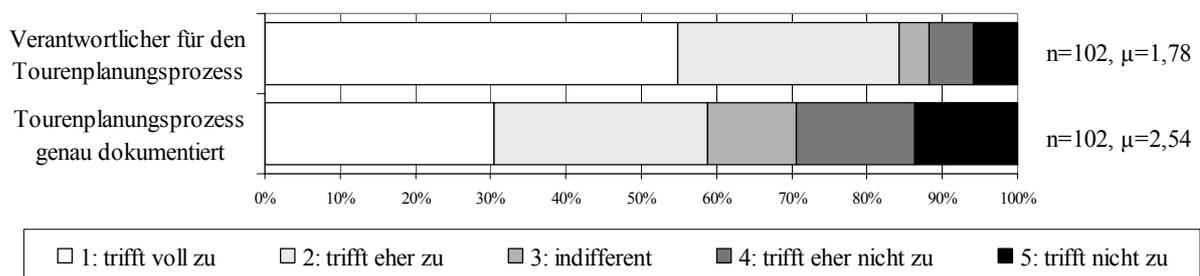


Abb. 2 Ausprägung der Dokumentation und der eindeutigen Prozessverantwortlichkeit in den befragten Unternehmen [in Anl. an WKSL06]

⁴ Die Unternehmen konnten dabei die Dokumentation des Tourenplanungsprozesses und die Existenz eines Gesamtverantwortlichen für den Tourenplanungsprozess im Fragebogen mittels fünfstufiger Likert-Skalen (trifft voll zu ... trifft nicht zu) bewerten.

Zwischen den untersuchten Branchen konnten dabei Unterschiede festgestellt werden: Speditionen und Entsorgungsunternehmen sind sowohl bei der Dokumentation als auch bei der Verantwortung im Mittel gut bis sehr gut aufgestellt. In der Nahrungsmittelindustrie und insbesondere im Baugewerbe ist die Prozessdokumentation bei in etwa gleich häufiger Benennung eines Verantwortlichen für den Planungsprozess jedoch wesentlich schwächer ausgeprägt.

2.1 Dokumentation des Tourenplanungsprozesses

Die Dokumentation von Prozessen zur Erhöhung der Logistikqualität⁵ im Allgemeinen wurde bereits in anderen Arbeiten untersucht und als integraler Bestandteil von logistischen Qualitätsmanagementansätzen identifiziert [Maue01]. In Bezug auf die Distributionslogistik konnte in einer empirischen Studie nachgewiesen werden, dass neben anderen die Prozessdokumentation einen starken Einfluss auf die Leistungsfähigkeit der Supply Chain hat [LoMc03].

Wie oben bereits dargelegt (Abb. 2), beträgt der Anteil der Unternehmen, die ihren Tourenplanungsprozess gut dokumentiert haben, trotz der in der Theorie dargestellten Notwendigkeit lediglich etwa 60%. Allerdings ist insbesondere bei Unternehmen mit hohem Transportaufkommen der Planungsprozess genau dokumentiert⁶. Für diese Unternehmen spielt die Transportplanung offensichtlich eine zentrale Rolle, worauf sie mit professionellem Prozessmanagement reagieren.

Die bereits in Abschnitt 1 erwähnte Notwendigkeit von umfassender Prozessdokumentation bei der Einführung und beim Einsatz von Softwaresystemen im Allgemeinen konnte auch für die Nutzung von Tourenplanungssoftware gezeigt werden. Unternehmen, die Tourenplanungssoftware einsetzen, haben ihren Planungsprozess genauer dokumentiert⁷.

⁵ Wobei bei [Maue01] der Begriff der Logistikqualität als das Maß für die Übereinstimmung zwischen Merkmalen und Merkmalsausprägungen des Logistikprozesses und den Kundenanforderungen an diesen definiert wird, und damit umfassender gewählt ist als der hier in Abgrenzung zur Kosteneffizienz verwendete Qualitätsbegriff.

⁶ Signifikante Korrelation: 0,237.

⁷ Hoch signifikante Korrelation: 0,425. Zur Nutzung von Tourenplanungssoftware in den vier untersuchten Branchen siehe auch Abb. 3.

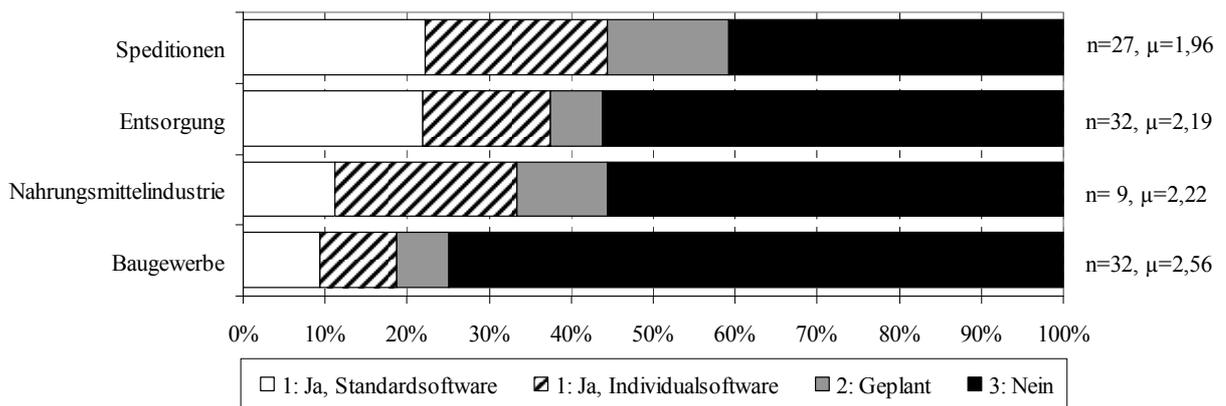


Abb. 3 Einsatz von Planungssoftware in den untersuchten Branchen [WKS06]

Gute Dokumentation wird durch einen für den Prozess verantwortlichen Mitarbeiter begünstigt⁸, so dass auch die Übernahme eindeutiger Prozessverantwortung im Modell berücksichtigt wird.

2.2 Prozessverantwortung

Im hier vorgestellten Modell zeigt sich die positive Wirkung eines Verantwortlichen für die Tourenplanung im Unternehmen auf die Gesamtzufriedenheit bezüglich der Prozessqualität. Insbesondere im Teilprozess „Überwachung der Touren“ führt eine klare Verantwortung für die Planung zu einer als hoch wahrgenommenen Zufriedenheit mit Kosteneffizienz und Qualität⁹. Die positive Wirkung eindeutig festgelegter Prozessverantwortung und der damit zusammenhängenden Kontrollmöglichkeiten und Weisungsbefugnissen zeigt sich auch daran, dass bei diesen Unternehmen größere Übereinstimmung zwischen geplanten und gefahrenen Routen erkennbar ist.¹⁰

Ein weiterer positiver Effekt klar zugeordneter Prozessverantwortung liegt in der höheren Kostentransparenz, die anhand der Differenz zwischen oberer und unterer Schranke der von den Unternehmen angegebenen Transportkosten gemessen wurde.¹¹

⁸ Dies lässt sich an der hoch signifikanten Korrelation von 0,372 zur Dokumentation des Tourenplanungsprozesses zeigen.

⁹ Hoch signifikante Korrelationen zwischen der Benennung eines Projektverantwortlichen und Kosteneffizienz bzw. Qualität bei der Überwachung der Touren: 0,287 bzw. 0,356.

¹⁰ Hoch signifikante Korrelation mit der Frage „Die Fahrer folgen den vorgegeben Routen“: 0,329.

¹¹ Die Transportkosten wurden über den Anteil der Transportkosten an den Gesamtkosten geschätzt. Die Differenz zwischen oberer und unterer Schranke weist eine signifikante Korrelation zu einer klaren Prozessverantwortung von 0,286 auf.

Diese Ergebnisse sind insbesondere vor dem Hintergrund einer Studie des Kompetenzzentrums für Geschäftsprozessmanagement [KoAc06] interessant. In dieser wurde erhoben, dass die Budgetverantwortung in Unternehmen¹², die Prozessverantwortliche haben, in über 95% der Fälle trotzdem beim Funktionsverantwortlichen verbleibt. Obwohl auch dort der Einsatz eines Prozessverantwortlichen als wichtigste Maßnahme zur Effizienzsteigerung der Prozessleistung eingeschätzt wurde, lässt die Verteilung der Budgetverantwortung vermuten, dass die Definition klarer Prozessverantwortung nur ein erster Schritt ist. Diesem muss noch die Umverteilung von Kompetenzen folgen.

3 Modellformulierung

Aus den in Abschnitt 1 und 2 zitierten Quellen¹³ lässt sich die Relevanz von Prozessverantwortung und Dokumentation für Qualität und Effizienz von Prozessen verschiedener Anwendungsdomänen entnehmen.

Aufbauend darauf soll das im Folgenden dargestellte Modell veranschaulichen, wie Prozessverantwortung und Dokumentation im Prozess der Transportplanung die Prozessqualität und -effizienz beeinflussen.

3.1 Ausgangspunkt der Modellbildung

Zur ersten Untersuchung dieser Zusammenhänge wurde ein einfaches Korrelationsmodell entwickelt, das in Abb. 4 dargestellt ist.

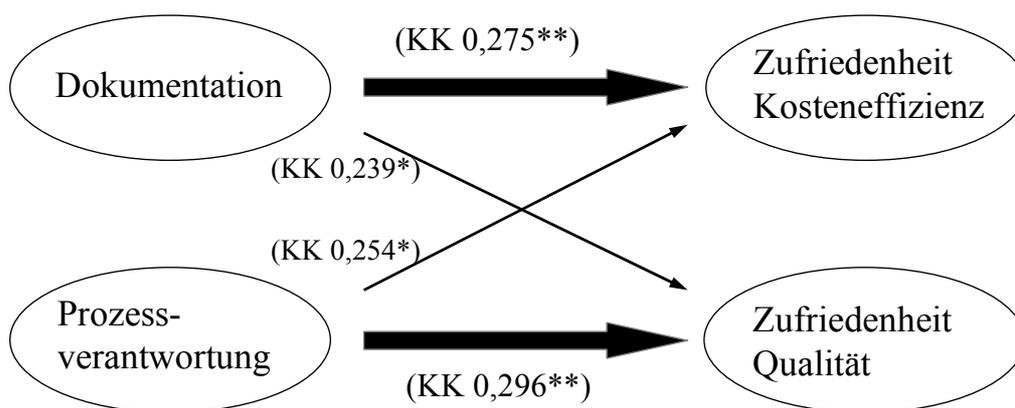


Abb. 4 Ausgangspunkt der Modellierung, Korrelationen der untersuchten Konstrukte (KK: Korrelationskoeffizient, *: signifikant auf dem 0,05-Niveau, **: signifikant auf dem 0,01-Niveau) [WKS06]

¹² In der zitierten Studie wurden Finanzdienstleister befragt.

¹³ [WeSW05; Tosc03; Maue01; LoMc03; KoAc06], darüber hinaus z.B. auch [SKGW04; Weit06].

Trotz der bereits in Abschnitt 2.1 erwähnten engen Verzahnung beider Faktoren konnten unterschiedlich starke Einflüsse dieser beiden Größen auf die Zufriedenheit bzgl. Qualität und Kosteneffizienz ausgemacht werden, die im folgenden Abschnitt anhand empirischer Daten in einem Regressionsmodell überprüft werden sollen.

Als Indikatoren für die Zufriedenheit wurden – wiederum auf Basis von Likert-Skalen – die Aussagen zur Zufriedenheit der Unternehmen mit Qualität und Kosteneffizienz innerhalb der fünf in Abschnitt 1.2 eingeführten Teilprozesse erfragt und sodann für das hier gezeigte Modell zur Gesamtzufriedenheit bezüglich Qualität bzw. Kosteneffizienz gemittelt.

3.2 Modellstruktur und Modellierungsergebnisse

Ziel des in diesem Beitrag vorgestellten Modells ist es, die in Abb. 4 vermuteten und lediglich durch Korrelationen gestützten Zusammenhänge zu belegen. Die aus Abb. 4 abgeleitete übergeordnete Hypothese „Prozessverantwortung und Dokumentation beeinflussen die Zufriedenheit mit Qualität und Kosteneffizienz in der Tourenplanung“ wurde gemäß der ersten Erkenntnisse aus dem Korrelationsmodell weiter spezifiziert:

- Die Dokumentation wirkt stärker auf die Zufriedenheit in Bezug auf Kosteneffizienz und
- die eindeutige Festlegung von Prozessverantwortung wirkt stärker auf die Zufriedenheit in Bezug auf Qualität.

Diese beiden Hypothesen wurden mittels Regressionsanalysen untersucht, die im Folgenden vorgestellt werden.

3.2.1 Regressionsanalyse zur Kosteneffizienz

Mit der ersten Regressionsanalyse soll die Hypothese untersucht werden, dass eine gute Dokumentation der Transportplanung stärker auf die Kosteneffizienz wirkt als die Benennung eines Prozessverantwortlichen.

Da der Einfluss der Variablen „Es existiert ein Verantwortlicher für den gesamten Tourenplanungsprozess.“ (*Verantwortung*) und „Unser Tourenplanungsprozess ist genau dokumentiert.“ (*Dokumentation*) auf die Variable „Bewertung der Kosteneffizienz“ (*Kosteneffizienz*) überprüft werden soll, werden *Verantwortung* und *Dokumentation* als unabhängige Variablen (Regressoren) und *Kosteneffizienz* als abhängige Variable (Regressand) definiert. Damit wird versucht,

den Regressanden $y_{\text{Kosteneffizienz}}$ als Linearkombination der Regressoren $x_{\text{Dokumentation}}$ und $x_{\text{Verantwortung}}$ auszudrücken: $y_{\text{Kosteneffizienz}} = \alpha + \beta_1 x_{\text{Dokumentation}} + \beta_2 x_{\text{Verantwortung}}$.

In einem ersten Schritt wird überprüft, ob die ausgewählten Regressoren überhaupt einen Beitrag zur Erklärung des Regressanden leisten. Als Maß für die Anpassungsgüte des durch die Regressionsgleichung formulierten Zusammenhangs zwischen den unabhängigen Variablen und der abhängigen Variable an die tatsächlichen Gegebenheiten wird das Bestimmtheitsmaß R^2 verwendet [ChPr77].

Es kennzeichnet den Varianzanteil des abhängigen Merkmals, der durch das lineare Modell erklärt wird, und lässt sich mit Hilfe folgender Definitionen berechnen [FKPT03]:

$$SQE = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \quad \text{erklärte Streuung}$$

$$SQR = \sum_{i=1}^N (y_i - \bar{y})^2 \quad \text{Reststreuung}$$

$$SQT = SQE + SQR \quad \text{Gesamtstreuung}$$

Dabei ist N der Umfang der Stichprobe, y_i der tatsächliche Wert, \hat{y}_i der geschätzte Wert und \bar{y} der Mittelwert von y_i .

Für das Bestimmtheitsmaß R^2 wurde im untersuchten Modell $R^2 = SQE/SQT = 3,728/35,334 = 0,1055$ ermittelt.

Das Bestimmtheitsmaß kann per Definition nur Werte zwischen 0 und 1 annehmen, wobei im Extremfall $R^2=1$ die komplette Streuung durch die Regressionsfunktion erklärt würde. Da im betrachteten Modell der Wert von R^2 bei nur 0,1055 liegt, würde dies bedeuten, dass dem vermeintlich gefundenen Zusammenhang zwischen den Regressoren und dem Regressanden keine Bedeutung geschenkt werden dürfte. [FKPT03]

Zur Verifikation dieser Bewertung wird ein „Goodness-of-fit-Test“ durchgeführt, der die Regressionsfunktion unter Berücksichtigung des Stichprobenumfangs sowie der Zahl der Regressoren nochmals überprüft. Dabei wird die Hypothese H_0 „Es besteht kein Zusammenhang zwischen den Regressoren und dem Regressanden“ (d.h. $\beta_i = 0 \forall i$) getestet. Dazu wird ein empirischer F-Wert F_{emp} ermittelt, der mit einem theoretischen F-Wert F_{tab} verglichen wird, der bei einem zufällig vorliegenden Zusammenhang angenommen würde. Ist nun $F_{\text{emp}} > F_{\text{tab}}$, so wird die Hypothese H_0 abgelehnt und die über R^2 ermittelte Abhängigkeit des Regressanden von den Regressoren bestätigt. [ChPr77]

Das Signifikanzniveau, auf dem die Hypothese H_0 verworfen werden kann, beträgt im betrachteten Modell 0,7%. Somit ist der Zusammenhang, den R^2 in der Stichprobe ausdrückt, nicht zufällig. Vielmehr ist der Zusammenhang der Variablen auf einem sehr hohen Signifikanzniveau abgesichert.

Nach der Prüfung der Regressionsfunktion durch den F-Test werden im nächsten Schritt die einzelnen Regressoren auf ihre Signifikanz überprüft. Für jeden Regressor wird die Hypothese H_0 „Der Regressor hat keinen Einfluss auf die abhängige Variable“ (d.h. $\beta_i = 0$ für $i=1, 2$) getestet. Im untersuchten Modell wird die Hypothese für beide Regressoren verworfen, für die Variable *Dokumentation* mit einer sehr guten Signifikanz von 0,027, für die Variable *Verantwortung* allerdings nur mit einer Signifikanz von 0,125, was in vorliegendem Fall aber noch als vertretbar angesehen werden kann.¹⁴ [DrSm66]

Zuletzt wird in einer Kollinearitätsdiagnose untersucht, ob zwischen den unabhängigen Variablen eine stark ausgeprägte lineare Abhängigkeit besteht. Für das untersuchte Modell konnte keine Kollinearität zwischen den Regressoren ermittelt werden. Es konnten auch die Forderungen [DrSm66] bestätigt werden, dass bei der Verteilung der Residuen (der Fehler der Schätzung) keine Heteroskedastizität vorliegen darf¹⁵ und dass die Residuen der Normalverteilung folgen.

Daher kann die Gleichung

$$y_{\text{Kosteneffizienz}} = 1,991 + 0,104 x_{\text{Dokumentation}} + 0,087 x_{\text{Verantwortung}}$$

mit vertretbarer Signifikanz (0,027 bzw. 0,125) zur Stützung der Annahme herangezogen werden, dass sich eine gute Prozessdokumentation stärker auf die Kosteneffizienz auswirkt als die Benennung eines Prozessverantwortlichen.

3.2.2 Regressionsanalyse zur Qualität

Mit der in Abschnitt 3.2.1 eingeführten Methodik wird eine zweite Regressionsanalyse durchgeführt, die zeigen soll, dass die Benennung eines Prozessverantwortlichen größeren Einfluss auf die wahrgenommene Qualität der Transportplanung hat als die Prozessdokumentation.

Wie bereits im Modell zur Kosteneffizienz werden die Variablen „Es existiert ein Verantwortlicher für den gesamten Tourenplanungsprozess.“ (*Verantwortung*) und „Unser Tourenplanungsprozess ist genau dokumentiert.“ (*Dokumentation*) als unabhängige Variablen (Regressoren) gewählt, als abhängige Variable (Regressand) wird jetzt *Qualität* gewählt. Somit soll eine lineare

¹⁴ Eine umfangreichere Stichprobe würde die Signifikanz wahrscheinlich erhöhen.

¹⁵ Das heißt, die Varianz der Residuen für den gesamten Wertebereich muss homogen sein.

re Funktion zwischen dem Regressanden $y_{\text{Qualität}}$ und den beiden Regressoren $x_{\text{Dokumentation}}$ und $x_{\text{Verantwortung}}$ bestimmt werden: $y_{\text{Qualität}} = \alpha + \beta_1 x_{\text{Dokumentation}} + \beta_2 x_{\text{Verantwortung}}$.

Das Bestimmtheitsmaß R^2 ist – wie schon in der Regressionsanalyse der Kosteneffizienz – auch in diesem Modell mit 0,102 recht klein, jedoch kann die mittels eines F-Tests überprüfte Hypothese „Es besteht kein Zusammenhang zwischen den Regressoren und dem Regressanden“ mit der sehr hohen Signifikanz von 0,01 verworfen werden.

Die sich anschließende Überprüfung der Signifikanz der einzelnen Regressoren liefert auch hier zufrieden stellende Ergebnisse. Mit einer Signifikanz von 0,077 (*Dokumentation*) bzw. 0,071 (*Verantwortung*) haben beide Variablen einen bedeutenden Einfluss auf die Wahrnehmung der Qualität des Tourenplanungsprozesses (*Qualität*).

Die abschließend durchgeführten Untersuchungen zu Kollinearität, Homoskedastizität und der Normalverteilung der Residuen rechtfertigen es, auf Basis der Gleichung

$$y_{\text{Qualität}} = 1,89 + 0,099 x_{\text{Dokumentation}} + 0,12 x_{\text{Verantwortung}}$$

mit vertretbarer Signifikanz (0,077 bzw. 0,071) zu behaupten, dass eine eindeutige Prozessverantwortung stärkeren Einfluss auf die wahrgenommene Qualität hat als eine gute Prozessdokumentation.

4 Diskussion der Ergebnisse und weitere Gestaltungsempfehlungen

Ausgangspunkt der Untersuchung war die Analyse der Wirkung von Prozessverantwortung und -dokumentation auf die Effizienz und Qualität der Transportplanung im Vorfeld einer Optimierung der Tourenplanung durch Planungssoftware.

Die Ergebnisse der beiden in diesem Beitrag vorgestellten Regressionsanalysen bestätigen zum einen, dass sowohl die Prozessdokumentation als auch die eindeutige Festlegung der Prozessverantwortung positiven Einfluss auf Prozesseffizienz und -qualität haben. Zum anderen konnte gezeigt werden, dass die Dokumentation stärker auf die Zufriedenheit im Hinblick auf Kosteneffizienz und die Prozessverantwortung stärker auf die Zufriedenheit im Hinblick auf Qualität wirkt.

Im nächsten Abschnitt soll folglich zunächst die Verbesserung der organisatorischen Rahmenbedingungen für die Tourenplanung als erster Schritt bei der Optimierung der gesamten Tourenplanung vorgeschlagen werden. Darauf aufbauend wird im Ausblick das Optimierungspotenzial durch den Einsatz von Tourenplanungssoftware diskutiert.

4.1 Optimierung der organisatorischen Rahmenbedingungen

Prozessverantwortung beinhaltet insbesondere die Verantwortung für die operative Durchführung des Planungsprozesses. Der direkte Zusammenhang zwischen Prozessqualität und Verantwortung konnte in Abschnitt 3.2.2 gezeigt werden. Weitere mögliche Faktoren, durch die ein Prozessverantwortlicher die Zufriedenheit mit der Transportplanung steigern kann, sind insbesondere technologische Effizienzsteigerungen und bessere Kontrollmechanismen der Planungsergebnisse.

Unternehmen sollten bei der Benennung eines Verantwortlichen für den Transportprozess darauf achten, dass der entsprechende Mitarbeiter sowohl über effiziente Kontrollinstrumente (z.B. GPS-Tracking) als auch über den organisatorischen Rahmen zur effizienten Steuerung des Prozesses verfügt. In diesem Zusammenhang sei nochmals auf die bereits in Abschnitt 2.2 zitierte Studie verwiesen [KoAc06], in der gezeigt wurde, dass die Benennung von Prozessverantwortlichen allein nicht ausreicht. Erst wenn auch die Verantwortung stärker von den Funktions- auf die Prozessverantwortlichen übertragen wird, kommt die Prozessorientierung wirklich zum Tragen.

Eine gute Dokumentation des Transportplanungsprozesses zahlt sich für Unternehmen aus. Neben der bereits dargestellten Wirkung auf die Kosteneffizienz führt sie im Detail zu einer vollständigeren Erfassung der Ausgangsdaten.¹⁶ Diese ist bei gut dokumentierten Planungsprozessen wesentlich stärker ausgeprägt, da viel besser aufgezeichnet wurde, welche Planungskriterien erforderlich sind.¹⁷

Obwohl die eindeutige Vergabe der Prozessverantwortung und die Dokumentation des Planungsprozesses eng miteinander verknüpft sind und im Idealfall verzahnt angegangen werden sollten, können Unternehmen die Intensität und Reihenfolge der beiden Maßnahmen danach bewerten, ob sie größere Defizite in ihrer Kosteneffizienz oder ihrer Prozessqualität sehen.

Im vorgestellten Modell wurden zwei Merkmale im Kontext der Zufriedenheit von Unternehmen mit Effizienz und Qualität der Tourenplanung untersucht. Im Falle des Einsatzes von Software sind dies jedoch nicht die einzigen Faktoren, welche auf den Erfolg der Transportplanung wirken. Ergänzend zu dieser Analyse wird daher im Folgenden kurz ein weiteres Modell vorgestellt, welches vier zentrale Einflussfaktoren für den erfolgreichen Einsatz von Software zur Tourenplanung untersucht.

¹⁶ Hoch signifikante Korrelation: 0,528.

¹⁷ Hoch signifikante Korrelation: 0,293.

4.2 Prozessoptimierung durch Tourenplanungssoftware

In der bereits erwähnten Studie „Transportplanung der Zukunft“ wurden in einem logistischen Regressionsmodell vier Maßnahmenpakete zur Optimierung der Transportplanung ermittelt [WKSL06]:

- Prozessverantwortung und -dokumentation,
- Berücksichtigung individueller Planungskriterien,
- zentrale Planung und kooperativer Planungsansatz und
- Prozessoptimierung durch Tourenplanungssoftware.

Insofern ist das hier gezeigte Modell nicht losgelöst von einer Softwareimplementierung zu sehen, sondern im Gegenteil als Plädoyer für eine gründliche Beschäftigung mit der Dokumentation von Prozessen und der Vergabe von eindeutiger Prozessverantwortung vor oder während der Veränderung des Planungsprozesses und der Einführung von Tourenplanungssoftware. Neben der Qualität der verwendeten Planungsdaten ist besonders bei der Softwareeinführung die detaillierte Dokumentation des Tourenplanungsprozesses von hoher Relevanz. Sie sollte von den im Unternehmen verantwortlichen Mitarbeitern gemeinsam mit Softwareexperten, die die Auswirkungen spezifischer Restriktionen auf den Optimierungsprozess analysieren, geplant und durchgeführt werden.

Weiterhin setzt der erfolgreiche Einsatz von Software eine genaue Anpassung an die individuellen Voraussetzungen des jeweiligen Unternehmens und die Integration von effizienten Kontrollinstrumenten voraus. Und auch wenn eine solche Berücksichtigung individueller Planungsziele und -kriterien geplant ist, setzen sich Unternehmen verstärkt mit der Dokumentation ihres Planungsprozesses auseinander, um den Bedarf an zu erfassenden Kriterien bestmöglich zu identifizieren.¹⁸

Zusammenfassend ist zu konstatieren, dass der erste Schritt zur Verbesserung der Transportplanung in den meisten Unternehmen eine Aktualisierung oder Neuerstellung der Dokumentation des Transportplanungsprozesses sein sollte. Auf dieser aufbauend können organisatorische Verbesserungspotenziale identifiziert werden und – falls die Effizienz einer software-basierten Planung den Verlust an Flexibilität aufwiegt – kann der Planungsprozess auf die Einführung einer Softwarelösung vorbereitet werden.

¹⁸ Signifikante Korrelation: 0,214.

Literaturverzeichnis

- [Bund05] Bundesamt für Güterverkehr (Hrsg.): Jahresbericht 2004. Köln 2005.
- [Bund05b] Bundesministerium für Verkehr, Bau- und Wohnungswesen (Hrsg.): Weitere Informationen zur Lkw-Maut. <http://www.bmvbs.de/artikel,-22466/Weitere-Informationen-zur-Lkw-.htm#3>, Abruf am 2006-07-20.
- [Bund06] Bundesamt für Güterverkehr (Hrsg.): Jahresbericht 2005. Köln 2006.
- [ChPr77] Chatterjee, Samprit; Price, Bertram: Regression Analysis by Example. Wiley, New York 1977.
- [DrSm66] Draper, Norman R.; Smith Harry: Applied Regression Analysis. Wiley, New York 1966.
- [FKPT03] Fahrmeir, Ludwig; Künstler, Rita; Pigeot, Iris; Tutz, Gerhard: Statistik - Der Weg zur Datenanalyse. 4. Aufl., Springer, Berlin Heidelberg 2003.
- [Kani02] Kanitz, Frieder: Kennzahlenbasierte Fehleridentifizierung in der Beschaffungslogistik, Dissertation. Universität Hannover 2002.
- [Kern02] Kerner, Axel: Modellbasierte Beurteilung der Logistikleistung von Prozessketten, Dissertation. Universität Hannover 2002.
- [KoAc06] Kompetenzzentrum für Geschäftsprozessmanagement GbR & Acrys Consult GmbH & Co. KG (Hrsg.): Ergebnisse der Studienumfrage Status Quo Geschäftsprozessmanagement 2005. http://www.acrys.com/en/PDF/GPM_Umfrageergebnisse_2005.pdf, Abruf am 2006-07-10.
- [Like32] Likert, Rensis: A Technique for the Measurement of Attitudes. Archives of Psychology, New York 1932.
- [LoMc03] Lockamy III, Archie; McCormack, Kevin: The Effect of Process Maturity on Supply Chain Performance: An Empirical Study, in: Metters, Rich (Hrsg.): Proceedings of the POMS-2003 Production and Operations Management Society, 4-7 April 2003, Savannah, Georgia 2003.

- [Maue01] Mauermann, Helmut: Leitfaden zur Erhöhung der Logistikqualität durch Analyse und Neugestaltung von Versorgungsketten, Dissertation. Universität-GH Paderborn 2001.
- [ScDK98] Scholl, Armin; Domschke, Wolfgang; Klein, Robert: Logistik: Aufgaben der Tourenplanung, in: WISU 27 (1998) 1, S. 62-67.
- [SKGW04] Skiera, Bernd; König, Wolfgang; Gensler, Sonja; Weitzel, Tim; Beimborn, Daniel; Blumenberg, Stefan; Franke, Jochen; Pfaff, Donovan: Financial Chain Management- Prozessanalyse, Effizienzpotenziale und Outsourcing. BoD 2004.
- [Stat04] Statistisches Bundesamt (Hrsg.): Statistischer Wochenbericht, 51. KW. Wiesbaden 2004.
- [Stra05] Straube, Frank (Hrsg.): Trends und Strategien in der Logistik - Ein Blick auf die Agenda des Logistik-Managements 2010. Deutscher Verkehrs-Verlag 2005.
- [Tosc03] Toschläger, Markus: Situativer Methoden- und Werkzeugeinsatz für das Management von IT-Projekten in kleinen und mittleren Unternehmen: Konzeption einer Methodik und Entwurf eines webbasierten Entscheidungsunterstützungssystems, Dissertation. Universität Paderborn 2003.
- [Weit06] Weitzel, Tim: Process governance and optimization for IT Reliant Business Processes: an empirical analysis of financial processes in Germany's Fortune 1000 non-banks, in: In: Sprague, Ralph H. (Hrsg.): Proceedings of the 39th Hawaii International Conference on System Sciences 4-7 January 2006, Kauai, Hawaii. IEEE Computer Society Press, Los Alamitos, CA 2006.
- [WeSW05] Wendt, Oliver; Stockheim, Tim; Weiß, Kilian; Intelligente Tourenplanung mit DynaRoute, in: WIRTSCHAFTSINFORMATIK 47 (2005) 2, S. 135-140.
- [WKSL06] Wendt, Oliver; König, Wolfgang; Stockheim, Tim; Lanninger, Volker; Weiß, Kilian: Transportplanung der Zukunft - Prozess- und Kostenanalyse, Optimierungspotenziale und Outsourcing. BoD 2006.

Einführung in den Track

Mobile Business and Communications

Prof. Dr. J. Felix Hampe

Universität Koblenz-Landau

Prof. Dr. Kai Rannenber

Universität Frankfurt a.M.

Prof. Dr. Klaus Turowski

Universität Augsburg

Nach wie vor erschließen sich durch die hochdynamischen Entwicklungen im Bereich mobiler und allgegenwärtiger Technologien neue Forschungs- und Anwendungsfelder für mobile Geschäftsszenarien und mobile Informations- und Kommunikationssysteme. Die breite Verfügbarkeit von Standards, von Infrastrukturen für drahtlose Dienste und nicht zuletzt die hohe Verbreitung von mobilen Endgeräten sorgen zudem für ein großes Marktpotenzial.

Ziel dieses Tracks ist die Auseinandersetzung mit allen Aspekten des Mobile Business von Enabling Technologies und Infrastrukturen bis hin zu Geschäftsmodellen, Anwendungen, Wirtschaftlichkeitsbetrachtungen und methodischen Ansätzen.

Programmkomitee:

Prof. Dr. Michael Amberg, Universität Erlangen-Nürnberg

Prof. Dr. Martin Breunig, Universität Osnabrück

Prof. Dr. Rony G. Flatscher, Wirtschaftsuniversität Wien

Prof. Dr. Birgitta König-Ries, Universität Jena

Prof. Dr. Franz Lehner, Universität Passau

Prof. Dr. Florian Matthes, TU München

Dr. Key Pousttchi, Universität Augsburg

Prof. Dr. Claus Rautenstrauch, Universität Magdeburg

Dr. Ingo Schneider, T-Mobile International

Prof. Dr. Detlef Schoder, Universität zu Köln

Privatsphärenfreundliche topozentrische Dienste unter Berücksichtigung rechtlicher, technischer und wirtschaftlicher Restriktionen

Mike Radmacher, Jan Zibuschka, Tobias Scherner, Lothar Fritsch, Kai Rannenberg

Lehrstuhl für Mobile Commerce und Mehrseitige Sicherheit

University of Frankfurt

60054 Frankfurt am Main

{mike.radmacher, jan.zibuschka, tobias.scherner, lothar.fritsch, kai.rannenberg}

@m-lehrstuhl.de

Abstract

Die Entwicklung neuer Produkte im dynamischen Markt mobiler Datendienste stellt Anbieter und Netzbetreiber vor besonderen Herausforderungen. Kurzlebige Produkte, unbekannte Akzeptanz potentieller Nutzer und der Zwang neue Dienste auf vorhandene Infrastrukturen aufzusetzen, führt zu Schwierigkeiten bei der Umsetzung. Gerade die Datenschutz-Regulierung der Telekommunikation und die Durchdringung des Alltags mit allgegenwärtigen Kommunikations- und Ortungsmitteln stellt die Produktentwickler vor die Herausforderung geschäftliche Interessen, Datenschutzvorgaben und Nutzerpräferenzen zu modellieren, auszugleichen und in den Basisinfrastrukturen neuer mobilen Datenprodukte zu implementieren. Der vorliegende Artikel präsentiert einen Entwurf und die Implementierung eines Prototyps, der im Rahmen eines Forschungsprojektes mit Industriebeteiligung entwickelt wurde.

1 Einleitung

Die große Verbreitung, die Mobiltelefone auf Basis der GSM- oder UMTS-Standards heute erreicht haben, macht die Verwendung dieser Infrastruktur für die Ortung des Benutzers im Rahmen der Erbringung topozentrischer Dienste außerordentlich attraktiv, da so eine große Benutzerbasis angesprochen werden kann [Bund05]. Endgeräte mit integrierter Ortungsfunktion - wie GPS oder Galileo - sind nach wie vor selten [BuHE00]. Dabei kann der Netzbetreiber sel-

ber als Anbieter topozentrischer Dienste auftreten oder nur die Ortsdaten für die eigentlichen Dienstanbieter liefern.

1.1 Problemstellung

Die Verwendung sensibler Kundendaten wie die Position des Benutzers ist mit rechtlichen Anforderungen verbunden, die den Schutz der Privatsphäre des Kunden betreffen. Obwohl Netzbetreiber sehr gute Voraussetzungen haben, die Ortungstechnik für die Erbringung topozentrischer Dienste zu stellen, verbleiben in diesem Geschäftsfeld hinsichtlich der Umsetzung der rechtlichen Vorgaben noch offene Fragen, etwa nach der Verantwortlichkeit bei einem möglichen Missbrauch der benutzerbezogenen Daten. Um nicht für Verfehlungen eines Dienstanbieters verantwortlich gemacht zu werden, ergibt sich die Motivation, die tatsächliche Erbringung der Dienste auszulagern und eine Schnittstelle zu schaffen, die eine saubere Trennung von Netzbetreiber und Dienstanbieter erleichtert. Um die Weitergabe der Position eines Benutzers unabhängig vom konkret erbrachten topozentrischen Dienst zu erlauben, muss die Schnittstelle weiterhin zu einem erheblichen Umfang vom Nutzer konfigurierbar sein, um mit grundsätzlichen Datenschutzerfordernungen in Einklang zu stehen. Daher ist eine Lösung erforderlich, die neben den wirtschaftlichen und rechtlichen Restriktionen die Privatsphäre des Nutzers respektiert und es ihm ermöglicht, die Herausgabe und Verwendung seiner persönlichen Daten zu kontrollieren. Ein System zum Identitätsmanagement, wie es ein aktuelles Forschungsprojekt implementiert, wird unerlässlich.

1.2 Szenario

Zur genaueren Analyse des Problems wurde ein spezielles Szenario gewählt. Dabei findet ein typischer topozentrischer Suchdienst Betrachtung, der prototypisch implementiert ist, um unter möglichst realitätsnahen Bedingungen die Eigenschaften der hier vorgestellten Lösung zu evaluieren:

Ein Handelsreisender (John) ist in einer neuen Stadt angekommen. Er stellt fest, dass er eine neue Dosis eines wichtigen Medikaments benötigt. Daher entscheidet er sich einen topozentrischen Dienst zu nutzen, der es ihm erlaubt, die nächste Apotheke zu lokalisieren. Sein Endgerät nimmt Verbindung mit dem Dienstanbieter auf, während der Netzbetreiber die eigentliche Ortung vornimmt. Die festgestellte Position wird an den Dienstanbieter übermittelt, der sie mit

seiner Datenbank vergleicht. Die Resultate - in diesem Fall die k nächsten Apotheken – werden an Johns Endgerät zurückgeliefert,.

1.3 State-of-The-Art

Es existiert heute eine große Anzahl von Techniken, mit deren Hilfe die Privatsphäre des Anwenders geschützt werden kann, zusammenfassend Privacy Enhancing Technologies (PETs) genannt [BIBO03]. Für die privatsphären-respektierende Behandlung von Ortsdaten finden sich in der Literatur verschiedene Ansätze bzw. Architekturen. Es folgt eine kurze Betrachtung ausgewählter Ansätze und Architekturen, um eine Gegenüberstellung der mit in diesem Papier dargestellten Architektur einzuleiten und die Vorteile hervorzuheben.

Die Alipes-Plattform [SNP02] bietet dem Benutzer die Möglichkeit den Zugriff auf seine Ortsdaten mittels von ihm konfigurierter Richtlinien zu kontrollieren. Weiterhin können Ortsinformationen aus unterschiedlichen Quellen aggregiert werden. Alipes bietet keinen weitergehenden Schutz der Identität des Nutzers und keine Pseudonymisierung.

Ein weiteres Beispiel stellt der von T-Systems entwickelte T-Identity Protector dar, der den Identitätsschutz unterstützen soll [Wage06]. Diese Lösung konzentriert sich auf die Pseudonymisierung von personenbezogenen Daten, bevor diese an eine weitere Instanz zur Verarbeitung weitergegeben werden. Die De-Pseudonymisierung von Nutzern ist für strittige Fälle vorgesehen. Der T-Identity Protector setzt ausschließlich auf Pseudonymisierung und umgeht die notwendige Einholung der Zustimmung des Nutzers vor der Weitergabe seiner Daten.

Durch das in [BöLR04] dargestellte Konzept wird einem Nutzer die Möglichkeit gegeben, für jeden Dienst explizit festzulegen, ob dieser eine Lokalisierung vornehmen darf oder nicht. Das Konzept liefert allerdings keinen Architekturvorschlag oder geht auf technische Details ein, wie beispielsweise die Verwendung von Pseudonymen.

In [RePr04] stellen die Autoren eine Architektur vor, die zur Reduktion der Ortsauflösung verwendet wird, um Dienste für die weniger genaue Informationen ausreichend sind, nur mit dem Notwendigen zu versorgen. Weder auf eine Zugriffskontrolle der Ortsinformationen, die Verwendung von Pseudonymen, die Einbettung in anderen Systemen noch die Verwendung kryptografischer Verfahren wird eingegangen.

Die in [JoBe04] vorgestellte Architektur adressiert topozentrische Dienste in Mobilfunkumfeld. Eine einfache Zugriffskontrolle und die Verwendung von Pseudonymen zur Verschleierung der

MSISDN werden diskutiert. Darüber hinaus findet allerdings keine Betrachtung der Informationsflüsse zwischen den an der Kommunikation beteiligten Parteien und dessen Schutz statt.

In [Oino02] werden Prozeduren zur Sicherstellung des Datenschutzes bei der Erbringung topozentrischer Dienste vorgeschlagen. Der Beitrag fokussiert sich auf die Einhaltung der gesetzlichen Vorgaben im Rahmen des Mobilfunks. Darin wird eine zentrale Entscheidungsfunktion auf Regelbasis vorgeschlagen, welche zwischen Ortsquelle und die nutzende Anwendung geschaltet ist. Genauer wird nicht spezifiziert, insbesondere keine Sicherheitsarchitektur.

Im Gegensatz dazu wird in [MyAD03] eine regelbasierte Sicherung der Ortsdaten vorgeschlagen. Beim Versuch des Zugriffs auf die Daten durch einen Diensterbringer evaluieren Entitäten namens „Validators“ die Regeln und fällen die Zugriffsentscheidung. Über die Strukturierung der Architektur liefert das Papier keine Angaben, insbesondere über die Installation der Validatoren als 3. Partei oder als Teil einer der anderen Entitäten.

Weiterhin gibt es Ansätze, die Modellierung von Systemen, die Datenschutz-Anforderungen erfüllen müssen, zu formalisieren, um eine Berücksichtigung der Interessen der verschiedenen Parteien sicherzustellen, siehe [FrSR06]. Wissenschaftler und Industriepartner untersuchen die Architektur und Anwendung dieser Technologien in unterschiedlichen Forschungsprojekten im Rahmen des FP6/IST Programms der Europäischen Union. Neben den technischen und wirtschaftlichen Aspekten wird der Fall untersucht, dass Ortsdaten eines Benutzers von einem Netzbetreiber zur Verfügung gestellt und für die Verwendung in topozentrischen Diensten weitergeleitet werden [KFKK05] [Dumo05].

Darüber hinaus wurde zur Standardisierung von PETs im April 2005 von ISO/IEC JTC1/SC27 eine Study Period mit dem Thema Privacy eingeleitet. Ein entsprechender Identitätsmanagement-Standard wird ebenfalls seit April 2005 entwickelt.

2 Anforderungen

Die Anforderungen an das Identitätsmanagement in topozentrischen Diensten in Mobilfunknetzen sind vielseitig. Im Rahmen dieses Beitrags werden vor allem die geschäftlichen Interessen des Mobilfunk-Netzbetreibers im Verhältnis zur gesetzlichen Regulierung des Datenschutzes und den Privatsphäreninteressen der Nutzer betrachtet. Die Anbieter topozentrischer Dienste spielen in ihrer starken Abhängigkeit von den Netzanbietern hier eine untergeordnete Rolle; im Basisszenario wird uneingeschränkt davon ausgegangen, dass die Datenkommunikation zwi-

schen Diensteanbieter und Nutzern ausschließlich über die Netzanbieter stattfindet. Des Weiteren ist von einer quasi Monopolstellung der Netzanbieter für die Ortung der Nutzer-Terminals auszugehen¹. Im resultierenden Szenario erbringt der Diensteanbieter, vermittelt vom Netzanbieter, topozentrische Datendienstleistungen auf Basis eigener Datenbestände. Diese Leistungen werden ganz oder teilweise mit dem Netzanbieter abgerechnet. Den Zugang zu Nutzern, Ortung und das Identitätsmanagement berechnet der Netzbetreiber. Ausprägungen und Diskussion dieser Konstellation wurden in [LFPR04] and [KFKK05] beschrieben.

2.1 Geschäftsmodell

Die vertriebliche Struktur von Mobilfunk-Zusatzdiensten wie Klingelton-Downloads, Handy-Logos und weiterer Dienste baut auf Wiederverkäufer-Strukturen auf. Netzbetreiber bieten die Zugangsschnittstellen zu Infrastruktur, Identitätsmanagement und Abrechnungsdienstleistungen an. Die Anforderungen des Industriepartners wurden zwischen verschiedenen Abteilungen abgestimmt und als Anforderungsdokument in das Prototyp-Projekt eingebracht [PRIM04a]. Als Grundszenario sollte ein System geschaffen werden, an welches analog zum Rufnummern- oder Klingeltongeschäft kaskadierende Wiederverkäufer für topozentrische Dienste ermöglicht, um die Vertriebsleistungen zu externalisieren und die Dienste nicht im Hause des Mobilfunkunternehmens installieren zu müssen. Die Aufstellung dieser Geschäftsvorgaben ist in Abbildung 1 dargestellt.

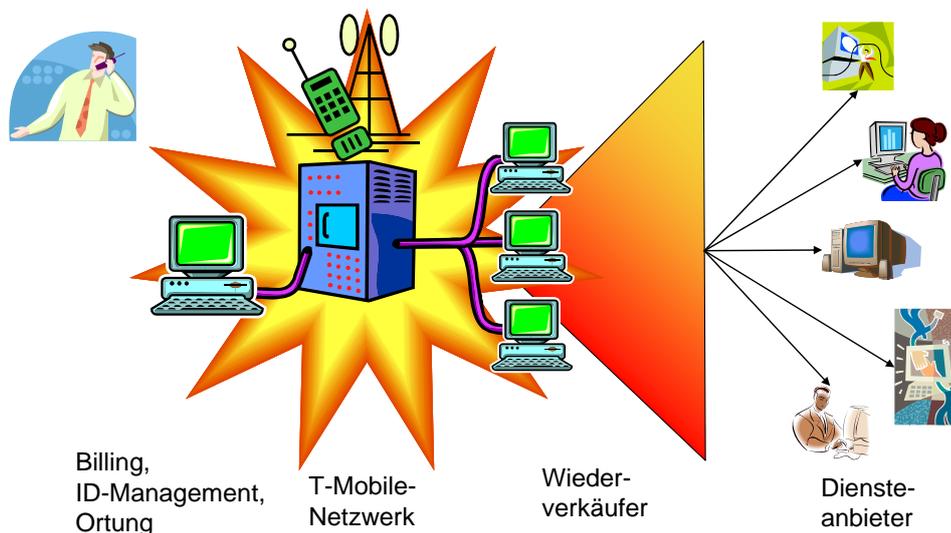


Abb. 1.: Struktur des Geschäftsmodells für topozentrische Dienste mit Wiederverkäufern.

¹ Es existieren zwar Terminals z.B. mit eingebauten Empfängern zur Satellitenpositionierung, diese sind aktuell jedoch nicht weit verbreitet. Diese Geräte stehen spezialisierten Nutzergruppen derzeit nur durch teure Neuan-schaffung zur Verfügung.

2.2 Intermediäre bei Informationsdiensten

Die pyramidenförmige Struktur der Dienstanbindung der Mobilfunkbetreiber legt einen Blick auf die Intermediärstheorie nahe. Neben der effizienten Organisation des Vertriebs über Dienstanbieter können Prozesse, Informationsprodukte und andere Leistungen an der zentralen Schnittstelle angeboten und gebündelt werden. Die Vorzüge von Intermediären bzgl. der Suchzeiten und Preisstrukturen von gehandelten Produkten sind wissenschaftlich erforscht. Einen guten Überblick für Intermediäre von Informationsprodukten findet man bei Rose in [Rose99] sowie in [ScLi02]. Zur Wertsteigerung durch Intermediation eignen sich im Kontext topozentrischer Netz-Mehrwertdienste besonders:

- Identitätsmanagement und Datenschutz nach Telekommunikationsrecht (siehe [KFKK05] oder in einem rudimentären Ansatz in [BöLR02])
- Bündelung von Abrechnungs- und Risikomanagementdienstleistungen
- Verkauf von Zusatzdiensten wie Geoinformationen, Profilierung (wie zum Beispiel in [Figg04])

Privatsphärenmanagement im Spannungsfeld von Regulierung und Kundenzufriedenheit ist ein kostenintensives Feld, wie in [Pone04] aufgezeigt wurde. Daher ist bei der Gestaltung der Wiederverkäufer-Szenarien die Berücksichtigung von Identitäts- und Privatsphärenmanagement als Posten in der Wertschöpfungskette besonders attraktiv.

2.3 Datenschutz

Neben den wirtschaftlich orientierten Anforderungen aus den Abschnitten 2.1 und 2.2 ergeben sich weitere Anforderungen aus der Regulierung. Telekommunikation in Europa wird durch Richtlinien reguliert, welche sich nach der Umsetzung in nationalen Gesetzen einzelner Ländern wieder finden. Relevant für die Implementierung ist die Richtlinie 2002/58/EC [Euro02], in welcher in Artikel 9 die unterschiedliche rechtliche Betrachtung der Verwendung personenbezogener Ortsdaten einerseits zum Zweck der Anruf-Durchstellung und andererseits zur weiteren Nutzung vorgeschrieben wird. Zur anderweitigen Nutzung, beispielsweise für topozentrische Dienste, muss die explizite Datenverarbeitung erklärt und eine juristisch wirksame Einwilligung der betroffenen Person eingeholt werden. Des Weiteren ist sicherzustellen, dass der Nutzer seine Einwilligung kurzfristig widerrufen kann. Unklar ist hierbei allerdings die Homogenität der lokalen Umsetzung in nationales Recht. Zudem gibt es bislang wenige einheitliche Rege-

lungen zur Ausprägung der Einwilligungen – besonders für Ad-hoc-Nutzung von Diensten auf Mobiltelefonen. Hier gibt es bislang zwar Lösungsideen wie [Ross04], aber noch keine Standards. Aus Sicht von Nutzern und Netzbetreibern ist es erforderlich, die Einwilligungen beweissicher einzuholen und nachweisen zu können. Da der Netzanbieter zunächst bei der Herausgabe von Ortsdaten haftet, muss er eine Einwilligungsentscheidung treffen, bevor der Anbieter eines topozentrischen Dienstes persönliche Daten übermittelt bekommt. Dienstanbieter werden im Rahmen ihrer individuellen Angebote weitere Datenschutzvereinbarungen für ihre Nutzer vorrätig halten.

Eine Detailanalyse der rechtlichen Voraussetzungen netzbasierter topozentrischer Dienste findet sich in einem Forschungsprojekt des FP6/IST Programms der Europäischen Union in zwei Anforderungsanalysen für privatsphären-respektierende Anwendungen wieder [PRIM04a] [PRIM04b].

Wichtig neben der Erfüllung der formalen Datenschutzerfordernungen ist in diesem Kontext für Netzbetreiber besonders die Flexibilität der Detailausprägung der Infrastrukturkomponente für Privatsphärenmanagement. Sowohl die Geschäftsstrategie als auch die Nutzer mobiler Dienste verstehen diese Dienste als über nationale Grenzen hinausgehende Dienstleistungen. Für Sprachtelefonie und Daten existiert hierfür bereits das „International Roaming“. Bei der Implementierung internationaler Infrastrukturen für privatsphärenfreundliche topozentrische Dienste müssen lokal unterschiedlich ausgestaltete Gesetzgebungen zu Datenschutz, Daten-Vorratsspeicherung oder Überwachung durch Bedarfsträger berücksichtigt werden. Zur Vermeidung von Neuprojektierungen für jedes einzelne Land, in dem ein Netzbetreiber ein Geschäftsfeld eröffnet, lohnt es sich, konfigurierbare Intermediärsdienstleistungen für Privatsphärenmanagement einheitlich bereit zu halten und zu konfigurieren. Aus diesem Blickwinkel betrachtet stützt die inhomogene Rechtslage die Abwicklung von Identitäts- und Privatsphärenmanagement über einen Intermediär.

3 Lösungsansatz

Basierend auf den Anforderungen aus Kapitel 2 ist im Rahmen eines europäischen Forschungsprojektes der folgende Lösungsansatz für die Umsetzung eines datenschutz- und privatsphärenfreundlichen topozentrischen Prototypen entwickelt worden.

Der Lösungsansatz beinhaltet die Integration eines Intermediärs als zusätzliche Partei neben dem Netzbetreiber, dem Anbieter topzentrischer Dienste und deren Nutzer in der bisherigen Wertschöpfungskette des Mobile Commerce [KFKK05].

Der Intermediär erfüllt in dem Szenario drei wesentliche Aufgaben. Die erste Aufgabe beinhaltet die Gewährleistung der Anonymität des Nutzers gegenüber den an der Kommunikation beteiligten Parteien. Damit ist gemeint, dass seine Identität gegenüber dem Dienstanbieter topzentrischer Dienste und der konkrete angefragte Dienst gegenüber dem Netzbetreiber verschleiert wird. Jede Kommunikation zwischen Netzbetreiber und Dienstanbieter erfolgt ausschließlich über den Intermediär. Die dritte Aufgabe des Intermediärs ist die Repräsentation des Nutzers und die damit verbundene Vertretung seiner Interessen während der gesamten Kommunikationsbeziehung, wie dargestellt in [PRIM06a]. Die Vertretung erfolgt durch die Definition feingranularer Regeln mit deren Hilfe die Dienstverwendung bestimmt und/oder eingeschränkt wird. Ziel ist die Transparenz der zu übertragenden persönlichen Daten zu gewährleisten.

Der im Rahmen der Projektarbeit entwickelte Prototyp basiert auf einer bereits existierenden Struktur eines Netzbetreibers und stellt einen ortsbasierten Dienst zum Finden von Apotheken zur Verfügung. Alle im Kommunikationsprozess involvierten Parteien, mit Ausnahme des Nutzers, integrierten Softwarekomponenten in ihre bestehenden internen Strukturen. Der Nutzer erhält über ein Webinterface direkten Zugriff auf den Intermediär, um mit der für ihn vorgesehenen Identitäts- und Zugriffskontrollinstanz zu interagieren [PRIM06a]. Im dargelegten Szenario ist der Intermediär physisch der Infrastruktur des Netzbetreibers zugeordnet. Die Architektur ist in Abbildung 2 dargestellt.

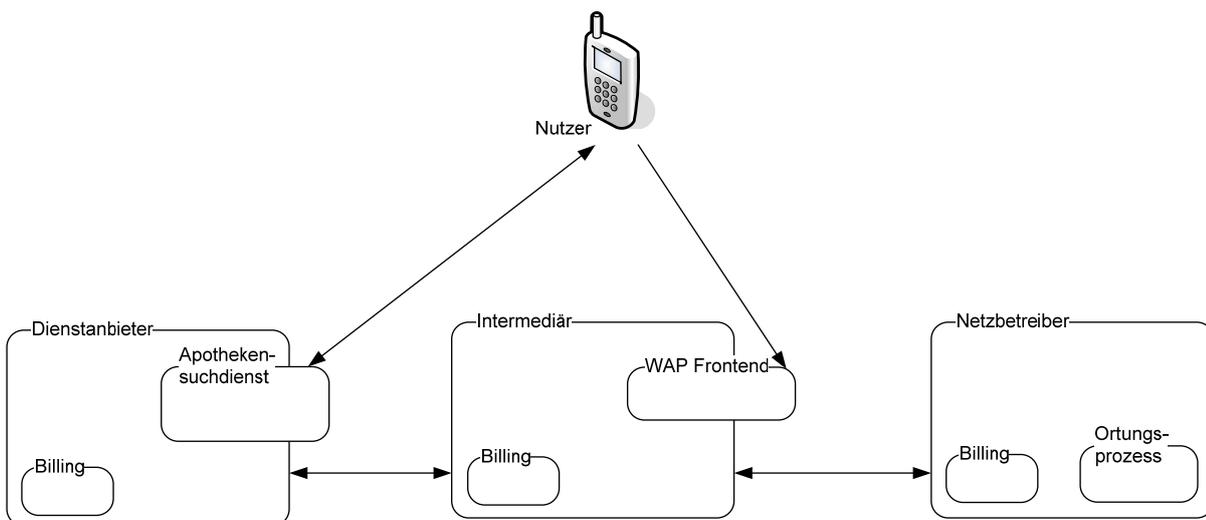


Abb. 2.: Architektur des Apothekensuchdienstes

Im Weiteren wird zunächst die Benutzerführung des Prototyps dargestellt.

3.1 Benutzerführung

Der Service begrüßt den Nutzer mit einem Bildschirm, von dem aus die Lokalisierung der nächstgelegenen Apotheke initiiert werden kann. Darüber hinaus ist es möglich, weitere Informationen über die Datenschutz-Funktionen des Prototyps abzurufen.



Abb. 3.: Hauptbildschirm

Wird der Dienst zum ersten Mal genutzt oder ist aus anderen Gründen keine gültige Richtlinie für die Handhabung der Ortsinformationen des Nutzers vorhanden, wird der Benutzer zu seiner Konsole beim Intermediär umgeleitet, wo er die Parameter der Datenschutzrichtlinie bearbeiten kann, bevor er sie bestätigt. Nach erfolgter Bestätigung wird der Benutzer – abhängig von seiner Eingabe – zurück zum Dienstanbieter oder aber zu einer Übersicht der eingestellten Datenschutzrichtlinien weitergeleitet. Sollte der Benutzer sich entscheiden, die neue Richtlinie nicht zu bestätigen, kann keine Nutzung des Dienstes erfolgen.



Abb. 4.: Einstellen einer Datenschutz-Richtlinie

Da nun ein durch die Benutzer-Richtlinie gesteuerter Datenaustausch zwischen Dienstanbieter und Intermediär erfolgen kann, ist eine Erbringung des Dienstes möglich - der Nutzer erhält das Ergebnis seiner Anfrage.



Abb. 5.: Ergebnis einer Ortung

3.2 Architektur

Die nachfolgenden Abschnitte betrachten den Nutzer, Dienstanbieter, Intermediär und Netzbetreiber im Detail. Neben der generellen Aufgabenbeschreibung wird auf die Beschreibung des Kommunikationsablaufs innerhalb des zugrunde liegenden Szenarios eingegangen. Abschließend wird in Abbildung 6 der gesamte Kommunikationsprozess zusammengefasst.

3.2.1 Nutzer

Bei einer Betrachtung aus der Perspektive des Nutzers ergeben sich besondere Anforderungen, die im Rahmen der Umsetzung berücksichtigt wurden. Der Nutzer des Apothekensuchdienstes ist im Besitz eines mobilen Endgerätes, das über ein kleines Display, eine beschränkte Eingabefunktionalität, geringe Rechenleistung und eine geringe Datenübertragungsrate verfügt. Diese Merkmale spiegeln die in der Masse der Bevölkerung vorzufindenden Endgeräte wieder [PRIM04a], die der Netzbetreiber als Zielgruppe adressiert. Des Weiteren ist vorgesehen, dass keine zusätzliche Software auf dem mobilen Endgerät des Nutzers zu installieren ist. Ein WAP Browser muss vorhanden sein, der zur Standardausstattung nahezu aller mobilen Endgeräte gehört [PRIM04a] [PRIM06a] [PRIM06b]. Der Zugriff auf die Identitäts- und Zugriffskontrolleinstanz des Nutzers ist durch ein WAP-Interface über den Intermediär möglich, dessen Verwendung dem Nutzer eine intuitive Handhabung erlaubt. Dadurch wird der Nutzer in die

Lage versetzt, Zugriffsregeln für Dienste, die er über einen Dienstanbieter in Anspruch nehmen möchte, zu setzen.

Verwendet ein Nutzer einen Dienst zum ersten Mal, ist keine Zugriffregel beim Intermediär definiert, die den Zugriff auf Ortinformationen oder Nutzerpräferenzen regelt. In diesem Fall wird eine Standardregel vorgeschlagen. Der Regelvorschlag wird beim Intermediär hinterlegt. Der Nutzer wird zum Intermediär weitergeleitet und erhält die Optionen den Regelvorschlag zu akzeptieren, zu überarbeiten oder abzulehnen [PRIM06a]. Dieser Mechanismus verhindert, dass personenbezogene Informationen über einen Nutzer ohne dessen explizite Einwilligung verarbeitet werden. Das Entgelt für die Dienstleistung wird durch den Netzbetreiber in einem monatlichen Intervall abgerechnet. Durch den Intermediär ist ein weiterer Schutzmechanismus geschaffen, der keinem Anbieter die Möglichkeit offeriert, einen Geldwert für die Inanspruchnahme in Rechnung zu stellen, ohne dass der Nutzer explizit, beispielsweise durch eine Zugriffregel dieser Dienstonutzung zugestimmt hat.

3.2.2 *Dienstanbieter*

Im vorliegenden Abschnitt wird der Blickwinkel des Dienstanbieters (Apothekensuchdienst) detailliert betrachtet. Dafür ist es notwendig Annahmen zu treffen. Der Dienstanbieter ist eine Partei innerhalb des gesamten Kommunikationsablaufs, die einer Nutzergruppe durch das Betreiben eines Webservers gegen Entgelt einen oder mehrere Dienste zur Verfügung stellt. Die Preise für die Nutzung eines Dienstes entsprechen definierten Preiskategorien des Netzbetreibers. Nach jeder autorisierten Transaktion (Dienstonutzung) belastet der Intermediär ein virtuelles Konto des Nutzers. Die dort auflaufenden Beträge werden einmal im Monat durch den Netzbetreiber abgerechnet [PRIM06a]. Damit der Intermediär die entstandenen Kosten des Dienstanbieters gegenüber dem Netzanbieter vertreten darf, sind an dieser Stelle weitere Verträge zwischen beiden Parteien notwendig. Als Identifikationsmerkmal des Nutzers wird die IP-Adresse verwendet. Um die Dienstonutzung aus Sicht des Dienstanbieters vollständig nachzuvollziehen, wird diese am Beispiel der Erstonutzung des Apothekensuchdienstes betrachtet.

Der Apothekensuchdienst ist durch einen beim Dienstanbieter betriebenen Webserver im Internet erreichbar. Der Nutzer meldet sich beim Dienstanbieter an, um einen Dienst zu verwenden. Dabei wird eine Instanz der sich beim Dienstanbieter befindlichen Softwarekomponente initialisiert. Diese sorgt für den datenschutz- und privatsphärenfreundlichen Austausch personenbezogener Daten. Nachdem sich der Nutzer beim Dienstanbieter für die Nutzung eines Dienstes angemeldet hat, benötigt dieser zusätzliche Informationen zur Dienstleistung, wie beispielswei-

se den Aufenthaltsort des Nutzers. Darüber hinaus ist sicherzustellen, dass der Anbieter des Dienstes den Geldwert für die Dienstnutzung abrechnen kann [PRIM06a].

Um Zugriff auf die Ortsinformation des Nutzers zu erlangen und den Geldwert für die Dienstnutzung einzufordern, hat der Dienstanbieter dem Intermediär den Dienstnamen und die Kosten für die Dienstnutzung mitzuteilen. Vorausgehend ist eine Authentifizierung des Dienstanbieters beim Intermediär erforderlich. Nach dem Abgleich der Dienstinformationen mit den durch den Nutzer hinterlegten Regeln (Intermediär) ist der Zugriff auf die Ortsinformationen gestattet oder versperrt. Für den Fall, dass der Zugriff auf die Ortsinformation und die Reservierung des Geldwertes, aufgrund fehlender oder sperrender Regeln, verwehrt ist, meldet die Softwarekomponente einen Fehler [PRIM06a]. Nach Erhalt der Fehlermeldung wird die Dienstanbieterseite einen Regelentwurf für die Nutzung des Dienstes generieren und an den Intermediär übertragen. Der Nutzer wird über den Regelvorschlag informiert und wahlweise direkt zum Intermediär weitergeleitet. Dort nimmt der Nutzer Zugriff auf seine Identitäts- und Zugriffskontrollinstanz. Die Regelvorlage des Dienstanbieters wird angezeigt. Der Nutzer hat die Möglichkeit diese Regel zu überprüfen, anzunehmen oder abzulehnen. Im Fall der Annahme erfolgt erneut eine Weiterleitung zum Dienstanbieter, der aufgrund der akzeptierten Regeln alle notwendigen Informationen für die Dienstleistung erhält [PRIM06a]. Ein virtuelles Konto des Nutzers, hinterlegt beim Intermediär, wird mit dem Geldwert der Dienstnutzung belastet, ein weiteres virtuelles Konto für Dienstanbieter mit den Kosten für die Ortung. Auf Basis der übermittelten Daten ist der Dienstanbieter in der Lage, seine Datenbank zu durchsuchen und dem Nutzer die nächstliegende Apotheke mitzuteilen. Alle benutzerrelevanten Daten werden nach Beendigung der Dienstleistung mit der Zerstörung der Session verworfen.

3.2.3 *Intermediär*

Der Intermediär ist eine zentrale Komponente der Architektur und befindet sich zwischen Dienstanbieter, Netzbetreiber und Nutzer [KFKK05]. Zu den ihm zugeteilten Aufgaben (siehe Kapitel 3 Einleitung) gehört die Repräsentation des Nutzers. Der Intermediär erhält Anfragen eines Dienstanbieters, dessen Dienst der Nutzer verwenden möchte. Eine Überprüfung der Nutzung hinsichtlich der vom Nutzer hinterlegten Regeln und der möglichen Zahlungsfähigkeit wird durchgeführt. Der detaillierte Ablauf der Kommunikation wird im Folgenden dargestellt.

Der Dienstanbieter, der mit dem Nutzer verbunden ist und diesen über eine IP-Adresse identifiziert, erfragt beim Intermediär die Ortsinformation der zugeordneten IP-Adresse an. Darüber hinaus teilt der Dienstanbieter dem Intermediär den festgesetzten Preis für die Inanspruchnahme

des Dienstes mit. Der Intermediär löst mit Hilfe des Netzbetreibers die IP-Adresse auf und überprüft die vom Nutzer beim Intermediär hinterlegten Regeln für die Dienstnutzung. Die aktuelle IP-Adresse stellt ein temporäres Identifikationsmerkmal dar, das sich nach erneutem Verbindungsaufbau verändert. Aus diesem Grund wird die IP-Adresse in ein Transaktionspseudonym überführt, das zwischen Intermediär und Dienstanbieter kommuniziert wird. Falls keine für die Dienstnutzung adäquate Regel existiert, ist die Nutzung ausgeschlossen und dem Dienstanbieter wird eine Fehlermeldung gemeldet. Die Übermittlung eines Regelentwurfs für die Nutzung des Dienstes erfolgt. Nach Zustimmung wird die Regel angewendet. Die Ortsinformation des Nutzers wird über den Intermediär beim Netzbetreiber angefragt, ohne Informationen über den zu verwendenden Dienst mitzuteilen. Ferner wird überprüft, ob die Zahlungsfähigkeit des Nutzers gegeben ist. Ist dies der Fall, wird die Ortsinformation über den Intermediär an den Dienstanbieter übermittelt. Die Kosten für die Ortung werden dem Dienstanbieter in einem virtuellen Konto beim Intermediär in Rechnung gestellt. Auf Basis der übertragenden Ortsinformation wird der Dienst durch den Dienstanbieter erbracht. Dem Nutzer werden die Kosten für den Dienst in Rechnung gestellt. Das virtuelle Konto des Nutzers wird mit den Kosten der Dienstnutzung belastet [PRIM06a]. Am Ende des Monats stellt der Intermediär zwei Rechnungen. Eine Rechnung, die dem Netzbetreiber zugestellt wird, beinhaltet die Kosten der Dienstnutzung ohne Bekanntgabe des Dienstes. Die zweite Rechnung über die Anzahl der Ortungen in Verbindung mit dem jeweiligen Transaktionspseudonym erhält der Dienstanbieter.

3.2.4 Netzbetreiber

Der Netzbetreiber ist die Partei, die eine traditionelle Geschäftsbeziehung mit dem Nutzer unterhält. Diese ermöglicht die Kommunikation, die Ortung und die Abrechnung zwischen Nutzer und Dienstanbieter. Die Softwarekomponente, die beim Netzbetreiber integriert ist, dient einerseits zur Übertragung der personenbezogenen Daten und andererseits als Kontrollinstanz zum Nachweis autorisierter Transaktionen [PRIM06a] [PRIM06c].

Der Netzbereiter wird im Szenario des Apothekensuchdiensts mehrfach involviert. Zu Beginn erfragt der Intermediär die Ortsinformation der ihm bekannten IP-Adresse. Im Anschluss wird erfragt, ob der Nutzer in der Lage ist, den zu verwendenden Dienst zu bezahlen. Die Antwort auf beide Fragen wird an den Intermediär übertragen. Da der Netzbetreiber den Dienstanbieter nicht kennt, werden die Kosten der Ortung dem Intermediär in Rechnung gestellt. Nach erfolgreicher Dienstnutzung informiert der Intermediär den Netzbetreiber, so dass dieser dem Nutzer

die Kosten der Dienstnutzung in Rechnung stellen kann [PRIM06a]. Die Abrechnung erfolgt einmal im Monat.

Zusammenfassend veranschaulicht das UML-Sequenzdiagramm in Abbildung. 6 nochmals den Kommunikationsablauf.

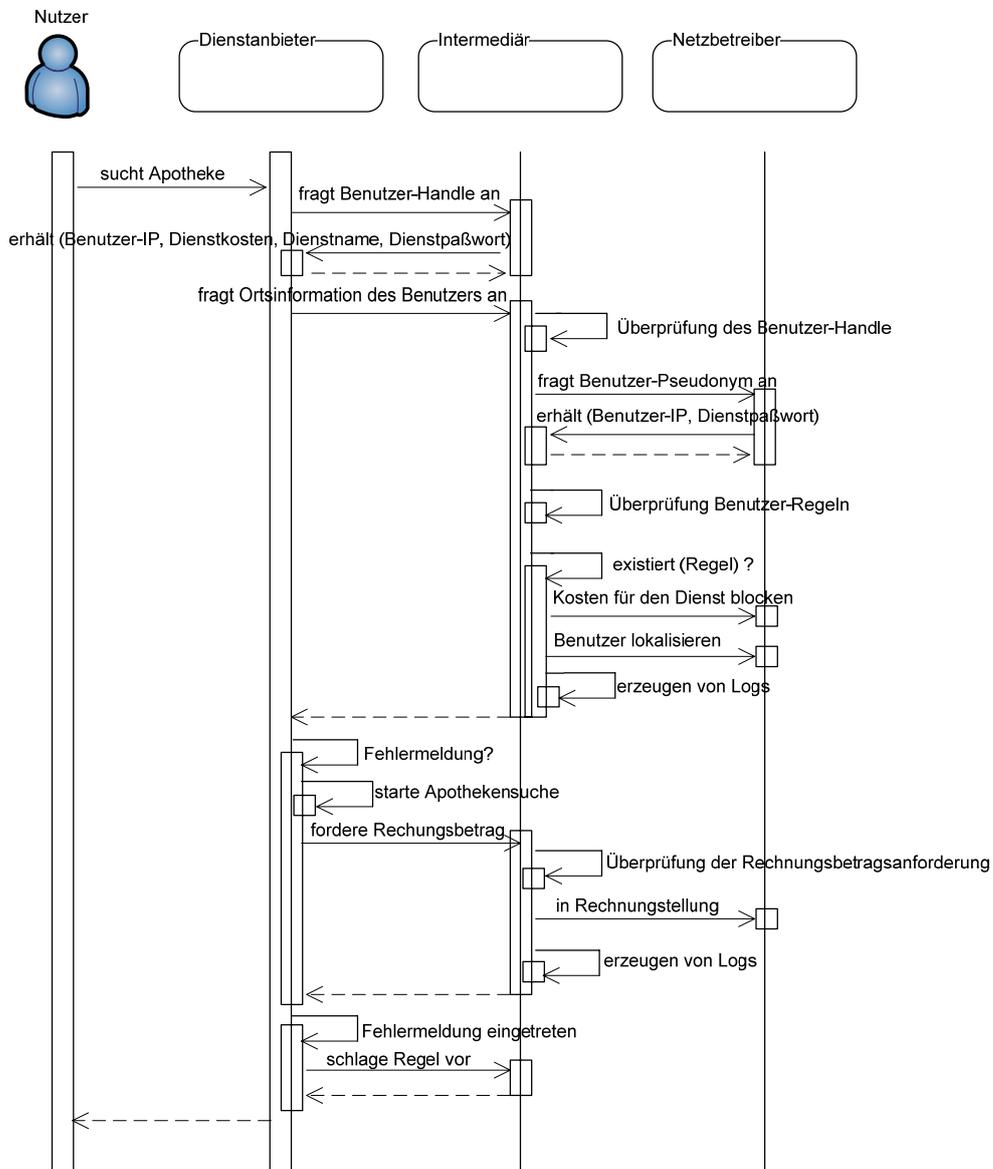


Abb. 6.: Sequenzdiagramm

4 Zusammenfassung und Ausblick

Wie gezeigt wurde, lässt sich unter den besonderen Rahmenbedingungen mobiler Datenprodukte auf bestehenden Infrastrukturen ein Ausgleich der Interessen der Beteiligten erreichen. Im Rahmen des bestehenden Prototyps für privatsphären-respektierende topozentrische Dienste

wurde eine Architektur erprobt, welche die Geschäfts- und Privatsphäreninteressen der Beteiligten umsetzt. Die wurde aus der Sicht verschiedener Disziplinen evaluiert [PRIM06d]. Die Spezialisten aus den Bereichen Recht, Assurance, HCI, und Sozioökonomie hatten die Aufgabe, in ihrem Spezialbereich die Prototypen auf Anwendbarkeit für Nutzer und beteiligte Businesspartner zu untersuchen. Die dabei gewonnenen Erkenntnisse wurden in einem Evaluationsbericht [PRIM05] zusammengefasst.

Die beteiligten Disziplinen sind in ihrem Resümee zu dem Urteil gekommen, dass der Prototyp für topozentrische Dienstleistung eine sehr gute Balance zwischen den Interessen der einzelnen Beteiligten darstellt. Insbesondere konnte gezeigt werden, dass die Berücksichtigung der Privatsphäre eine positive Voraussetzung für Anwendungsfälle sein kann, ohne dass die zugehörigen Prozesse für Nutzer und Dienstleister hochkomplex in Betrieb und Nutzung und somit unattraktiv für Anbieter und Nutzer werden.

Eine Standardisierung der Intermediärsschnittstelle würde eine vom Netzbetreiber unabhängige Verwendung topozentrischer Dienste, etwa im Falle von International Roaming, erlauben. Zusätzlich könnte eine solche Identitätsmanagement-Funktionalität dem Benutzer eine konsistente Kontrolle seiner persönlichen Informationen ermöglichen, was - gerade bei kritischen Nutzern - eine höhere Akzeptanz nach sich ziehen könnte. Daher regen die Autoren an, die Vereinheitlichung von Intermediationsschnittstellen für Identitäts- und Privatsphärenmanagement im internationalen Rahmen anzustreben und diese im industriellen Rahmen umzusetzen. So könnte ein internationaler Durchbruch für topozentrische Dienstleistungen erreicht - und beispielsweise von aktuellen europäischen Bestrebungen bei der Ortung von Notrufen bereits eingesetzt werden oder der Verbreiterung des Szenarios im Sinne von [FrSc05] dienen.

Danksagung

Die Inhalte dieses Papiers wurden im Rahmen der Projektarbeit im PRIME Projekt erarbeitet, repräsentieren allerdings nur die Ansicht der Autoren.

Literaturverzeichnis

- [BIBO03] Blarkom, G.W.; Borking, J.; Olk, J.G.: Handbook of Privacy and Privacy-Enhancing Technologies. College bescherming persoonsgegevens, The Hague. 2003.
- [BöLR02] Böhm, A.; Leiber, T.; Reufenheuser, B.: Trust and Transparency in Location-Based Services: Making Users lose their Fear of Big Brother In: Workshop on Location Systems Privacy and Control, 2004.
- [BuHE00] Bulusu, N.; Heidemann, H.; Estrin, D.: GPS-less low cost outdoor localization for very small devices. In: Technical report 00-729, Computer science department, University of Southern California. 2000.
- [Bund05] Bundesnetzagentur: Jahresbericht. 2005.
- [Comm05] Common Criteria Project: The Common Criteria Part 1 - Introduction and general model, Version 2.3, similar to IS 18045. 2005.
- [Dumo05] Dumortier, J.: Combining Personalised Communications Services with Privacy-Friendly Identity Management. In: Proceedings of the 44th FITCE Congress, 2005, 142-146.
- [Euro02] European Parliament: Directive 2002/58/EC of the European Parliament and of the council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). Brussels: 2002.
- [Euro95] European Parliament: Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Luxembourg. 1995.
- [Figg04] Figgie, S.: Situation-dependent services - a challenge for mobile network operators. In: Journal of Business Research, 57, 2004, 1416-1422.

- [FrSc05] Fritsch, L. und Scherner, T.: A Multilaterally Secure, Privacy-Friendly Location-based Service for Disaster Management and Civil Protection. In Proceedings of the AICED/ICN 2005, (Springer Lecture Notes on Computer Science LNCS 3421), Berlin, Heidelberg, New York 2005 S. 1130-1137.2005, 1130-1137.
- [FrSR06] Fritsch, L.; Scherner, T.; Rannenber, K.: Von Anforderungen zur verteilten, Privatsphären-respektierenden Infrastruktur. In: Praxis in der Informationsverarbeitung und Kommunikation (PIK), 29, 2006, 37-42.
- [Inter05a] International Organization for Standardization (ISO): The Common Criteria Part 2 - Security functional requirements, Version 2.3, similar to IS 18045:2004. 2005.
- [Inter05b] International Organization for Standardization (ISO): The Common Criteria Part 3: Security assurance requirements - Version 2.3, similar to IS 18405. 2005.
- [JoBe04] Jorns, O, -Bessler, S.: PRIVES: A privacy enhanced location based scheme. In: Workshop on Location Systems, Privacy and Control, 2004.
- [KFKK05] Koelsch, T.; Fritsch, L.; Kohlweiss, M.; Kesdogan, D.: Privacy for Profitable Location Based Services. In: Proceedings of the Security in Pervasive Computing Workshop (SPC), 2005, 164-179.
- [LFPR04] Lindner, T.; Fritsch, L.; Plank, K.; Rannenber, K.: Exploitation of Public and Private WiFi Coverage for New Business Models. In: Proceedings of the 4th IFIP Conference on E-Commerce, E-Business, and E-Government (I3E),2004.
- [MyAD03] Myles, G., Friday, A. and Davies, N.: Preserving Privacy in Environments with Location Based Applications. In: IEEE Pervasive Computing 2, (2003) 1, S. 56-64.
- [Oino02] Oinonen K. TR101 - LIF Privacy Guidelines. 2002.
- [PRIM04a] PRIME Project: Application Requirements. In deliverable 1.1.a part 3, 2004.
- [PRIM04b] PRIME Project: Legal Requirements. In part 1 of deliverable 1.1a of IST PRIME EU. Arizona, 2004

- [PRIM05] PRIME Project: Framework V.1. In deliverable 14.1.a. 2005
- [PRIM06a] PRIME Project: Pharmacy Search - High Level Design. In deliverable 4.1.a.3.8, 2006.
- [PRIM06b] PRIME Project: LBS Installation Guide. In deliverable 4.1.a.3.10, 2006
- [PRIM06c] PRIME Project: User documentation on the LBS application prototype – pharmacy search. In Deliverable 4.1.a.3.10, 2006
- [PRIM06d] PRIME Project: Evaluation of initial Application Prototypes. 2006.
- [Pone04] Ponemon Institute: The Cost of Privacy - Study, Tucson, Arizona, 2004.
- [RePr04] Cheng, Reynold; Prabhakar, Sunil: Using Uncertainty to Provide-Preserving and High-Quality Location-Based Services, in: Workshop on Location Systems, Privacy and Control, 2004.
- [Rose99] Rose, F.: The economics, concept and design of information intermediaries. Physica-Verlag, Heidelberg, 1999
- [Ross04] Rossmagel, H.: Mobile Qualified Electronic Signatures and Certification on Demand, in: Proceedings of the 1st European PKI Workshop - Research and Applications. 2004.
- [ScLi02] Schackmann, J.; Link, H.: Intermediaries for the Provision of Mass Customized Digital Goods in Electronic Commerce, in: Moving into Mass Customization, 2002.
- [SNP02] Synnes, K.; North, J.; Parnes, P.: Location Privacy in the Alipes Platform. Institutionen för Systemteknik; Lulea University of Technology; 2002
- [Wage06] Wagner, F.: T-Identity Protector – Functional operation. In W3C Workshop on Languages for Privacy Policy Negotiation and Semantics-Driven Enforcement, Ispra. 2006

Allokation und Bepreisung kontextbezogener Kundenkontakte im Mobile Marketing

Rahmenbedingungen für Einsatz und Gestaltung von multi-attributiven Auktionen

Andreas Albers

Professur für M-Commerce und Mehrseitige Sicherheit
Goethe Universität Frankfurt a. M.
60054 Frankfurt a. M.
andreas.albers@m-lehrstuhl.de

Abstract

Die Einbindung von Identitäts- und Kontextinformationen in Mobile Marketing Aktivitäten eignet sich besonders für Kundenkontakte, die durch Zugriff auf mobile Portale generiert wurden. Allerdings erfordert es die Etablierung von Informationsintermediären, um diese *kontextbezogenen Kundenkontakte* effizient an Werbetreibende zu vermitteln. Aus dem individuellen Wert von mobilen Kundenkontakten für Werbetreibende sowie der Gewährleistung relevanter werblicher Kommunikation, ergeben sich deshalb besondere Anforderungen für die Allokation und Bepreisung dieser Kundenkontakte. Der vorliegende Beitrag adressiert diese Problemstellung mit einer Diskussion der Rahmenbedingungen für den Einsatz und die Gestaltung eines interaktiven Preismodells in Form von multi-attributiven Auktionen.

1 Einleitung

1.1 Die mobile Datenkommunikation zur Umsetzung von Marketing Aktivitäten

Die Nutzung mobiler Datenkommunikation zur Umsetzung und Unterstützung von Marketingaktivitäten im Business-To-Consumer (B2C) Bereich, insbesondere zur *personalisierten* werblichen Kommunikation, wird meist unter *Mobile Marketing* bzw. *Mobile Advertising* subsumiert. Wenn auch in der Literatur bzgl. dieser Begrifflichkeiten keine Einigkeit besteht

[LeSK06], so herrscht hinsichtlich der besonderen Eignung des mobilen Kanals zur Ausbringung absatzfördernder Maßnahmen ein allgemeiner Konsens [BaSt02, SaTä05, Wohl01]. Der Personenbezug fast ständig mitgeführter [Rösc06] mobiler Endgeräte in Kombination mit der Interaktivität des mobilen Kanals, erlaubt Werbetreibenden eine orts- und zeitunabhängige, bidirektionale Kommunikation mit ihren Rezipienten. Die Möglichkeit zur Identifikation bzw. Wiedererkennung mobiler Endgeräte schafft zudem die Voraussetzungen für eine Zuordnung von ggf. hinterlegten Personenprofilen sowie der Bestimmung des Aufenthaltsorts und Nutzungszeitpunkts mobiler Nutzer [DBHS06]. Diese Informationen, dynamisch zusammengeführt zu einem kontextbezogenen Kundenprofil, ermöglichen eine gezielte Auswahl und personalisierte Ansprache von Rezipienten und damit die Grundlage zur Umsetzung des als idealtypisch angesehenen One-to-One Marketing Paradigmas [PeRo97]. Zudem verschafft die allein in Deutschland Ende 2005 vorhandene Mobilfunkpenetration von 96% [BUNA06] dem mobilen Medium eine potentiell große Reichweite in der Bevölkerung.

Der Blick auf die aktuelle Marktsituation in Deutschland zeigt jedoch, dass nur vergleichsweise einfache Implementierungen von Mobile Marketing Anwendungen existieren, welche nur einen Teil des Potentials der mobilen Datenkommunikation ausnutzen [DBHS06]. Umsetzungen auf Basis des *Push-Ansatzes*, bei denen Nachrichten ohne vorherige Aufforderung übermittelt werden, nutzen überwiegend die Möglichkeit Rezipienten personenbezogenen sowie orts- und zeitunabhängig anzusprechen. Über andere Marketing Aktivitäten bzw. Medien werden dazu Kundenkontakte generiert und profiliert in einer Datenbank gespeichert. Im Rahmen einer Direktmarketing Kampagne können dann Rezipienten durch diese Personenprofile ausgewählt und Werbebotschaften per SMS-Nachricht an sie versandt werden [Lipp04].

Bei Nachrichten, die erst nach vorausgegangener Aufforderung durch einen Rezipienten versendet werden, wird vom *Pull-Ansatz* gesprochen. Diese Form der Kommunikation wird hauptsächlich durch andere Medien initiiert. Beispielsweise werden spezielle Codes auf Produktverpackungen gedruckt, die der mobile Nutzer gegen den Erhalt von Sachleistungen per SMS-Nachricht an Werbetreibende einsendet und damit den werblichen Dialog anstößt [Rösc06].

Die Differenz zwischen theoretischem Potential und der dargestellten Marktsituation im Kontext des Mobile Marketings zeigt vor allem, dass kaum von der Möglichkeit zur Identitäts- und Kontextdeterminierung mobiler Nutzer Gebrauch gemacht wird. Als Gründe dafür werden hauptsächlich die derzeit damit verbundenen Kosten und die komplexe Technologie genannt [Rösc06]. Lediglich beim Push-Ansatz findet der ortsunabhängige Versand von Nachrichten zu

präferierten Tageszeiten statt, benötigt aber komplexe Entscheidungsmodelle zur Einbeziehung weiterer Kontextinformationen [Trip03]. Zu dem steht der Push-Ansatz der bedürfnisgesteuerten Initiierung einer werblichen Kommunikation durch Rezipienten entgegen und generiert geringere Reaktionsraten. Dennoch hat er in der Praxis insbesondere für zeitkritische Nachrichten (z.B. Promotions) seine Berechtigung gefunden [Berl03].

Die in beiden Ansätzen letztlich durch SMS-Nachrichten gestaltete Interaktion stellt sich als meist umständlich dar und verhindert den Transport von multimedialen Inhalten [Lipp04].

1.2 Zukünftige Ansätze für Mobile Marketing Aktivitäten

Für die Zukunft von Mobile Marketing Aktivitäten sehen Anbieter die Nutzung von mobilen Portalen (Pull-Ansatz) als besonders geeignet und viel versprechend an [Rösc06]. Vor dem Hintergrund, dass die Mobilität eine der dominierenden Motivationen zur Befriedigung von Konsumbedürfnissen ist [Zäng00], erhöht sich damit die Wahrscheinlichkeit, eine nutzerseitig initiierte werbliche Kommunikation in eine geschäftliche Transaktion zu überführen. Die Einbeziehung von Identitäts- und Kontextinformationen ist dazu im Vergleich zum Push-Ansatz mit deutlich weniger konzeptioneller Komplexität verbunden [Trip03] und ermöglicht es, die Rezipienten gezielter auszuwählen und personalisiert anzusprechen. Zu dem kann die eingeschränkte, textbasierte Kommunikation auf Basis von SMS-Nachrichten durch eine intuitive, multimediale und webbasierte Benutzeroberfläche ersetzt und gleichzeitig der Medienbruch im Dialog zwischen Werbetreibenden und Rezipienten beseitigt werden.

Zur Integration von Identitäts- und Kontextinformationen in pull-basierte Mobile Marketing Aktivitäten, schlägt Figge in seiner Arbeit *Konzept eines Entwurfsmusters für werbefinanzierte Geschäftsmodelle im M-Commerce* [Figg06] vor, Mobilfunkbetreiber als Informationsintermediäre zu etablieren, um dadurch Werbetreibende und Rezipienten auf Basis ihrer gegenseitigen Anforderungen (Zielgruppe und Kundenbedürfnisse) individuell zusammenzuführen. Mobilfunkbetreiber sind nach Figge deshalb besonders geeignet, da sie bereits eine große Anzahl von bestehenden Kundenbeziehungen besitzen, die benötigte technische Infrastruktur vorhanden ist, Identitäts- und Kontextinformationen ihnen zu großen Teilen kostenfrei zur Verfügung stehen und sie ihre vorhandenen mobilen Portale als Kontaktfläche positionieren können.

Den stark vereinfachten Ablauf der Vermittlung mobiler Kundenkontakte im Intermediär- bzw. Mobilfunkbetreiber Szenario zeigt Abbildung 1.

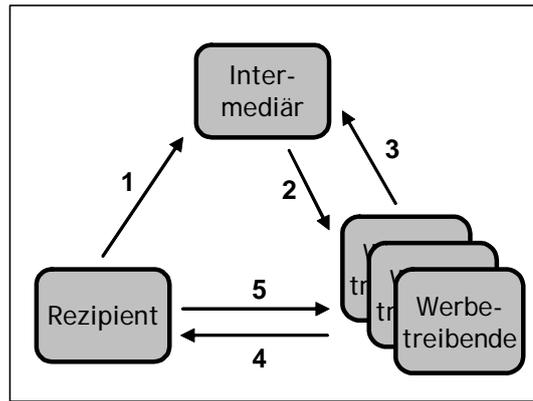


Abb. 1: Vermittlung kontextbezogener mobiler Kundenkontakte an Werbetreibende

Nach der Generierung und Erfassung des mobilen Kundenkontakts und der evt. Anreicherung mit Kontextinformationen (1), bietet der Intermediär diesen Kundenkontakt Werbetreibenden an (2). Sie prüfen die Relevanz des Rezipienten, bekunden ggf. Interesse und bekommen den Kundenkontakt gegen die Zahlung eines Entgelts zugeteilt. Im Anschluss beginnen sie ihre werbliche Kommunikation mit dem Rezipienten (4). Bei erwidertem Interesse des Rezipienten kommt es schließlich zu einem werblichen Dialog oder im weiteren zu einer geschäftlichen Transaktion (5).

1.3 Allokation und Bepreisung von mobilen Kundenkontakten

Der durch mobile Portale generierte und ggf. mit Identitäts- und Kontextinformationen angereicherte *kontextbezogene mobile Kundenkontakt* stellt für einen Intermediär eine Herausforderung hinsichtlich der Bepreisung und Allokation dar. So besitzt der Kundenkontakt für Werbetreibende in Bezug auf ihre definierte Zielgruppe einen individuellen Wert, der für Intermediäre nur schwer zu bestimmen ist. Im Weiteren besteht dieser Wert nur für die Dauer der Ansprache des Rezipienten und verfällt danach. Die starke Verankerung der mobilen Kommunikation in der Privatsphäre von mobilen Nutzern birgt dazu ein hohes Störungspotential [Meie02] durch unrelevante werbliche Nachrichten.

Ziel des vorliegenden Beitrags ist deshalb ein Preis- bzw. Allokationsmodell zu identifizieren, mit dem die genannten Herausforderungen adressiert werden können. Dazu diskutiert Kapitel 2 die Charakteristika kontextbezogener mobiler Kundenkontakte sowie relevante Identitäts- und Kontextinformationen zur personalisierten Kundenansprache und stellt den Einfluss von wirtschaftlichen, technologischen und datenschutzrechtlichen Aspekten auf die Erhebung dieser Informationen dar. In Kapitel 3 wird für die Allokation ein interaktives Preismodell in Form

einer multi-attributiven Auktion vorgeschlagen während in Kapitel 4 die Designaspekte für dessen Einsatz und Gestaltung diskutiert werden. Mit einer Zusammenfassung sowie einem Ausblick auf zukünftige Forschungsfragen schließt Kapitel 5.

2 Charakteristika kontextbezogener mobiler Kundenkontakte

Kontextbezogene mobile Kundenkontakte lassen sich gegenüber Kontakten aus den klassischen Massenmedien durch die folgenden Charakteristika abgrenzen:

- Mobile Kundenkontakte haben durch ihren möglichen Identitäts- und Kontextbezug für Werbetreibende jeweils einen privaten, individuellen Wert.
- Mobile Kundenkontakte besitzen aufgrund des möglichen Identitäts- und Kontextbezugs nur für die Dauer der Rezipientenansprache einen Nutzen für Werbetreibende.
- Mobile Kundenkontakte können nicht an beliebige Werbetreibende veräußert werden, da unrelevante werbliche Nachrichten ein hohes Störungspotential besitzen [Meie02].
- Der Umfang und die Qualität von Identitäts- und Kontextinformationen sind für jeden mobilen Kundenkontakt unterschiedlich.

Im Vergleich zu den klassischen Massenmedien, wo ein Kundenkontakt eine durch Qualitätsmerkmale beschriebene Gruppe von Rezipienten darstellt [Bogs01], adressiert der mobile Kundenkontakt nur einen Rezipienten, der aufgrund seines einzigartigen Kontexts (z.B. hinsichtlich des Aufenthaltsorts) für Werbetreibende einen individuellen Wert darstellt. Da nach Beendigung eines werblichen Dialogs mit einem Rezipienten sich dessen Kontext ändert (z.B. durch Verlassen des mobilen Portals), verliert der Kundenkontakt durch den veralteten Kontext seinen Wert. Durch den Stellenwert von mobilen Endgeräten als persönliche Accessoires ist die mobile Kommunikation stark in der Privatsphäre eines mobilen Rezipienten verankert. Je stärker also der Eingriff mobiler Kommunikation in die Privatsphäre von Nutzern, desto größer die Aufmerksamkeit für werbliche Nachrichten aber auch die Gefahr von deren Belästigung [Meie02]. In diesem Spannungsfeld nimmt die Nutzung von Identitäts- und Kontextinformationen mobiler Rezipienten zur Gewährleistung einer hohen Relevanz werblicher Kommunikation eine zentrale

Rolle ein [Meie02]. Der Umfang dieser Informationen variiert jedoch aufgrund verschiedener Einflüsse für jeden Kundenkontakt und wird deshalb für die, in einer Expertenumfrage [FiTh06] als relevant identifizierten Informationen *Identität, Personenprofil, Aufenthaltsort* und *Aktionskontext*, diskutiert. Der *Nutzungszeitpunkt* ist an dieser Stelle ebenfalls wichtig, kann aber jederzeit ohne besondere Voraussetzungen erhoben werden.

Die Ermittlung der Identität eines mobilen Nutzers durch das Mobilfunknetz ermöglicht die Zuordnung eines ggf. hinterlegten Personenprofils. Diese Profile enthalten derzeit meist soziodemographische Daten sowie Interessen eines Nutzers [Mind06]. Die Kenntnis des Aufenthaltsorts ermöglicht Werbetreibenden die Relevanz von mobilen Nutzern hinsichtlich eines ortsgebundenen Leistungsangebots zu ermitteln (z.B. Entfernung zu einem Point-of-Sales (POS)) während der Aktionskontext das Bedürfnis eines mobilen Nutzers in seiner aktuellen Situation (z.B. ableitbar aus der Eingabe eines Suchbegriffs) beschreibt. Bei der Erhebung dieser Identitäts- und Kontextinformationen beeinflussen *datenschutzrechtliche Aspekte*, die *verfügbare Infrastruktur* sowie entstehende *Kosten* deren Umfang und Genauigkeit und letztlich den Wert des Kundenkontakts für Werbetreibende (vgl. Abb. 2).

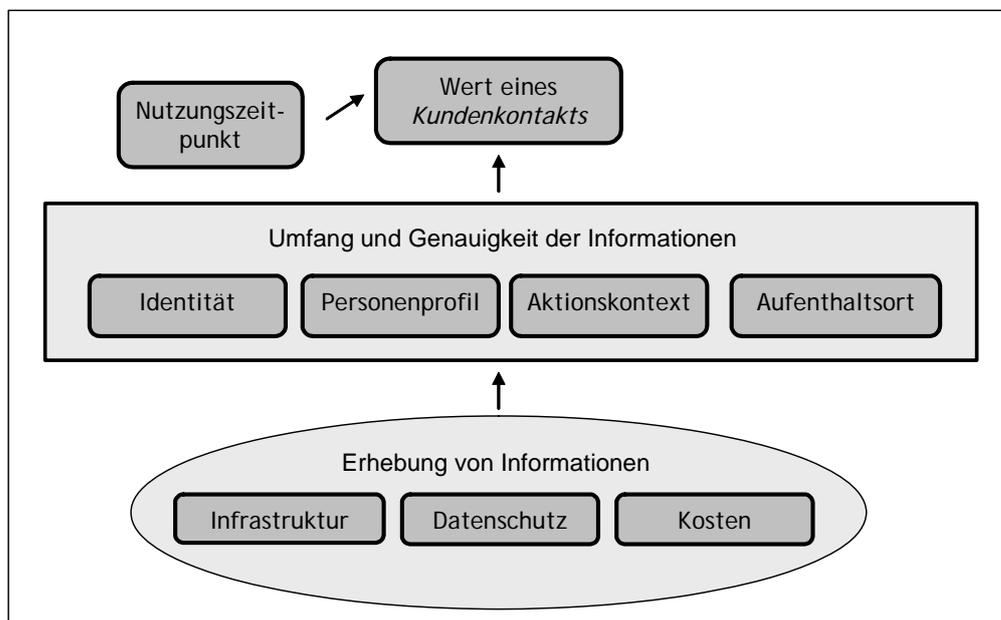


Abb. 2: Einflussfaktoren auf den Wert kontextbezogener mobiler Kundenkontakte

Aus Gründen des *Datenschutzes* müssen mobile Nutzer für die Erhebung von personenbezogenen Daten ihre Einwilligung erteilen und diese ggf. auch wieder zurücknehmen können [Figg06]. Daraus resultiert, dass der Umfang der zur Verfügung gestellten Daten (mit Ausnah-

me des Nutzungszeitpunkts) vom Bedürfnis nach Privatsphäre oder dem Nutzen, der aus der Preisgabe von persönlichen Daten resultiert, abhängt [NSRH02].

Die *verfügbare Infrastruktur* hingegen determiniert, welche Arten von Kontextinformationen zu welcher Qualität erhoben werden können. Für die Zuordnung eines Personenprofils muss das Mobilfunknetz die Identifikation eines mobilen Nutzers erlauben. Gleiches gilt für die Lokalisierung des Aufenthaltsortes, welche das Mobilfunknetz zwar immer zur Verfügung stellen kann, jedoch das klassische Cell-of-Origin (COO) Verfahren nur sehr ungenau arbeitet. Das Global Position System (GPS) liefert hingegen sehr präzise Daten, ist aber nur derzeit in wenigen mobilen Endgeräten vorhanden [AIFR05].

Letztlich können auch *Kosten* über den Umfang von Identitäts- und Kontextinformationen entscheiden. So steht beispielsweise die Abfrage von Identitäts- und Lokalisierungsdaten nur Mobilfunkbetreibern kostenlos zur Verfügung [Figg06].

3 Preismodelle für kontextbezogene mobile Kundenkontakte

3.1 Identifikation eines Preismodells für kontextbezogene mobile Kundenkontakte

Für die im vorausgegangenen Kapitel diskutierten Charakteristika kontextbezogener mobiler Kundenkontakte soll im Folgenden ein adäquates Preismodell identifiziert werden. Dazu ist zunächst zwischen zwei Ansätzen zu unterscheiden. Bei statischen Modellen wird der zu entrichtende Preis für ein Gut einseitig durch den Verkäufer festgelegt, während die interaktive Bepreisung Käufer, Verkäufer oder beide Parteien in die Preisfindung einbezieht [SkSW05].

In den klassischen Massenmedien wird der Wert einer Medienleistung durch die generelle Verbreitung des Mediums, die Menge der Rezipienten des Mediums sowie die Kontaktqualität (Merkmale der Rezipienten) definiert und *statisch* durch den Tausend-Kontakt-Preis (TKP)¹ bepreist [Bogs01]. Diese konstant verfügbaren Qualitätsmerkmale sind bei mobilen Kundenkontakten aufgrund der dynamischen Einflüsse von Datenschutz-, Infrastruktur- und Kostenaspekten auf die Qualität von Identitäts- und Kontextinformationen nicht vorhanden. Eine statische Bepreisung seitens des Verkäufers würde deshalb erfordern, dass für jede Qualität eines Kundenkontakts ein Preis ermittelt wird. Dies ist aber vor dem Hintergrund des entstehenden Planungsproblems völlig ineffizient. Zu dem schafft die individuelle Wertschätzung von kon-

¹ Preis um 1000 Rezipienten über ein Medium mit den beschriebenen Qualitätsmerkmalen zu erreichen

textbezogenen mobilen Kundenkontakten die Voraussetzung für eine Bepreisung analog zur statischen Preisdifferenzierung, d.h. entsprechend der individuellen Zahlungsbereitschaft von Werbetreibenden [SkSp05].

Aufgrund des durch Werbetreibende (Käufer) determinierten Nutzens mobiler Kundenkontakte wird auf die Betrachtung von Preismodellen mit Beteiligung des Verkäufers verzichtet und es verbleiben das *Reverse Pricing* und *Auktionen* [SkSW05]. Beim Reverse Pricing gibt ein Käufer Gebote für ein Produkt an den Verkäufer ab. Liegt dieses Gebot über der Preisschwelle des Verkäufers, so erhält der Käufer das Produkt. Im anderen Fall beginnt eine neue Gebotsrunde oder der Käufer muss vom Erwerb des Produkts absehen [BeSS05]. Da aufgrund der beschriebenen dynamischen Qualitätsausprägungen mobiler Kundenkontakte die Preisschwelle für jeden mobilen Kundenkontakt jeweils neu festgelegt werden muss, entsteht dadurch ein der statischen Preisdifferenzierung vergleichbares Planungsproblem.

Gegen die verbleibende Einsatzmöglichkeit von Auktionen können auf dieser Abstraktionsebene keine Einwände gefunden werden. Jedoch ergibt sich dadurch die Herausforderung, eine optimale Gestaltung für Auktionen zu finden, die neben dem Preis auch den Nutzen der werblichen Kommunikation für Rezipienten im Rahmen der Allokation berücksichtigt.

3.2 Einsatz eines interaktiven Preismodells in Form einer Auktion

Die zentrale Rolle für die Akzeptanz von Mobile Marketing Aktivitäten spielt der Nutzen eines werblichen Angebots für den Rezipienten [Meie02]. Dieser Nutzen setzt sich aus der Übereinstimmung der gegenseitigen Anforderungen von Werbetreibenden und Rezipienten (*Relevanz*), sowie der *Qualität der Rezipientenansprache* hinsichtlich der personalisierten Gestaltung und dem Inhalt der werblichen Nachrichten zusammen [EFGK00]. Die Relevanz ergibt sich dabei aus der Eignung von Werbetreibenden zur Befriedigung des aktuellen Bedürfnisses von mobilen Rezipienten. So hat beispielsweise ein Restaurant, das sich in der direkten Umgebung eines mobilen Rezipienten befindet, eine höhere Eignung als ein solches in weiterer Entfernung. Dieser Wert kann über den Abgleich der Identitäts- und Kontextinformationen mobiler Rezipienten und dem Zielgruppenprofil eines Werbetreibenden ermittelt werden. Die Qualität der Rezipientenansprache hingegen lässt sich nicht direkt bewerten, denn Inhalt und Gestaltung einer werblichen Nachricht lassen sich nicht quantitativ erfassen und objektiv bewerten [EFGK00]. Ein Beispiel für einen indirekten Ansatz liefert jedoch das im stationären Internet agierende Unter-

nehmen *Google*. Dieses nutzt die Click-Rate² für Werbeanzeigen zur Bewertung der Qualität einer werblichen Nachricht [Goog06]. Die Annahme dabei ist, dass aus einer erfolgreich gestalteten, relevanten werblichen Kommunikation auch eine höhere Click-Rate folgt.

Aus der Zusammenführung der zwei dargestellten Faktoren *Nutzen für Rezipienten* und *Zahlungsbereitschaft Werbetreibender* resultiert ein virtuelles Gebot, dessen Höhe über den Zuschlag für einen kontextbezogenen mobilen Kundenkontakt entscheidet (vgl. Abb. 3).

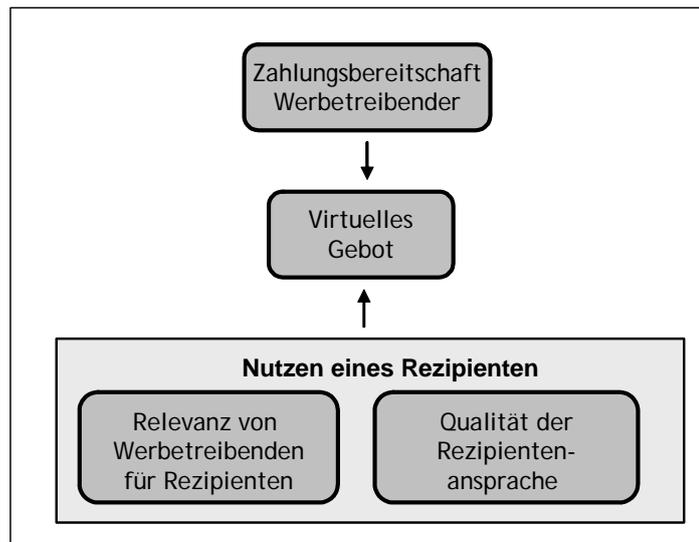


Abb. 3: Einflussfaktoren auf die Allokation kontextbezogener mobiler Kundenkontakte

Fasst man den Nutzen bzw. die Relevanz eines Werbetreibenden für einen Rezipienten als Qualitätsmerkmal auf und beschreibt sie durch eine Nutzenfunktion, so kann das dargestellte Allokationsproblem durch den Einsatz multi-attributiver Auktionen [Bich00] adressiert werden. Da eine konkrete Implementierung dieser Auktionsform von den jeweiligen Zielen des Intermediäres (Verkäufers) abhängig ist, werden in Kapitel 4 ausschließlich Designaspekte dieser Auktionsform erarbeitet. Anhand eines Anwendungsbeispiels soll jedoch der Ablauf der Allokation eines mobilen kontextbezogenen Kundenkontakts kurz illustriert werden:

Ein mobiler Nutzer sucht nach einer Apotheke in seiner näheren Umgebung. Dazu ruft er über seinen PDA einen mobilen Informationsdienst auf und gibt das Suchwort *Apotheke* ein. Daraufhin wird sein aktueller Aufenthaltsort ermittelt und alle in der Datenbank des Informationsdienstes gespeicherten relevanten Apotheken abgerufen. Durch von den Apotheken hinterlegten Zielgruppenprofilen und Werbebudgets, wird automatisch ihr monetäres Gebot für den Kun-

² Das Verhältnis von präsentierten und rezipientenseitig-aktivierten Werbeanzeigen [SkSp00]

denkontakt ermittelt (vgl. Kapitel 4.4). Der Informationsdienst berücksichtigt dann für die Ermittlung des Gewinners der Auktion, neben diesem monetären Gebot, auch die Distanz zwischen Apotheke und mobilen Nutzer. Die höchsten drei monetären Gebote, bewertet mit der jeweiligen Distanz, erhalten dann den Zuschlag und werden auf dem PDA des mobilen Nutzers angezeigt.

4 Designaspekte für multi-attributive Auktionen zur Allokation kontextbezogener mobiler Kundenkontakte

4.1 Multi-attributive Auktionen zur Allokation kontextbezogener mobiler Kundenkontakte

Auktionen werden in [McMc87] als “... market institution with an explicit set of rules determining resource allocation and prices on the basis of bids from the market participants” definiert. *Multi-attributive Auktionen* sind dabei eine spezielle Form von Auktionen, die es ermöglichen, neben dem Preis noch zusätzliche Attribute eines Gutes (z.B. Qualität) in Auktionsgebote einzubeziehen [Che93, Bich00]³. Interpretiert man die Relevanz zwischen mobilen Nutzern und Werbetreibenden als ein solches Attribut, so lässt sich dieses mittels einer multi-attributiven Auktion berücksichtigen. Nachfolgend werden dazu die, für die Allokation von kontextbezogenen mobilen Kundenkontakten, notwendigen Designaspekte vorgestellt. Dabei gilt es allgemeine Anforderungen für diese Auktionen zu diskutieren, Rahmenbedingungen für Nutzenfunktionen zur aggregierten Bewertung der Attribute (Relevanz und Preis) durch den Intermediär darzustellen und einen Ansatz zur automatischen Gebotsbildung für Bieter zu erarbeiten.

4.2 Allgemeine Anforderungen für die Auktion von kontextbezogenen mobilen Kundenkontakten

Die allgemeinen Anforderungen an die Gestaltung einer Auktion für kontextbezogene mobile Kundenkontakte werden anhand der von *Skiera / Spann* identifizierten und in Tabelle 1 dargestellten Designaspekte [SkSp04] diskutiert.

³ Einen weiterführenden Überblick zu Gestaltungsoptionen von Auktionen liefert Klempere [Klemp99]

Designaspekt	Entscheidungen
Auktionsgegenstand	Beschreibung des Auktionsgegenstands, Festlegung der Gebotsvariablen, Versteigerung von Optionen
Auktionsteilnahme	Beschränkte oder offene Auktionsteilnahme, Teilnahmegebühr
Dauer der Auktion	Gesamtdauer der Auktion, Fixes oder variables Auktionsende
Rahmenbedingungen	Mindestpreise, Flexibilität der Auktionsregeln, Verbindlichkeit der Gebote
Verkauf mehrerer Einheiten eines Produkts	Simultane oder sequentielle Versteigerung, einheitliche oder individueller Zuschlagspreis bei mehreren Einheiten identischer Produkte

Tab. 1: Designaspekte für Auktionen auf Basis von [SkSp04]

4.2.1 Auktionsformen

Die Formen von Auktionen lassen sich in englische, holländische, Vickrey- und Höchstpreisauktion aufteilen. Unterschiede bestehen in der Form der Gebotsabgabe und der Bestimmung des Transaktionspreises [SkSp04]. Die Allokation eines mobilen Kundenkontakts muss unmittelbar beim Zugriff auf ein mobiles Nutzenangebot (z.B. mobiles Portal) ohne Verzögerung erfolgen. Auktionsbedingte Wartezeiten, die die Präsentation des Werbemittels verzögern, würden zu einer sinkenden Akzeptanz oder Ablehnung seitens der Rezipienten führen. Als Auktionsformen sind daher nur die Höchstpreis- und Vickrey Auktion zulässig, da sie keine iterativen Gebotsrunden vorsehen [SkSp04].

4.2.2 Auktionsgegenstand

Auktionsgegenstand ist das digitale Informationsgut *kontextbezogener mobiler Kundenkontakt*, für dessen Vervielfältigung zwar keine weiteren Kosten entstehen würden, das aber nur an den *relevantesten* Werbetreibenden veräußert wird und dessen Wert danach verfällt. Die Beschreibung des Auktionsgegenstandes ändert sich dynamisch, hinsichtlich der in Kapitel 2 diskutierten datenschutzrechtlichen, technischen und kostenbezogenen Aspekte bei der Erhebung der Identitäts- und Kontextinformationen. Für jeden mobilen Kundenkontakt wird ein einmaliger Preis erhoben. Eine Versteigerung von Optionen für diesen findet nicht statt, da er nur zum aktuellen Zeitpunkt einen Wert besitzt und nutzbar ist.

4.2.3 *Auktionsteilnahme*

Zur Gewährleistung eines hohen Nutzens der werblichen Kommunikation werden die Zielgruppenprofile von Werbetreibenden mit den Identitäts- und Kontextinformationen von mobilen Rezipienten im Rahmen der Auktion verglichen. Mit einer redaktionellen Prüfung von grundlegenden Angaben Werbetreibender (z.B. Standort eines Point-of-Sale) können falschen Daten reduziert und nur geprüfte Unternehmen zugelassen werden. Da nicht relevante Angebote (z.B. die Empfehlung eines Restaurants an dessen Ruhetag) im mobilen Internet zu einer größeren Belästigung für Rezipienten führen [Meie02], hat eine solche Prüfung einen größeren Stellenwert als im stationären Internet.

4.2.4 *Dauer der Auktion*

Wie bereits in Kapitel 4.2.1 beschrieben, darf die Dauer einer Auktion nur unmerklich die Präsentation eines werblichen Nutzenangebots nach der Anforderung durch Rezipienten verzögern. Ein fixes Ende der Auktion ist daher erforderlich.

4.2.5 *Rahmenbedingungen der Auktion*

Die Generierung eines kontextbezogenen mobilen Kundenkontaktes verursacht je nach erhobenen Identitäts- und Kontextinformationen variable Kosten. Mit der Erhebung von Mindestpreisen besteht die Möglichkeit, die Deckung dieser Kosten im Rahmen der Auktion abzusichern, schafft aber auch die Gefahr, weniger oder keine Käufer zu finden [SkSp04]. Die verursachten Kosten wären in letzterem Fall verloren, da ein mobiler Kundenkontakt nur zum Zeitpunkt der Auktion einen Wert besitzt. Aus diesem Grund erlaubt die Auktion keine Flexibilität hinsichtlich der Auktionsregeln sowie der Verbindlichkeit von Geboten.

4.2.6 *Verkauf von mehreren Einheiten eines Produkts*

Jeder kontextbezogene mobile Kundenkontakt stellt ein Unikat dar, so dass nur die Reihenfolge für den Verkauf von mehreren *unterschiedlichen* Kundenkontakten relevant ist. Da aber für mobile Kundenkontakte von einem nicht-komplementären Verhältnis ausgegangen wird, können sie in diesem Kontext sequentiell versteigert werden.

4.3 Nutzenfunktion zur Abbildung der Relevanz von Werbetreibenden für mobile Rezipienten

Die Ergebnisse des Abgleichs der Identitäts- und Kontextinformationen von mobilen Rezipienten mit dem Zielgruppenprofil von Werbetreibenden determiniert die Nutzenfunktion im Rahmen der Auktion. Dazu kann die Güte des Abgleichs von einzelnen Informationen (z.B. des Personenprofils) zusätzlich jeweils einzeln gewichtet werden. Zu dem ist die Präzision der gesamten Angaben von Werbetreibenden zu ihrem Zielgruppenprofil zu gewichten. Dadurch wird vermieden, dass Werbetreibende durch ein breites, ungenaues Zielgruppenprofil ihre Wahrscheinlichkeit für einen besseren Abgleich erhöhen (vgl. Abb. 4).

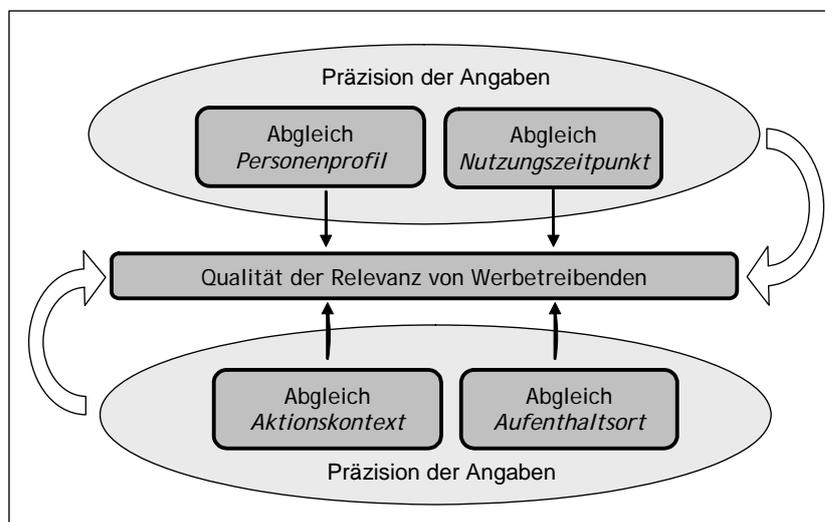


Abb. 4: Qualität der Relevanz von Werbetreibenden für mobile Rezipienten

Für den *Aktionskontext* können sich Werbetreibende beispielsweise einer oder mehreren Portal-kategorien oder Suchbegriffen zuordnen, die mit der Anfrage des mobilen Nutzers verglichen werden. Das *Personenprofil* enthält Informationen, die bei der klassischen Marktsegmentierung und -bearbeitung Anwendung finden und dementsprechend abgeglichen werden [Kotl01]. Der *Nutzungszeitpunkt* und der *Aufenthaltort* eines Rezipienten sind insbesondere bei einem physischen Leistungsangebot (z.B. Restaurant in der näheren Umgebung eines mobilen Nutzers) relevant.

Eine besondere Stellung nimmt jedoch der Abgleich des *Aufenthaltorts* eines mobilen Nutzers mit der Präferenz von Werbetreibenden ein. Nach Huff [Huff64] leitet sich die Bereitschaft eines Konsumenten, den Weg zu einem Point-of-Sale (POS) überwinden, aus der Attraktivität des

POS (z.B. aufgrund der Anzahl angebotener Produkte) und der Distanz zu diesem ab. Dabei nimmt diese Bereitschaft exponentiell mit steigender Distanz zum POS ab. Huff berücksichtigt zu dem in seinem Modell, dass diese Bereitschaft für unterschiedliche Leistungsangebote variieren kann und damit beispielsweise für teure Produkte größer ist als für Billige. Interpretiert man nun das Ergebnis des Abgleichs von *Aktionskontext*, *Personenprofil* und *Nutzungszeitpunkt* als *Attraktivität* eines Werbetreibenden nach Huff und ist der *Aufenthaltort* für die werbliche Ansprache relevant, so kann auch das Huff-Modell zur Bestimmung der gesamten Nutzenfunktion im Rahmen der Auktion genutzt werden (vgl. Abb. 5).

$\frac{\text{Abgleich (Aktionskontext, Personenprofil, Nutzungszeitpunkt)}}{\text{Distanz (Point-of-Sale)}} = \text{Relevanz eines Werbetreibenden}$
--

Abb. 5: Bewertung der Relevanz von Werbetreibenden für Rezipienten auf Basis des Huff-Modells

4.4 Automatische Gebotsbildung für Werbetreibende

Aufgrund der zu erwartenden hohen Anzahl von Transaktionen, der zeitkritischen Dauer sowie der Komplexität, kann die Auktion nur vollautomatisiert durchgeführt werden. Dazu müssen Werbetreibende in der Lage sein, für die Ermittlung eines Gebotes ihre eigene individuelle Nutzenfunktion für mobile Kundenkontakte zu definieren. Beispielsweise könnte der Nutzungszeitpunkt für ein Restaurant mit langen Öffnungszeiten weniger relevant sein, als für eine Bar, deren Öffnungszeiten sich oft auf den Abend konzentrieren. Hierfür kann die gleiche Funktionalität zum Vergleich von Identitäts- und Kontextinformationen und dem Zielgruppenprofil, wie in Kapitel 4.3 beschrieben, genutzt werden. Die automatische Ermittlung der Zahlungsbereitschaft bzw. des Gebots erfolgt dann in drei Schritten. Zunächst werden die Merkmale eines *idealen* Kundenkontakts in die Nutzenfunktion eingesetzt und deren Wert ermittelt, dann wird die Nutzenfunktion für die *tatsächlichen* Merkmale der vorliegenden Kundenkontakte berechnet. Fehlende, aber relevante Merkmale eines Rezipienten wirken sich dabei negativ auf den Nutzen für Werbetreibende und damit auf die Zahlungsbereitschaft aus. Das Verhältnis von *tatsächlichem* zu *idealem* Kundenkontakt multipliziert mit dem für einen Kundenkontakt jeweils zur Verfügung stehenden Werbebudget ergibt das Gebot des Werbetreibenden für den Kundenkontakt (vgl. Abb. 6).

$$\frac{\text{tatsächlicher Kundenkontakt}}{\text{idealer Kundenkontakt}} \times \text{Werbepudget pro Kundenkontakt} = \text{Gebot für Kundenkontakt}$$

Abb. 6: Ermittlung des Gebots für einen Kundenkontakt

5 Zusammenfassung und weitere Forschung

Im vorliegenden Beitrag wurde ein möglicher Ansatz zur Gestaltung der Allokation und Bepreisung von *kontextbezogenen mobilen Kundenkontakten* mittels multi-attributiver Auktionen vorgestellt. Aufgrund der Dynamik und Komplexität sowie die u.a. daraus resultierenden individuellen Wertschätzung dieses Informationsgutes durch Werbetreibende, wurden interaktive Preismodelle vorgeschlagen, um diese Anforderungen adäquat zu adressieren. Dadurch verlagert sich der verkäuferseitige Planungsaufwand im Rahmen der Preisfindung hin zu einem Problem der optimalen Gestaltung des Preismodells. Aufgrund der zentralen Rolle des Nutzens bzw. der Relevanz werblicher Kommunikation und der direkten Implikationen auf die Akzeptanz für mobile Rezipienten, sind multi-attributive Auktionen zur Adressierung dieser Anforderungen ausgewählt worden. Anstatt ein konkretes Auktionsdesign vorzustellen, wurden Rahmenbedingungen für Einsatz und Gestaltung dieser Auktionsform erarbeitet.

Ausgehend von dieser Grundlage, ist für die weitere Forschung geplant, die dargestellten Designaspekte für multi-attributive Auktionen aufzugreifen und gestützt durch Case Studies in ein konkretes Auktionsdesign zu gießen und prototypisch zu implementieren. Daneben soll der Nutzen, der von Sachleistungen (z.B. in Form von Coupons) für Rezipienten ausgeht, in die Nutzenfunktion der Auktion eingebunden werden.

Literaturverzeichnis

- [AIFR05] Albers, Andreas; Figge, Stefan; Radmacher, Mike: LOC³ - Architecture Proposal for Efficient Subscriber Localisation in Mobile Commerce Infrastructures. In: Proceedings of 2nd IEEE International Workshop on Mobile Commerce and Services (WMCS'05), München 2005.

- [BaSt02] Barwise, Patrick; Strong, Colin: Permission-based mobile Advertising. *Journal of Interactive Marketing* 16 (2002) 1, S. 14-24.
- [Berl03] Berlecon: Basisreport Mobile Marketing. http://www.berlecon.de/output/studien.php?we_objectID=115, 2003, Abruf am 2006-07-01.
- [BeSS05] Bernhardt, Martin; Spann, Martin; Skiera, Bernd: Reverse Pricing. In: *Die Betriebswirtschaft (DBW)* 65 (2005), S. 104-107.
- [Bich00] Bichler, Martin: An experimental analysis of multi-attribute auctions. In: *Decision Support Systems* 29 (2000), S. 249-268.
- [Bogs01] Bogs, Nicolas: Werbeagenturen im Mediamarkt und allokativer Effizienz. Peter Lang, Frankfurt a. M. 2001.
- [BUNA06] Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen: Jahresbericht 2005. <http://www.bundesnetzagentur.de/media/archive/5278.pdf>, 2006, Abruf am 2006-07-01.
- [Che93] Che, Yeon-Koo: Design competition through multidimensional auctions. In: *Journal of Economics* 24 (1993) 4, S. 668-680.
- [DBHS06] Decker, Michael; Bulander, Rebecca; Högler, Tamara; Schiefer, Gunther: m-Advertising: Werbung mit mobilen Endgeräten – Ein Überblick. In: *Proceedings der 1. Fachtagung "Mobilität und Mobile Informationssysteme"*, Passau 2006.
- [EFGK00] Engelhardt, Bernhard; Frick, Hans-Adolf; Gemeinhardt, Roland; Kreuzer Christine: *Werbung in Theorie und Praxis*. M&S Verlag, Waiblingen 2000.
- [Figg06] Figge, Stefan: Entwurfsmuster eines werbefinanzierten Geschäftsmodells für dem Mobile Commerce auf Basis eines semantischen Identitätskonzepts. Dissertation, Universität Frankfurt a. M., Frankfurt a. M. 2006.
- [FiTh06] Figge, Stefan; Theysohn, Sven: Quantifizierung IKS-basierter Marktleistungen. In: *WIRTSCHAFTSINFORMATIK* 48 (2006) 2, S. 96-106.

- [Goog06] Google Inc.: Adwords. <https://adwords.google.de/select/comparison.html>, 2006, Abruf am 2006-07-20.
- [Huff64] Huff, David: Definition and Estimating a Trading Area. In: Journal of Marketing. 29 (1964) 3, S. 34-38.
- [Klem99] Klemperer, Paul: Auction Theory: A Guide to the Literature. In: Journal of Economic Surveys 13 (1999) 3, S. 227-286.
- [Kotl01] Kotler, Philip: Marketing Management: Analyse, Planung und Verwirklichung. Schäffer-Poeschel, Stuttgart 2001.
- [LeSK06] Leppäniemi, Matti; Sinisalo, Jaakko; Karjaluoto, Heikki: Mobile Marketing Research 2000 -2005: Emergence, Current Status, Future Directions. In: Proceedings of the CMC 2006 11th Conference on Corporate and Marketing Communications, Ljubljana, Slowenien 2006.
- [Lipp04] Lippert, Ingo: Mobile Marketing. In: Gerstner, R.; Hunke, G.; Sabel, H. (Hrsg.): Innovatives Marketing. Fritz Knapp, Frankfurt am Main 2004, S. 109-123.
- [McMc87] McAfee, Preston; McMillan John: Auctions and Bidding. In: Journal of Economic Literature 25 (1987), S. 699-738
- [Meie02] Meier, Roland: Generierung von Kundenwert durch mobile Dienste. Dissertation, Technische Universität München 2002.
- [Mind06] Mindmatics AG: Mr. AdGood. <http://www.misteradgood.com/de/public/registry.form.go?to=2&level=0>, 2006, Abruf am 2006-07-01.
- [NSRH02] Ng-Kruelle, Grace; Swatman, Paul; Rebne, Douglas; Hampe, Felix: The Price of Convenience: Privacy and Mobile Commerce. In: 3rd World Congress on the Management of Electronic Commerce. Hamilton, Ontario, Canada 2002.
- [PeRo97] Peppers, Don; Rogers, Martha: The one to one future. Doubleday, New York, USA 1997.

- [Rösc06] Rösch, Bert: Auf dem Weg zum Mobile Marketing 2.0. In: ONEtoONE (2006) 7, S. 8-13.
- [SaTä05] Salo, Jari; Tähtinen, Jaana: Retailer Use of Permission-Based Mobile Advertising. In: Advances in Electronic Marketing. Idea Publishing, USA, U.K 2005, S. 139-155.
- [SkSW05] Skiera, Bernd; Spann, Martin; Walz Uwe: Erlösquellen und Preismodelle für den Business-to-Consumer-Bereich im Internet. In: WIRTSCHAFTSINFORMATIK 47 (2005) 4, S. 285-293.
- [SkSp00] Skiera, Bernd; Spann, Martin: Preisdifferenzierung im Internet. In: Controlling 12 (2000), S. 417-423.
- [SkSp04] Skiera, Bernd; Spann, Martin: Gestaltung von Auktionen. In: Backhaus, K.; Voeth, M. (Hrsg.): Handbuch Industriegütermarketing, Gabler, Wiesbaden 2004, S. 1039-1056.
- [SkSp05] Skiera, Bernd; Spann, Martin: Preisdifferenzierung im Internet. In: Schögel, Tomczak, Belz (Hrsg.): Ro@dmapp to E-Business. Thexis, St. Gallen 2002, S. 270-284.
- [Trip03] Tripathi, Arvind: Decision Models in Wireless Advertising. Dissertation, Universität von Connecticut, Connecticut 2003.
- [Wohl01] Wohlfahrt, Jens: Wireless Advertising. In: Silberer, G.; Wohlfahrt, J.; Wilhelm, T. (Hrsg.): Mobile Commerce. Gabler, Wiesbaden, 2001.
- [ZAW05] Zentralverband der deutschen Werbewirtschaft: Werbung in Deutschland 2005. edition ZAW, Berlin 2005.
- [Zäng00] Zängler, Thomas: Mikroanalyse des Mobilitätsverhaltens in Alltag und Freizeit. Springer, Berlin 2000.

Adoption and Impact of Mobile-Integrated Business Processes – Comparison of Existing Frameworks and Analysis of their Generalization Potential

Key Pousttchi, Bettina Thurnher

Mobile Commerce Working Group
Chair of Business Informatics and Systems Engineering (WI-SE)
University of Augsburg
Universitaetsstrasse 16, 86135 Augsburg, Germany
{key.pousttchi, bettina.thurnher}@wi-mobile.de

Abstract

The integration of mobile workplaces in the (electronically mapped) intra-enterprise value chain is a major and still increasing corporate IT issue. Although the usage of mobile technologies for this purpose is far behind expectations and numerous failures can be observed, still little work has been done on theory building in this area.

In this contribution we identify and compare existing frameworks for adoption and impact of mobile technology to support mobile business processes. The hypotheses underlying these frameworks are challenged with experiences from three long-term case studies which are diverse in industry, company size and other factors in order to scrutinize their potential for generalization. The outcome is a set of hypotheses that show robustness against variation of major parameters and thus may be suitable to serve as a basis for a generalized and unified framework on mobile-integrated business processes.

1 Introduction

Usage and relevance of information technology (IT) in enterprises have changed fundamentally in recent decades. Instead of isolated IT usage limited to an operational level (e.g. production planning or machine-tool steering), vast electronification and networking in many enterprises have led to a strategic and area-wide IT usage through the electronic mapping of the intra-enterprise value chain [TuPo04]. In larger companies this was typically achieved through

the introduction of complex ERP systems while in smaller companies rather through the use of smaller industry-specific software packages. Resulting potentials include for instance common data management, integrated workflow management or contemporary availability of aggregated, edited corporate data (business intelligence). However, a major problem occurs as soon as single business processes (or even only parts of these) are executed mobile and thus can no longer be fully supported by the use of stationary IT. As a result, the integration of mobile workplaces is a major and still increasing corporate IT issue. While mobile information and communication technologies are available to fill in the gap, their actual usage for this purpose is far behind expectations. What is more, among the actually activated projects numerous failures can be observed (e.g. [Cogn05; Aber04]). Despite the high relevance of the topic theory building is not advanced at all. Our extensive research in high-ranked relevant literature¹ showed only little results (see section 2). Two frameworks were proposed: (a) [WaCh04] is deductively suggesting an *adoption framework* and examining it with an industry-specific survey (for a detailed description see section 2.1) and (b) [GeSh04] is deductively suggesting an *impact assessment framework* and examining it with one specific case study (for a detailed description see section 2.2).

The research objective of this paper is to investigate these two frameworks and scrutinize their potential for generalization. In order to do this, the underlying hypotheses are one by one challenged with the experiences from three in-depth case studies on mobile-integrated business processes which are diverse in industry, company size and other factors. The outcome is a set of hypotheses that show robustness against variation of major parameters and thus may be suitable to serve as a basis for a generalized and unified framework on mobile-integrated business processes. In the context of this paper, the term *mobile business processes* (MBP) refers to any business process which is (partly or completely) executed mobile and thus can not be fully supported by the use of stationary IT [GuPo05]. The term *mobile-integrated business processes* (MIBP) refers to any MBP that is fully supported by mobile IT [PoTh05].

The contribution is organized as follows: In section 2 the *adoption framework* from [WaCh04] and the *impact assessment framework* from [GeSh04] are introduced. In section 3.1 the

¹ Literature research comprised the following resources: MIS Quarterly (MISQ), Journal of the AIS (JAIS), International Journal of Electronic Commerce (IJEC), Communications of the ACM (CACM), Information Systems and e-Business Management (ISeB), International Journal of Mobile Communication (IJMC), Wirtschaftsinformatik (WI), Zeitschrift für Betriebswirtschaft (ZfB), Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung (zbf), Schmalenbach Business Review (sbr), Die Betriebswirtschaft (DBW), and the proceedings of "Internationale Tagung Wirtschaftsinformatik".

method and the used case studies are explained. In section 3.2, the two frameworks are compared and in sections 3.3 and 3.4 the respective hypotheses of the two frameworks are examined. In section 4 the conclusion is drawn.

2 Existing Research

2.1 The Adoption Framework

The adoption framework of [WaCh04] is derived from previous research in the field of strategic choice theory and upper-echelon theory. Upper-echelon theory emphasizes the role of the firm's management in actively shaping organizational strategies [Chil72]. These research lines provide a theoretical foundation for understanding organizational decisions about innovation adoption. [WaCh04] identified critical factors that influence e- and mobile (m) business adoption by travel agencies and suggested a three level framework that states that differences between the technology adoption behaviours of firms can be fully captured only if the influential factors at multiple levels (environmental, organizational, and managerial) are taken into consideration [ToKl82; Roge95]. Moreover a firm's strategic orientation and the personal characteristics of its top managers influence its approach to innovation adoption. The framework was tested according to 7 hypotheses in a case study. In the study 137 completed questionnaires of the Taipei Association of Travel Agents were returned. Additionally a second study was depicted in this paper where they investigated to what extent the idea of mobile e-business is understood and accepted by travel agency CEO in Taiwan.

The findings show that no firm had a clear plan for mobile e-business. However, the CEO did point out that mobile advertising through the Short Messaging Service (SMS) would be the first mobile application despite the technological limitations of currently available mobile devices. Although the CEO agreed that the expansion to mobile e-business is only "a matter of time," [WaCh04, p. 56].

The suggested framework of [WaCh04] consists of the following elements:

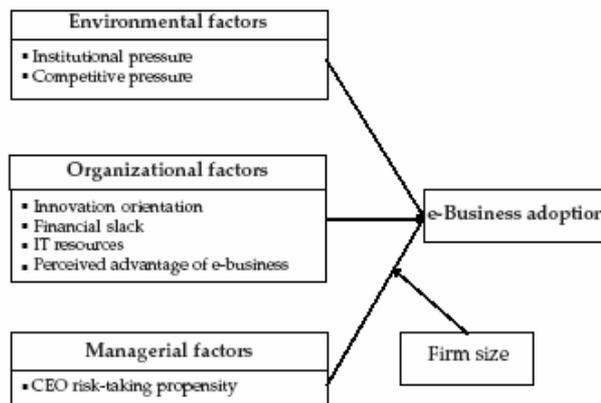


Fig. 1 Framework of e- and m-Business Adoption for Service Firms

The analysis of the data has been executed via hypothesis testing with statistical methods e.g. R-square and mean centering method proposed by Cronbach. Seven independent and two dependent variables (DV) have been defined: DV1 measures the firm's intention to develop e- and m-business applications and DV2 measures the degree of e- and m-business implementation.

2.2 The Impact-Assessment Framework

The impact-assessment framework of [GeSh04] is derived from the task/technology fit model [GoTh95] and the technology acceptance model [Davi89]. Moreover research on technology innovation [ToKl82] has been used by [GeSh04] to point out the importance of matching information systems with the organizational tasks to be supported or automated, as a precursor to system use and subsequent benefits. Recent research of Goodhue and Thompson's work has been extended and integrated with the technology acceptance model [DiSt98; DiSt99; MaKe98]. Other researchers have applied Zigurs and Buckland's framework to gain further insights on the success factors of group work [HoMO93]. The research framework combines the idea of task/technology fit with the general notion of impacts of information technology on organizations. The framework consists of four factors: technology, task, usage and impact to which a mobile business application is investigated. The parameters of each factor are depicted in **Fig. 2**. [GeSh04] validated the framework according to 9 propositions in a case study on a mobile e-procurement platform at a Fortune 100 company. The authors interacted with members of the project management group on a regular basis during the period from August 2001 to March 2002 to obtain information about the project, including objectives, progress, and organizational and technical details. Responses were collected from a total of 17 end-users and

used WAP-enabled cell phones. The study was conducted in Asia, USA and Europe. The suggested framework of [GeSh04] consists of the following elements:

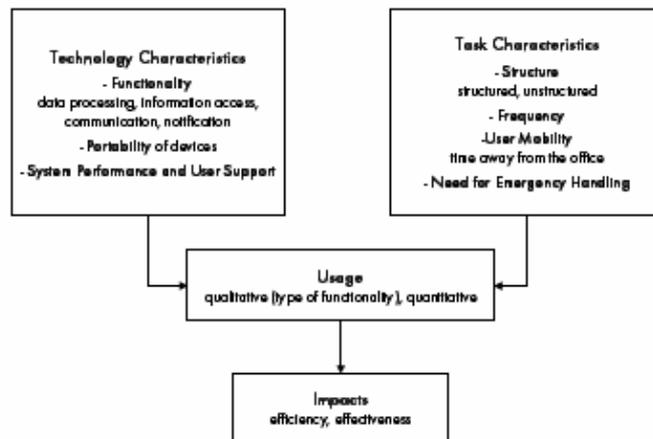


Fig. 2 The Impact-Assessment Framework [GeSh04]

The analysis of the data has been executed mainly via qualitative results from the interviews and quantitative measures like the amount of mobile tool usage for specific tasks. Moreover [GeSh04] analysed the correlations of the variables of their propositions.

2.3 Other Related Work

More recent research shows that mobile technologies and research in mobile applications are of increasing importance in various fields of applications e.g. [ScPL06; FoGK05]. [ChNa04] investigated the value of mobile business applications in domains like fast food restaurants, facility management, maintenance and housekeeping, insurance companies, support of sales force, logistics, mobile-marketing campaigns, real time weather information for the army and even the application of mobile technology for the efficient handling of formula one races. Obviously there are differences in the judgment about the impact of mobile-business applications in different branches. Whereas [WaCh04] found out that there was no need for mobile applications in 2004 for Taiwanese travel agencies [ChNa04] found out that mobile applications are widely applied in various branches. The amount of increasing research [ScPL06; WaCh04; ChNa04; GeSh04] and the improved mobile technology (device, connectivity, performance) indicate that mobile business applications are gaining importance. Therefore we want to verify in this paper if the hypothesis and the framework suggested by [WaCh04] are applicable to mobile business applications.

3 Comparison and Validation of existing frameworks

3.1 Method

Our research approach is based upon theory building from case studies [Eise89]. This approach is especially appropriate for obtaining complex details and novel understandings about a specific phenomenon under investigation. Case study research is appropriate for examining practice-based problems, since it allows a researcher to capture the knowledge of practitioners and use it to generate theory [BeGM87; Cres03; Eise89].

The overall aim of our research is to contribute to theory building in the area of MIBP. We consider case study research as a suitable approach for the investigation of MIBP impact and adoption factors as it provides the opportunity for longitudinal research and in-depth investigation especially of socio-technical interdependencies. Both [WaCh04] and [GeSh04] suggest further research should test the framework applicability in different industry domains and different organizations. Thus, we selected case studies to be most diverse in order to validate the limits of the frameworks. As a comparison of the case studies themselves is not intended negative effects of this diversity do not occur. As a common characteristic, the examined MBP in all three case studies is technical customer service provision. Most important differences are industry and organization size. All three case studies have been longitudinal studies and lasted from 6 to 15 months. Within the studies, methods like expert and user interviews as well as questionnaires have been applied to collect data. Interviews have been transcribed and fed back to the interviewees in order to reduce possible errors. For data gathering CEO, CTO, project managers and end users of the mobile application have been questioned with semi-structured questionnaires in personal or via-phone interviews. In the large company four persons have been interviewed, in the medium-sized company two and in the small one again four. Moreover, in the small IT service company we conducted an in-depth usability and tool acceptance study with 30 users [PoTh06]. Table 1 provides an overview of the examined case studies.

Case Study 1: Telecommunication Service Provider

For the first case study we investigated a mobile dispatching and order handling system of a major European Telecommunication Service Provider. 12,000 technicians have been equipped with standard commercial PDA. The introduction of MIBP was conducted in iterative steps from the year 2000 until now. Legacy systems have been substituted by mobile client-server architecture (although service order generation and handling are still conducted on separate information systems for private and business customers). Main tasks of the field technicians

include installation, maintenance and repair of telecommunication products. Service orders are dispatched automatically considering skill level, location, optimal routing, urgency and the respective Service Level Agreements (SLA). Order information is then transmitted via GPRS on the PDA. 50% of the orders are dispatched prior to the working day, the rest during the day.

Case Study 2: Municipal Utility Company

For the second case study we investigated a major German Municipal Utility Company and its mobile service, maintenance and invoicing process. About 1,000 technicians have been equipped with standard industry PDA. MIBP were introduced a one year project with the support of a specialized consulting company. Typical tasks of the utility technicians are e.g. installation and maintenance of meters for gas or electricity in private, public or industry buildings. Due to special restrictions and the need of managerial control, dispatching stayed manually: Technicians come to the office, receive their schedule and then head off to customers. During the day they are now able to receive new jobs, accept or deny them and send status reports.

Case Study 3: IT Service Provider

For the third case study we investigated an Austrian IT Service Provider for hard- and software installation and maintenance. 40 IT service technicians have been equipped with mobile devices. The system was implemented with an Application Service Provider (ASP), implementation took 6 month. Field technicians use a browser-based solution on the mobile device in order to receive their jobs for the day which can be rescheduled via an implemented push-mail function. Main tasks of an IT service technician include repair, maintenance and installation of different hard- and software. A major driver for MIBP introduction was to shorten issuing the invoice, for which especially the record of working/driving time and used spare parts is essential.

Company	Service	Business	Company Size	Architecture	Implementation process	Dispatching
Telecommunication Service Provider	Technical customer service	B2B and B2C	Large	Client/server system; separate IS B2B/B2C	Backend and client	Automatically
Municipal Utility Company	Technical customer service	B2B and B2C	Medium	Client/Server	Backend and Client	Manually
IT-Service Provider	Technical customer service	B2B	Small	ASP solution, browser-enabled client	Server	Manually

Tab. 1 Case Study Overview

3.2 Comparison

This section will provide a comparison of the two frameworks introduced in section 2. The following table opposites the most relevant characteristics:

E-Business Adoption Framework	Impact-Assessment Framework
Measures the adoption behaviour of new technologies according to the three factors: environment, organization and managerial	Suggest Propositions which are tested according to validate the framework
Aims at identifying critical factors that influence e-business adoption in travel agencies	Aims at identifying factors that influence successful m-business application usage
Approach: Case Study Research	Approach: Case Study Research
Hypothesis testing	Qualitative Research which quantitative implications
Limitations	Limitations
Further empirical studies in other application domains should be executed to validate the generalizability of the framework	Further empirical studies in other application domains should be executed to validate the generalizability of the framework
No evolution of the framework has been done considering mobile applications	Respondents have been partially self selected

Tab. 2 Comparison of the e-business adoption- and impact-assessment framework

[WaCh04] applied quantitative validation of hypothesis with 137 completed questionnaires from different companies whereas [GeSh04] applied a more qualitative approach using interviews among 17 partially self-selected respondents within one case study.

3.3 Validation of the adoption framework

In this section we will first introduce the hypotheses and the findings of [WaCh04] and then state differences and common findings in comparison with the three case studies.

The 7 hypothesis of [WaCh04] are:

Hypothesis 1: Institutional pressure is positively related to a firm's adoption of e-business.

Institutional pressure can come from governmental bodies, trade associations, consulting companies, business media and other stakeholders. [WaCh04] found no significant relationship between institutional pressure and either intention to develop e-business or degree of e-business implementation. Consequently, H1 is not supported.

Case Study findings: In our three case studies we could not find any concrete evidence for institutional pressure. As the technical service market is not regulated by governmental bodies but under private company management institutional impacts play a minor role. However for practical implications we found some anecdotal evidence that increased institutional pressure would increase DV1 and DV2 especially if pressure comes from financial stakeholders.

Hypothesis 2: Competitive pressure is positively related to a firm's adoption of e-business.

Competitive pressure has a significant positive relationship with e-business implementation. It has only a marginally significant relationship ($p < 0.10$) with the intention to develop e-business. Arguably, H2 is supported.

Case Study findings: In our case studies DV1 and DV2 have been supported. Due to very short Service Level Agreement (SLA)-fulfilment times in the technical service branch in general efficient and effective process organisation for the dispatching of technicians, spare parts and invoicing are essential for longitudinal market viability. We could further investigate that over the period of investigation (in the longest case 15 months) more and more technical service companies adopted mobile technology and in the same time the amount of mobile solution providers increased. Mobile technology is a critical success factor for long-term market viability; especially for small companies technology leadership is vital.

This applies to all three case studies and company size did not have an influence on these findings.

Hypothesis 3: A firm with a high level of innovation orientation is more likely to adopt e-business than a firm with a low level of innovation orientation.

Innovation orientation is positively associated with intention (as proposed in H3), but not with degree of implementation. Consequently one can conclude that innovation orientation has little to do with a firm's "real" involvement in e-business.

Case Study findings: The case studies revealed that the DV1 can be supported and DV2 can only be supported partially in small and medium-sized (SME) companies. This is due to the fact that smaller companies often lack the resources for innovation (e.g. no specific R&D department) and the investment power is reduced. In large organizations we could investigate the same phenomena as [WaCh04]. Due to the relatively slow decision processes in large companies it is difficult to implement innovative approaches like e- and mobile business even though the intention to adopt e-business might be very high. Especially in large organizations the role of the management and the amount of management commitment and sometimes pressure to implement mobile and e-business technology plays an important role. If management commitment or pressure is given DV1 and DV2 can be supported.

Hypothesis 4: The perceived advantage of e-business is positively related to a firm's adoption of e-business.

This hypothesis is partially supported, a significant positive relationship was found between perceived advantage of e-business and intention to adopt e-business.

Case Study findings: In our case study we could observe that in the SME companies the degree of e-business implementation is higher than in large organizations. The small and medium sized companies implemented e- and mobile business applications once the intention to adopt e-business was given. In the large technical service providing company more hesitation to implement e- and mobile business applications could be observed.

Hypothesis 5: Financial slack is positively related to a firm's adoption of e-business.

In the present study DV1 was supported while DV2 was not. A significant positive relationship was found between perceived advantage of e-business and intention to adopt e-business. Financial slack is positively associated with the degree of e-business implementation. However, the proposed positive relationship between financial slack and intention to adopt e-business is not supported. Instead, there is a negative correlation between the two variables. The more financial slack, the more implementation can be done.

Case Study findings: In our case study mobile application projects have been tackled when sufficient financial resources have been available. Once the resources could be provided DV1 and DV 2 had a positive correlation and the intention to develop e- and mobile applications lead to higher degree of application implementation.

Hypothesis 6: IT resources have a positive effect on a firm's adoption of e-business.

The last organizational factor, IT resources, has positive relationships with both DVs and H6 is supported.

Case Study findings: This could also be observed in the case studies and is mainly due to the presence of IT knowledge and already (partially) existing infrastructures (e.g. Servers). Once IT knowledge about the implementation of an e- or mobile business application was present the reluctance towards mobile tool implementation decreased as the CEO and CTO have been more confident that the investment will pay off rather quickly.

Hypothesis 7: CEO risk-taking propensity has a positive effect on a firm's e-business adoption. The effect is moderated by firm size; it is stronger in smaller firms than in larger firms.

The firm size moderates the relationship between CEO risk-taking propensity and each DV. The relationship is positive when firm size is small and negative when firm size is large.

Case Study findings: These findings are align with the results of our case studies. We identified that the more direct a CEO can decide upon an e- or mobile commerce strategy the more likely it was that those strategies have been adopted. Moreover for SME the competitive advantage gained through faster service processes is vital to keep up with competitors and large organizations.

3.4 Validation of the impact-assessment framework

In this section we will first introduce the propositions and the findings of [GeSh04] and than state differences and common findings in comparison with the three case studies.

The nine propositions of [GeSh04] are:

Proposition 1: Users will first use mobile business applications for notification and communication purposes rather than for data processing and information access.

For this proposition limited support could be found in the Case Study. The users showed willingness to use mobile devices (cell phones) for innovative purposes which means the users

applied the mobile device for tasks they never did before e.g. the notification about waiting approval requests.

Case Study findings: In the case studies we could support this findings and identified that typically mobile devices like mobile phones or PDA are first used for communication purposes and then for e.g. to access the company ERP systems or CRM systems, but this depends on the management directives within a company. If the managerial pressure is rather high and mobile tool usage is fostered strongly we observed that in SME a former paper-based process can be replaced through mobile tool usage in a short period of time. From a business process reengineering perspective this is a bomb-dropping strategy [HaCh93] and can cause employee reluctance as the process change might be too fast for stepwise adoption.

Proposition 2: There is a trade-off between portability and usability of mobile devices, effectively limiting the usage of mobile business applications to simple activities.

This proposition could be supported by the Case Study findings. While users acknowledged the portability of the device, its usability drew criticism. Over two-thirds of the respondents indicated that the small screens and limited keyboards were significant or very significant limiting factors. [GeSh04] investigated that the usability judgement depends highly on the used device. “While a number of users commented that screens and keyboards were too small, one of them experienced a different situation. This participant, who had access to a phone with a larger colour screen and full keyboard, was in fact positive about the usability of the application” [GeSh04 p. 30]. Users preferred to use the tool for simple tasks, such as notification and approval requests.

Case Study findings: The findings in the three case studies support proposition 2. Tool usability is an enabler for innovative applications and especially for mobile applications as they are used while the user is in a mobile condition (walking, standing, and driving). Therefore usability requirements are even more sophisticated than for standard desktop applications [PoTh06].

Proposition 3: System performance and user support have a positive impact on the usage of mobile business applications.

This proposition could be supported. System performance and user support had a positive impact on the usage of the mobile business application, but in the present case poor performance was found to have a negative impact on usage. Some user reported that the downloading time is longer on the mobile device than on the desktop-based application which made it hard to justify the effort.

Case Study findings: System performance factors are highly connected to usability perceptions of a system [PoTh06]. When new mobile applications are introduced to the end users in a prototype state, where performance is usually slower due to the test setting, tool reluctance can increase. Therefore it is important to state clearly that

the system is in a preliminary status and performance rates are not the same as in the real-life running system. In our case studies usability and performance have been identified as a major facilitator for mobile tool acceptance.

Proposition 4a: Employees performing highly structured tasks tend to use mobile business applications for data processing.

The approval tasks of the finance and accounting approvers were much more structured than the tasks of the approving managers. Compared to approving managers, finance and accounting approvers indicated a higher percentage of approvals that could be submitted via the mobile application, providing evidence for proposition 4a.

Case Study findings: In the application domain of technical service support proposition 4a holds as well. The tasks of service technicians are highly structured and therefore mobile tool support can easily be implemented. The typical task set of a service technician consist of order acceptance/denial, documenting the repair,- driving time as well as the used spare parts and transmit the data to the head office. In case of requests (e.g. further details to the order or unclear information) the technician calls the dispatcher or back office. Well structured tasks are easy to be executed via a mobile tool. As the modality functions are limited on a mobile device the workflow should be supported efficiently with a mobile tool.

Proposition 4b: Employees performing unstructured tasks tend to use mobile business applications to access information and for communication purposes.

In the study of [GeSh04] approving managers indicated a higher percentage of approvals where the mobile application could support communication and access to information this provides evidence for proposition 4b.

Case Study findings: The case study findings support this proposition as well. In our study sells representatives especially in the small company used the mobile applications to access current prices and information about the customer e.g. prior communication. The tasks of sells personnel are far less structured than the tasks of technical support staff. The necessary information depends on the actual information and can hardly be planned in advance. Consequently the case study findings support the findings of [GeSh04] in proposition 4.

Proposition 5: The frequency with which a task is performed has a positive impact on the usage of mobile business applications.

In the study of [GeSh04] finance and accounting approvers generally indicated a higher number and dollar volume of approvals than managing approvers. This difference in volume reflected the fact that finance and accounting approvers basically performed approvals on a full-time basis, whereas managers had a broader set of responsibilities. In line with Proposition 5, one would expect finance and accounting approvers to be more optimistic about the mobile e-procurement application than approving managers. For the data set, however, the correlation between repetitiveness and possible usage was negative (-0.4), indicating that higher volume

approvers found the mobile application less useful than approvers who had lower levels of repetitiveness. Evidence gained from user comments indicated that a combination of frequency of task performance and mobility can have greater explanatory power of usage than each factor alone. However it is unclear if this proposition can be dismissed but it is likely that the judgment of the users has been influenced by factors like general enthusiasm for the new technology and difficulties in establishing system access.

Case Study findings: In our case studies we found clear evidence for proposition 5 in the SME companies in the large organisation tool usage was hindered in the beginning due to a high mobile tool reluctance of the technicians. In the large organization tool acceptance and workflow change needed far more time than in the SME. One might argue that this is only due to the size of an organization but we could also investigate that the degree of employee involvement in the design and implementation of a mobile application, especially the workflow design have highly positive influence of tool acceptance and the frequency of tool usage. Another factor that could be observed as relevant was that the system was fully implemented and not in a prototype state when the end user started to use it for operative tasks. Therefore performance and access issues to the system did not arise.

Proposition 6: Employees who are more mobile tend to use mobile business applications more often to perform their tasks than employees who are less mobile.

As mentioned above some of the 17 users stated that a combination of frequency and mobility can have greater explanatory power of usage than each factor alone. Unfortunately [GeSh04] do not indicate how many users made such statements. However there was a positive correlation (0.3) between mobility (time out of the office) and intended usage of the mobile e-procurement application for approval processes.

Case Study findings: In our case studies this could be verified and especially the CEO and CTO of all three case study companies used their PDA or mobile phones extensively when on business trips.

Proposition 7: The use of mobile business applications is positively related to the perceived need to handle emergency situations.

For this proposition a variation among users regarding the question of what constituted an emergency situation could be revealed. However only some anecdotal evidence could be found to support this proposition.

Case Study findings: In our case studies we could find clear evidence for this proposition in every company. The domain of service technicians is highly structured according to SLA priorities and sometimes urgent incident handling e.g. a virus attack. For efficient reaction and fast incident handling mobile tools offer the necessary flexibility to increase the SLA fulfilment rate and consequently increase customer satisfaction and the positive perception of the technology leading companies in the market.

Proposition 8: The use of mobile business applications increases employee productivity and operational efficiency.

This proposition could be supported but [GeSh04] suggest that future research should consider both direct effects on the user and indirect effects on employees interacting with the user.

Case Study findings: From a process management view we followed a holistic approach and investigated the whole process from the order generation to the invoice delivery to the customer [TuPo04]. Within this process the service technician is not integrated into the process chain while doing repair and maintenance jobs at the customer's site. Mobile technology provides the possibility that a service technician stays integrated in the business process e.g. company order generating system, ERP-Systems while being physically mobile for task execution. Consequently we could investigate that large gains in efficiency and effectiveness could be gained in the invoicing and dispatching processes. Invoicing used to be cumbersome as the necessary data (working-, driving time and amount of used spare parts) used to take one working week (5 days) to be transferred from the paper-based forms to the desktop application and to be available for further processing. Due to reduced redundant data entry the error prone process of work protocol transcription could be optimized e.g. about 10 min per day per technician in the small company. Additionally the dispatching (allocation of jobs to a specific service technician) was inefficient as the dispatcher had to make three calls on average to reach a free technician as the status e.g. repair progress and location was only traceable via phone calls. After the mobile tool integration dispatching could be done far more efficient in 85% of all calls via the mobile tool without phone calls and the whole invoicing cycle could be reduced from 10 days down to 4 working days on average.

Proposition 9: The use of mobile business applications improves organizational flexibility and the ability to handle emergency situations.

[GeSh04] found out that no clear understanding of the value of increased flexibility and the ability to handle emergency situations existed among the users. Some anecdotal evidence could be found that users appreciated the fact that the application allowed to keep in touch with co-workers while out of the office.

Case Study findings: The case studies revealed the same as for proposition 7 and proposition 9 can therefore be supported.

Proposition	Evidence form case study
Proposition 1: use for notification and communication purposes	Limited
Proposition 2: trade-off between portability and usability	Yes
Proposition 3: impact of performance and user support on device usage	Yes (strong evidence)
Proposition 4a: highly structured task increase foster mobile tool usage for data processing	Yes
Proposition 4b: unstructured tasks foster mobile tool usage for information and communication purposes	Yes
Proposition 5: task frequency has an impact on mobile tool usage	Unclear
Proposition 6: mobile employees tend to use mobile tools more often	Yes
Proposition 7: mobile tool usage is positively related to the need of handling emergency situations	Sometimes (anecdotal evidence)
Proposition 8: mobile tool usage increases employee productivity and efficiency	Yes
Proposition 9: mobile tool usage increases organizational flexibility and the ability to handle emergency situations	Sometimes (anecdotal evidence)

Tab. 3 Summary of the findings for [GeSh04]

3.5 Suitability for generalization

In general, generalization potential for both frameworks is extremely good. Not only has to be stated that a vast majority of the hypotheses could be validated with the three diverse case studies, but even case study research revealed that some propositions (like e.g. [GeSh04] proposition 7 and 9) could be supported whereas the original study with its limited validation effort only showed anecdotal evidence.

Case Study research is less rigour than an experiment design with hypothesis testing and statistical analysis. However, correctly applied it is able to provide convincing evidence and valid implications with regard to practice, especially if long-term in-depth studies are possible. The detailed results of the validation are shown in Tab. 4.

Moreover the frameworks can be used in industry context for the analysis of business processes to evaluate their mobilization potential and therefore serve as decision support for management.

Frame- work/Findings [WaCh04]	Case Study Findings [WaCh04]	Framework/Findings [GeSh04]	Case Study Find- ings [GeSh04]
H1: no evidence	anecdotal evi- dence	P1: limited evidence	Support
H2: DV2	DV1 + DV2	P2: support	Support
H3: DV1	DV1	P3: strong evidence	Strong evidence
H4: DV1	DV1 + DV2	P4a: support	Support
H5: no evidence	DV1	P4b: support	Support
H6: DV1 + DV2	DV1 + DV2	P5: unclear	Clear evidence in SME; partially in large organization
H7: DV1 + DV2	DV1 + DV2	P6: support	Support
		P7: anecdotal evidence	Support
		P8: support	Support (strong evidence)
		P9: anecdotal evidence	Support

Tab. 4 Comparison of [WaCh04], [GeSh04] with the Case Study Findings

Beyond the validation of the hypotheses 1 to 7 and the propositions 1 to 9 the difficult and critical issues within the field of mobile technology are to measure the degree and speed of technology adoption within the mobile arena. Again in depth, long-term case study research can provide further insights and deeper understanding of the socio-technological interdependencies. Additional research in this area is necessary and we would like to encourage researchers in the field of mobile-integrated business processes to conduct further joint studies.

4 Conclusion

The research objective of this paper was to investigate the *adoption framework* from [WaCh04] and the *impact assessment framework* from [GeSh04] and scrutinize their potential for generalization. In order to do this, we challenged the underlying hypotheses one by one with the experiences from three in-depth case studies which are diverse in industry, company size and other factors.

The outcome is a set of hypotheses (shown in table 4) that shows robustness against variation of major parameters.

Above the level of single hypotheses, both frameworks showed an unexpected high potential for generalization in the area of technical customer support over diverse industries and company sizes. While on a tactical level, further validation outside of technical customer support is supposed, the next step is to form a generalized and unified framework on mobile-integrated business processes using the validated hypotheses and validate this with a large-scale empirical study in order to come to valid theories and help practitioners to be successful in this domain.

Literature

- [Aber04] The Mobile Field Service Benchmark Report - Bridging the Chasm Between the Field and Back Office. <http://www.aberdeen.com/summary/report/benchmark/mobileFS-093004.asp>, Last Access: 2006-07-12.
- [BeGM87] Benbasat, Izak; Goldstein, David. K.; Mead, Melissa: The case research strategy in studies of information systems. In: *MIS Quarterly* 11 (1987) 3, pp. 369-386.
- [Chil72] Child, John: Organizational structure, environment, and performance: The role of strategic choice. In: *Sociology* (1972) 6, pp. 1-22.
- [ChNa04] Chen, Lei-da; Nath, Ravi: A framework for mobile business applications. In: *The International Journal of Mobile Communication* 2 (2004.) 4, pp. 368-381.
- [Cres03] Creswell, John W.: *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 2nd edition, Sage Publications Inc. Chennai India 2003.
- [Davi89] Davis, Fred D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. In: *MIS Quarterly* 13 (1989) 3, pp. 319-341.
- [DiSt98] Dishaw, Mark.T.; Strong, Diane M.: Supporting software maintenance with software engineering tools: A computed task-technology fit analysis. In *Journal of Systems and Software* 44 (1998) 2, PP. 107-120.
- [DiSt99] Dishaw, Mark.T.; Strong, Diane M.: Extending the technology acceptance model with task-technology fit constructs. In *Information & Management* 36 (1999) 1, pp. 9-21.
- [Eise89] Eisenhardt Kathleen M.: Building Theories from Case Study Research. In *Academy of Management Review* 14 (1989) 4, pp. 532-550.
- [FoGK05] Fouskas, Konstantinos G.; Giaglis, Georg M.; Kourouthanassis, Panos E.: A roadmap for research in mobile business: In: *International Journal of Mobile Communications*. 3 (2005) 4, pp. 350-373.
- [GeSh04] Gebauer, Judith; Shaw, Michael: Success Factors and Impacts of Mobile Business Applications: Results from a Mobile e-Procurement Study. In: *IJEC*. 8 (2004) 3, pp. 19-41.
- [Gogn05] Cognito Limited: *Mobile Data: Avoiding the costly pitfalls*. <http://www.cognito.co.uk>, Last Access: 2006-07-24.
- [GuPo05] Gumpp, A.; Pousttchi, K.: The Mobility-M-framework for Application of Mobile Technology in Business Processes. In: Cremers, A. B.; Manthey, R.; Martini, P.;

Steinhage, V. (Hrsg.): INFORMATIK 2005 - Informatik LIVE! Band 2 - Beiträge der 35. Jahrestagung der Gesellschaft für Informatik e. V. (GI) 19.-22. September 2005 in Bonn. LNI. Bd. P-68, Bonn 2005, pp. 523–527.

- [GoTh95] Goodhue, Dale L.; Thompson, Ronald L.: Task-technology fit and individual performance. In: MIS Quarterly 19 (1995) 2, pp. 213–236.
- [HaCh93] Hammer, Michael; Champy, James: Reengineering the Corporation – A Manifesto for Business Revolution. New York 1993.
- [HoMO93] Hollingshead, Andrea B.; McGrath, Joseph E.; O'Connor, Kathleen M.: Group task performance and communication technology: A longitudinal study of computer-mediated versus face-to-face groups. In: Small Group Research 24 (1993) 3, pp. 307–333.
- [MaKe98] Mathieson, Kieran; Keil, Mark: Beyond the interface: Ease of use and task/ technology fit. In: Information & Management 34 (1998) 4, pp. 221–230.
- [Roge95] Rogers, Everett M.: Diffusion of Innovations, 4th edition, Free Press New York 1995.
- [PoTh05] Pousttchi, K., Thurnher, B.: Einsatz mobiler Technologien zur Unterstützung von Geschäftsprozessen (WCI). Berlin (2005), pp. 101 – 120.
- [PoTh06] Pousttchi, K., Thurnher, B.: Understanding Effects and Determinants of Mobile Support Tools: A Usability-Centered Field Study on IT Service Technicians. ICMB. IEEE. Copenhagen, Denmark 2006, S.1-8.
- [ScPL06] Scornavacca, Eusebio; Prasad, Mishul; Lehmann, Hans: Exploring the organisational impact and perceived benefits of wireless Personal Digital Assistants in restaurants. In: The International Journal of Mobile Communication 4 (2006) 5, pp. 558-567.
- [ToKl82] Tornatzky, Louis G.; Klein, Katherine J.: Innovation characteristics and innovation adoption-implementation: A meta-analysis of findings. In: IEEE Transaction on Engineering Management 29 (1982) 1, pp. 28–45.
- [TuPo04] Turowski K. ; Pousttchi, K.: Mobile Commerce – Grundlagen und Techniken. Heidelberg 2004.
- [WaCh04] Wang, Sophia; Waiman, Cheung: E-Business Adoption by Travel Agencies: Prime Candidates for Mobile e-Business: In: International Journal of Electronic Commerce. 8 (2004) 3, pp. 43–63.

Critical success factors for mobile field service applications: A case research

Andreas Birkhofer, Sina Deibert, Franz Rothlauf

Lehrstuhl für Wirtschaftsinformatik I
Universität Mannheim
68131 Mannheim
abirkhof@rumms.uni-mannheim.de ,
{deibert, rothlauf}@uni-mannheim.de

Abstract

This paper presents a multiple case research concerning success factors and issues of mobile field service implementations. Based on Eisenhardt's explorative research design, five successful mobile technology implementations are compared and analysed. Important aspects of the implementation strategy and the used technology are examined and discussed. The results indicate that the success of mobile field service implementations is influenced by the implementation strategy, project management, change management, and technological restrictions of the used mobile technology. The analysis identifies the complexity of technological components and their limitations as well as aspects influencing user acceptance as central challenges for mobile technology implementations.

1 Introduction

Many publications proclaim great potential for efficiency improvements in business processes by using mobile technology [Kala03, Pflu02]. The characteristics of mobile technology like mobility, reachability or localisation offer new possibilities to contact mobile workers and organize their tasks more efficiently [PoTW03, SiSh03]. A typical example of a mobile worker is a field service technician who has to perform maintenance activities at different locations. Due to this mobility of the task, the worker is unable to work at a stationary workplace [KrLj98]. Therefore, efficiency gains could be realised in the coordination and documentation of the work

using mobile technology [VaHe02]. For example, redundant work due to paper-based documentation and manual data input can be reduced.

While the potential impact of the mobile technology is hardly questioned, the realization of these potentials is rarely addressed in IT research. Furthermore, many failed IT projects show that the implementation process is a crucial factor for the success of the integration and use of a new technology [Brow04]. So far, there is only limited research on mobile technology implementations. There are some case studies analysing benefits and possible barriers in multiple case analysis [BrVe05, ScMc04]. Nah, Siau and Shen presented a single case study and derived some fundamental objects and aspects influencing the success of mobile technology [NaSS04]. A more conceptual approach is applied by Wang, van de Kar and Meijer, who try to analyse the implementation process using the method of the way of working, thinking, designing and controlling [WaKM05]. The single case study of this report is enlarged to a multiple case study by Wang [Wang05]. Though these are good descriptions of the application development, a structured description of success factors and potential problems is still missing. To identify success factors and issues of current mobile technology implementations, a multiple case study in the area of mobile technologies is needed.

The goal of this paper is to analyse implementation strategies and critical success factors of successful mobile field service applications comparing different cases in one study. Studying different cases in one study allows drawing more relevant conclusions as a direct comparison between different cases is possible. In this study, we also compare the different success factors between mobile technologies and other standard IS implementations.

In the next section, the general research methodology and analysis framework is described. After a short description of the cases in section three, the key findings of the study are presented in section four. In section five, the main findings are discussed and also related to other studies from the literature. The paper ends with a short summary and some conclusions.

2 Methodology

This section describes the methodology we used in our study. We discussed the analytical framework and the way we conducted our empirical research.

In this study, we identify the strategic patterns of the mobile technology implementation by analyzing some selected real-world cases. For the analysis, we used successfully implemented mo-

mobile applications. For the design of our studies we follow the structure of qualitative research proposed in [Kepp00]. The used method of multiple case research is based on Eisenhardt's explorative research approach [Eise89]. For the data collection, we used guided semi-structured interviews [Yin02]. These interviews were conducted either face to face or by telephone. The direct interaction with the interview partner reduced problems of possible misunderstandings [FrMe95]. The impact of the direct contact on the argumentation is minimized in the data analysis [Atte2003, ScHE05]. The different cases are compared using a process-based approach [MaRo88].

We analysed five successful implementations of mobile technology as exemplary case studies. By focussing on one specific type of mobile technology implementation (mobile field services), we reached a homogenous setting of the study. To consider different technologies as well as different stakeholders, we chose cases from three different software providers. We interviewed the project manager of the provider and the customer as well as a user of the application. Overall, we conducted 16 interviews; at least three for each case. The interviews were structured in three parts. The first part consisted of questions concerning the background and the trigger for the implementation of the mobile technology. The second part analysed the proceeding of the implementation and the used technology. The third part evaluated the results of the projects as well as the problems and success factors of the implementation. The length of each interview was about one hour. All interviews were based on a semi-structured interview guideline with open answers. The analysis is based on the transcripts of the interviews and, if available, background material provided by the interview partner.

3 Case Descriptions

In this section, we will shortly depict the five cases under consideration for this research paper. First, general aspects will be explained. In addition, a tabular listing will take a closer look on the different examples.

For this research five different companies in the German speaking area which recently implemented a mobile solution have been analysed. In three cases the analysed implementation of the mobile solution was a technological novelty. For the other two companies the analysed projects were replacements for already existing mobile solutions. Installing mobile solutions for the first, the goals of the solution were basically reductions of media disruptions and process efficiency,

the replacements should lead to a cost reduction concerning the maintenance of the mobile solutions.

Though all five solutions are implemented to facilitate field service activities of the field service technicians the solutions differed concerning the standardization of the application, the mobile device and the synchronisation of the mobile application and the back-end system. Table 1 shows some central aspects of the five case studies.

	Case A	Case B	Case C	Case D	Case E
Occupation	Heating systems	Facility management	Utility	Heating systems	Industrial gas
Type of usage	Field services	Field services	Field services	Field services	Field services
Existing mobile solution	No	No	Yes	No	Yes
Count of users	130	1000	100	200	100
Offline/Online	Offline	Offline	Offline	Offline	Offline
Synchronisation	GPRS/GSM	GPRS	LAN, DSL, ISDN	GPRS	LAN, DSL, ISDN, GPRS
Type of solution	Individual	Individual	Standardized	Standardized	Standardized
Mobile devices	Pocket PC	PDA and Laptop	Laptop	Tablet PC/Laptop	Laptop
Reason for the mobile solution	<ul style="list-style-type: none"> - Elimination of media disruptions - Process efficiency - Centralisation of administration 	<ul style="list-style-type: none"> - Elimination of media disruptions - Process efficiency 	<ul style="list-style-type: none"> - Reduction of the maintenance costs - New functions 	<ul style="list-style-type: none"> - Efficiency of documentation - Elimination of media disruptions - Process efficiency 	<ul style="list-style-type: none"> - End of maintenance contract - Low user acceptance of old solution

Table 1 – Description of the analysed cases

4 Key Findings

This section depicts the key findings of the case research. The results address factors that are important during implementation and which are critical for the success of the implementation. First, as the mobile technology is a key aspect for our research, we will analyse the technological situation and identify success factors and problems regarding the different components of the mobile solution. In section 4.2 we analyse the implementation strategy of the projects. Therefore we describe the key strategy and identify critical success factors.

4.1 Technology

A mobile solution is a complex system consisting of different components. In this subsection, we analyse success factors and issues of the mobile solution and some of its components. As this is no technology analysis but an implementation research, we focus on the most important issues mentioned in the interviews:

- interfaces to back-end systems,
- mobile middleware,
- mobile application,
- synchronisation medium/network,
- mobile device, and
- mobile solution.

As all cases intend to use the data and information from the mobile application in the back-end system, *interfaces between mobile solution and back office system* have to be designed. All cases state this aspect as important and were satisfied with their application. Therefore, the design of interfaces is not viewed as a major problem. One reason for this aspect might be that all cases use standardized back-end systems and these are designed by the same provider (three out of five cases).

The next component of the solution is the *mobile middleware* which is used in a standard version. Therefore, the middleware was not viewed as a challenging aspect in most cases. However, in one case about 300,000 equipments should be provided by the mobile middleware. This *amount of data* could not be executed by the middleware the provider distributed with the solution. As the provider could not offer a different middleware a preliminary CD-based procedure had to be added to the solution to circumvent this problem. This solution offers the possibility to provide the necessary data for the equipments and changes can be quickly included by a new version of the CD. A final solution is intended to be implemented soon.

The counterpart to the middleware is the mobile application which runs on the mobile device. We can split the mobile application in two parts. The first part is the business logic that is implemented on the mobile device. The second part is the user interface for the data input. The features and the flexibility of the business logic design depend on both, the standardization of the software and the flexibility of the solution provider. In the five cases missing functionalities and/or software problems were viewed as challenging for the implementing process of the mobile application. However, with updates and new releases most of those problems are solved. Although the *stability of the business logic* is essential, the *user interface* is the most important aspect for the interview partners. For them, the design of this part of the solution is most critical for the acceptance of the user. It should be intuitive, easy to use, and only few clicks should be needed. This is important because field service technicians are usually not used to work with computers. Therefore, the ease of use is viewed as crucial for the acceptance of the whole appli-

cation. In all cases, the user interfaces are customer-designed and often inherit components known from the paper-based process to help the user to accustom to the new technology.

The data has to be transmitted from the application to the back end; hence the *synchronisation* is also mentioned as an important aspect by the interview partners. The field service technicians often work in cellars and other areas without constant network access. Consequently, none of the cases used an always-online solution. All applications are basically offline solutions with a synchronisation mechanism to update the information. The synchronisation mediums differ among the cases. While two companies use wired connection like LAN, DSL, ISDN, or WLAN for synchronisation, the other three solutions synchronise via GPRS. The medium, frequency, and initiation of the synchronisation depend on the need for actuality. Among other reasons, failed synchronisations can have their origin in network problems since the *availability of the mobile network* (GPRS) is limited. It might be worthwhile to test different telecom providers and network cards as the availability and the operational functions can differ. For example, in two cases the replacement of the network card resulted in more reliable communication.

In most interviews, the *mobile devices* are seen as critical success factors. It is important to select a mobile device that fits well to the task and the environmental circumstances. The interview partners suggest to carefully analyse the working environment and the previous working process and then to decide which device meets these requirements. In all cases, problems with mobile devices occurred. In two cases the problem occurred due to the insufficient *robustness of the devices* which resulted in a high failure rate. Additional equipment for robustness like a special plastic cover or replacing the device by a more rugged device was needed to reduce this problem. Another problem was the limited *screen size* of the mobile devices which result in problems for inexperienced users. This makes the data input complicated and cumbersome. Even with devices specially manufactured for the customer, there might be problems as one case showed. The provider of the device was not able to provide an acceptable solution in time. Consequently, the scheduled time of the implementation had to be postponed. Finally, the customer had to use a different device from a different provider.

Another problem related to mobile devices is the limited *amount of data* which can be synchronised or stored on the mobile device (especially when dealing with “mass data”). A recommendation to reduce this problem is to focus on the relevant information needed for the process.

In the analysed cases, the applications using laptops seem to have fewer problems than the cases with PDAs and tablet-PCs. This may be because of fewer restrictions regarding size and capac-

ity. However, this does not mean that laptops are always the better solution. Limitations concerning data input, size, and weight might not allow using laptops.

Finally, an integrated view on the complete mobile solution and the *interoperability of the different components* is important. A fundamental expectation of all cases is that *the application should run stable and secure* without much effort for support and maintenance. Depending on the confidentiality of data, different encoding mechanisms and systems are used to guarantee data and system security in all five cases. The stability of the systems is seen positive in all cases (this is expected as we only study successful applications). However, the *complexity of the different components* led sometimes to difficulties in the traceability of the origin of failure. To solve such problems, a high knowledge on the different components and the underlying business processes was seen to be important.

4.2 Implementation Strategy

This subsection discusses the strategy of the mobile technology implementation. Different strategies are compared and critical success factors and issues of the implementation are identified. Furthermore, we study the influence of user experience on the *implementation strategy*.

We have been able to identify the following phases in the implementation strategy: analysis of the current situation, conceptual design, business blue print, software engineering, software tests, pilot phase, roll-out and adoption. These are typical steps of *phase models known from standard software implementations* [Balz82, Boeh86, StHa05]. Although the phases slightly vary between companies and projects, the key procedure is the same. We have not identified any differences between standard software solutions and individually-engineered software. This might be due to the fact that two of the cases with standard software solutions were projects in the launching period of a new software version and all providers had at least a standardized technological framework. Furthermore, the need for individual designs of user interfaces and adjustments based on the individual business processes might have led to the similar approaches. This aspect is supported by statements of most providers that two mobile technology implementations can hardly be compared to each other.

In the following paragraphs, we describe the success factors and problems of the implementation process observed in the five cases. We discuss differences in the process and split the analysis in four parts by aggregating some of the phases identified above. These are:

- concept,

- development and testing,
- roll-out, and
- adoption.

4.2.1 *Concept*

The first stage of a mobile technology implementation is the conceptual phase. In this part a proper setting for the whole implementation has to be defined. The interview partners regard this stage as the most important part of a mobile technology implementation. Hidden problems and not included functionalities can result in severe consequences for the rest of the implementation as later adjustments will be complicated and expensive.

The interview partners started with a *detailed description of the relevant processes* by analysing the present situation and describing the expected new situation. All relevant processes and sub-processes are verified and, if necessary, synchronized. In some cases not every process was modelled in full detail. This resulted in adjustments and extensions in the pilot phase. These adjustments were time and cost intensive.

Furthermore, the *cooperation between the provider and the customer* was estimated to be essential for the success in all cases. Issues and problems have to be solved together and a common idea of the goals and needs regarding the solution has to be developed. This demands flexibility on both sides and the provider has to understand the way of working and thinking of the customer to create a satisfactory solution.

In all cases a new mobile solution was introduced the first time, a large portion of users existed which are not used to work with mobile devices. Many of them were afraid of the new technology. Concerns were possible supervision by the management based on the characteristics of the mobile technology or to lose their jobs based on the streamlining of the business process. There were also concerns of not being able to cope with the new technology. Some *user resistance* was also based in the unwillingness of some employees to change their way of working. *Communication and information* was estimated as necessary to reduce these fears and resistances. The goals and background of the implementation have to be explained to the users as early as possible and in direct conversation. In addition, as much information as possible should be spread during the project. The information has to be accessible for all users. In one case, there was the problem that relevant information was available in the intranet but not all users had access to this. In those cases mobile solutions have already been used for some years, no

fears or resistances have been observed as users were used to work with mobile devices and hoped for efficiency gains.

The following aspect is also related to user acceptance. According to some of the interview partners, the *user should be involved in the project* already in the conceptual design of the solution. The users know the environment and workflow and have their own expectations and needs concerning the mobile application. In the first draft, the integration is essential for a good solution in the first draft and results in fewer later adjustments of the application. Furthermore, users integrated in an early stage were more devoted to “their” project and experiences from motivated users helped to prevent resistances and to reduce fears of their colleagues. However, the level of user integration strongly varied among the cases. While some cases integrated selected users from the very beginning as fully accepted team members, other cases include the users just for advisory comments on the user interfaces and for software tests. The reason of these different approaches is not based on technology, but on the philosophy of the project leaders. The interviews showed that involved users show a higher commitment towards the solution and interest in a further development and extensions.

The project managers emphasized that, besides user involvement, it is necessary to either integrate or to regularly *inform all stakeholders* that are affected by the implementation. This also includes those departments in the back-office which are indirectly affected by the implementation of a mobile solution. Especially the management board and the workers’ council should be kept updated. In two cases, the information flow to the workers’ council was not sufficient. Consequently, they tried to stop the project which resulted in a prolongation of certain phases. There are many persons involved in such an implementation, thus the task for the project leader is quite demanding and therefore *a detailed project plan with strict deadlines and defined milestones* is important. Furthermore, the interview partners demand enough *power for the project manager*. The quicker a project leader can decide the better he/she can react on problems and necessary adjustments. In the cases observed, there existed minor problems concerning deadlines, which could be solved due to buffer periods and the flexibility of the parties.

4.2.2 *Software Engineering and Tests*

When the concept is finished, the software has to be engineered or adjusted according to the needs of the customer. Then, the final version has to be tested, first by the provider and then by the customer. In almost all cases, the customer was already able to test preliminary versions and check if the expectations are matched, especially regarding the user interface.

In all cases, the most relevant test was a *pilot period*. Therefore, some user had to test the new solution in real-world environments. Based on the high number of settings and equipments of the field service activity, the pilot phase was essential to early identify problems and eliminate these. The interview partners suggested using a hand-selected group of users, consisting of some who are used to work with computers and some inexperienced users. They should test the solution and report problems. As aforementioned, a stable solution is vital for the success of the implementation, therefore software problems and other issues have to be eliminated in the pilot period. To guarantee this, a vital cooperation between pilot users, IT department and the solution provider is necessary as the improvements have to be realized quickly.

The cases revealed two problems based on strategic decisions. In one case, the customer's *decision for a mobile device* was not made until the test period of the software started. Thus, the test of the mobile device process was parallel to the test period of the software. Consequently, the error tracing was sometimes difficult, as the problem might have originated in the different settings of the hardware as well as in the software. This ended in additional effort for testing and a prolongation of the test period. Thus, the hardware has to be chosen either before the test period or before the software engineering is about to start.

In another case, ill-defined *responsibilities* between the provider and its subcontractor caused some problems, as both tested their own part separately, but none took the responsibility for the whole solution. This led to problems in software testing, as the components worked fine alone but did not work properly when combined. This incident resulted in a prolongation of the whole test period as neither of the parties wanted to be responsible for the malfunction and blamed the other party. Therefore the responsibilities have to be well-defined at the very beginning.

4.2.3 Roll-Out

After eliminating problems in the pilot period, the application has to be rolled-out. An essential part of this is the *training of the user*. The experiences from the case examples showed that it is useful to educate computer inexperienced users first on the new device and the operating system before starting with the training for the mobile application. To separate groups with respect to their technological knowledge might also be reasonable, however, not always possible. The users should be trained with the interfaces they will actually use. The time necessary for training increases with the complexity of the application. In addition to trainings, almost all cases introduced *key-users* to help their colleagues with operational questions. Especially inexperienced users accept the help of a colleague more easily than an advice of an IT-specialist. These per-

sons have to be selected carefully, as the interview partners state that certain users caused a lot of additional work. An impatient treatment would discourage the user and reduce the acceptance of the new technology.

One case example allowed inexperienced users to use their old paper-based documents parallel to the new technology. Furthermore, the technicians started with using only parts of the functionalities and extend the use of the new solution step by step. This *slow changeover* allows the user to slowly adapt to the new solution and to get more confidence in their technological capabilities.

Problems in this stage of the implementation were mainly based on an overly optimistic estimation of the capabilities of the users. In some cases the length of training sessions was too short and had to be extended. In one case, due to software problems, the training period started about two weeks before the actual go-live. This delay between the trainings and the actual use of the new knowledge resulted in additional support from their colleagues to use the new application, as some of the users could not remember all aspects concerning the new application. Thus, trainings should not be conducted until the software is stable enough to go live.

4.2.4 Adoption

The interviews revealed that the users accepted the new tool if it does not impede his/her daily life. Ideally, it offers some improvements in comparison to the situation before the solution was implemented. Consequently, the *speed of eliminating problems and difficulties is essential* for the acceptance of the application of the user. Problems that occur must be solved rapidly. On the one hand software problems have to be fixed on short-hand notice by the IT department and the provider, respectively. On the other hand, *fallback scenarios* should be planned for hardware problems and alternative devices should be available in short period of time. Unproductive waiting and driving times to get the system running again will reduce acceptance, especially among critical users. Problems with user support were low. In one case, the late assignment of the person responsible for support and maintenance represented a challenge. Although the application was defined at the beginning of the project, in one case it was not possible to find a suitable person for support until the pilot users started. This could have resulted in major problems as at this time usually the new application needs the highest amount of support. However, in the actual case, other persons from the project team could fill the gap. Nevertheless, this person should be assigned at the very beginning and be an important part of the project team. The key aspects of the last section are summarized in figure 1.

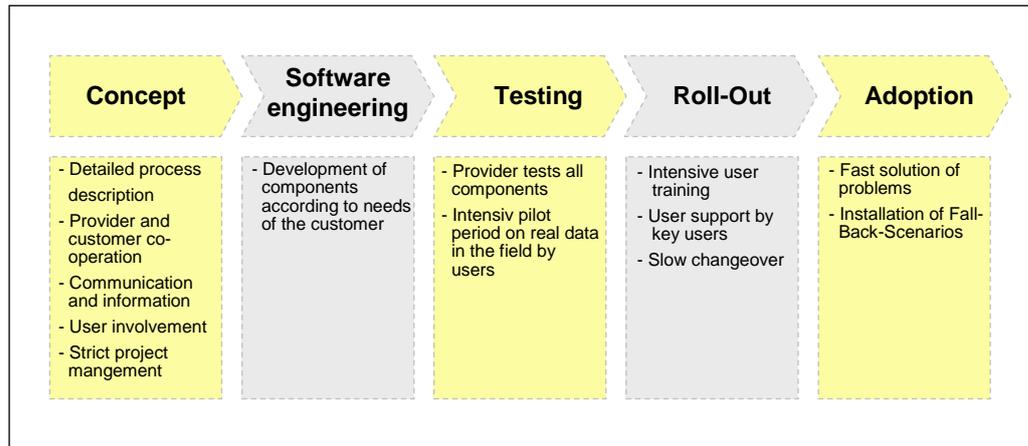


Figure 2 – Aspects of the implementation procedure

5 Discussion

We discuss the key findings of the case research on mobile field service implementations. The previous listing of the success factors and problems might be helpful as guideline, but we also relate the findings to existing observations from literature. We also examine the limitations of the study due to the research design.

The implementation strategies for the mobile applications in our cases are based on standard phase approaches [Balz82], [Boeh86] [StHa05]. The cases do not indicate any peculiarities which would suggest a special implementation strategy for mobile technology.

Regarding the success factors for a mobile technology implementation, the analysis revealed a number of issues and problems originating from different aspects. According to Riemenschneider [Riem01], success factors could be classified into the groups implementation research, project management, change management and technology specific aspects. We can use this categorization for classifying the results observed in the study.¹ Figure 2 shows some aspects according to this classification.

¹ Riemenschneider studied the implementation of integrated management systems. In his work the technological aspects are labelled as issues concerning integrated management systems instead of technological aspects.

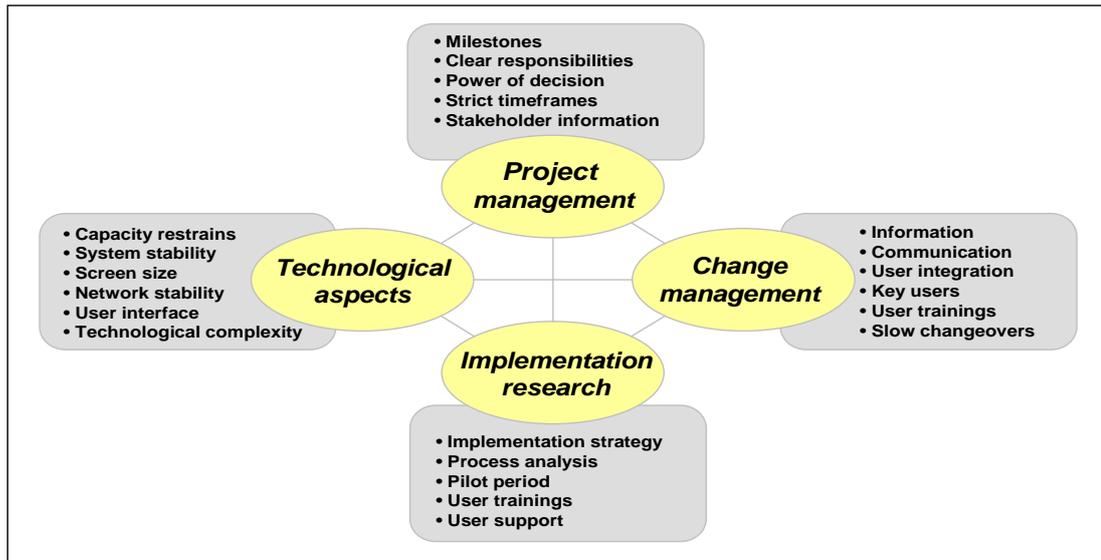


Figure 2 – Important aspects for mobile field service implementation

The identification of these success factors and the different influential areas is not surprising. Regarding the situation when the implementation research started, computer and information systems represented technological innovations and potential users were not used to work with computers. Therefore the implementation research for information systems was often centred on user acceptance [MaOp83, Riem01]. The situation for first-time mobile technology implementations for field service technicians is similar because most of the technicians are also confronted with the new technology for the first time. These service technicians experience the mobile solution as change in their way of working. Change management is focused on such aspects as it addresses the resistance of users against organizational changes [Reis97ab], [BoPi92]. Reiss describes four aspects that influence the resistance towards organizational change. These are lack of information, lack of knowledge/capabilities, lack of motivation and organizational boundaries [Reis97a]. As described above, the studied cases of first time implementation showed a great amount of resistance originating from all four areas. In some cases, the percentage of users with a negative attitude towards the new solution was up to seventy percent. However, it was possible to reduce these resistances by communication, training, support, creation of personal benefits, and integration [Reis97b]. Although the user resistance was lower in the two cases with technology replacements, the interview partners remarked that the expectations of the users concerning the stability and usability of the solution still define the central and most crucial aspect for a successful application. As all cases used a project organization for the im-

plementation, the management and coordination of the different persons and interests involved in this temporary collaboration also represents a success factor [Lech96].

Analysing the specific aspects of mobile technology, we identified a couple of restrictions and limitations based on the different components like capacity restrictions, connection problems, or size and usability of the mobile device. These effects are already observed in previous publications [KrLj99, LeBe04]. The case examples showed that a central aspect for a successful mobile technology implementation is to master the restrictions of the devices based on the mobile technology. This means to accept the limitations of the different components and create a stable solution suitable to the specific requirements of each company. In addition, the design of the user interface signifies a central aspect for a successful mobile field service implementation.

Therefore, the main challenge of a mobile technology implementation is the high complexity of the different components and the interaction between technology, user and other stakeholders in the implementation process.

This study offers an overview on success factors and issues of mobile technology implementations based on five case studies. However, there are some limitations due to the used research design: based on the explorative research there is a potential bias due to the direct influence of the interviewer [Atte2003]. We tried to eliminate the distortion in the data analysis; however, it is difficult to fully eliminate this bias.

Furthermore, the used information is a significant issue. As the results of our study are based on the answers of the interviews, we could only make use of information provided by the interview partners. Therefore, problems or success factors that are not addressed in the interviews are not included. In addition, the different interview partners were not integrated into the project to the same amount.

The setting of the study offers two other problems. First we analysed only successful mobile technology implementations. This eliminates the possibility to study which factors lead to a failure of projects. The second problem of the setting is that in most of the projects the mobile applications were used mainly for reporting. Therefore, the consequences of a system breakdown are limited as the information is not time critical and technicians perform most of his work without the solution.

6 Summary and Conclusions

This case research illustrates success factors and issues of mobile field service implementations. The paper describes the lack of such an analysis and motivates the need for a direct comparison of different mobile field service implementations. We discussed the explorative research methodology based on multiple case studies. After a short description of the cases and the situation before the mobile technology was implemented, we presented the key findings of the case studies. We illustrated success factors and issues of a mobile technology implementation originated from the technology and the implementation strategy. Finally, the key findings are discussed and compared to existing results from the literature on software and innovation implementations.

We summarize the key findings of this study: first, the results can be used as a guideline for implementing mobile applications. We showed relevant success factors, issues, and actual problems of implementing mobile technology. Second, a mobile field service implementation is based on aspects regarding project management, implementation research, change management, and mobile technology. A successful implementation addresses success factors from the different fields and respects the existing limitations. Especially the technological limitations have to be mastered as the mobile technology is still a developing area. Third, a first time implementation often has to face resistances of the users against the new technology. These resistances have to be encountered by communication, integration, trainings, and support. We found resistance lower if the users do not view the introduction of mobile applications as an organizational change (e.g. when replacing existing mobile solutions).

Further research including failed implementations is necessary to confirm the results of this study. Furthermore, we develop in future research a ranking of the importance of the different aspects regarding a failure of the introduction of mobile applications.

References

- [Atte03] Atteslander, Peter: Methoden der empirischen Sozialforschung, 10. Auflage, Berlin, New York: de Gryter, 2003.

- [Balz82] Balzert, Helmut: Die Entwicklung von Software Systemen: Prinzipien, Methoden, Sprachen, Werkzeuge, Mannheim, Wien, Zürich: Bibliografisches Institut, 1982.
- [Boeh86] Boehm, Barry W.: Wirtschaftliche Software-Produktion, Wiesbaden: Forkel Verlag, 1986.
- [Brow04] Brown, Wayne: Enterprise Resource Planning (ERP) Implementation Planning and Structure: A Recipe for ERP Success, User Services Conference Proceedings of the 32nd annual ACM SIGUCCS conference on User services, Baltimore, pp. 82-86, 2004.
- [BrPi92] Bromann, Paul; Piwinger, Manfred: Gestaltung der Unternehmenskultur: Strategie und Kommunikation, Stuttgart: Schäffer-Poeschel, 1992.
- [BrVe05] Brodt, Thorsten L.; Verburg, Robert M.: Managing Mobile Work – Insights from European Practice, eChallenges, Ljubljana, 2005. URL: <http://www.alexandria.unisg.ch/Publikationen/29013> (20.04.2006).
- [Eise89] Eisenhardt, Kathleen M.: Building Theories from Case Study Research. In: Academy of Management Review, 14, 4, pp. 532–550, 1989.
- [FrMe95] Frey, James H.; Mertens Oishi, Sabine: How to conduct interviews by telephone and in person, Thousand Oaks, London, New Delhi 1995.
- [Kala03] Kalakota, Richard: Mobile Business: Moving from Vision to Strategy. In: E-Business Review, pp. 49-55, 2003.
- [Kepp00] Kepper Gaby: Methoden der qualitativen Marktforschung, in: Hermann, Andreas; Homburg, Christian (Hrsg.): Marktforschung: Methoden, Anwendungen, Praxisbeispiele, 2. Auflage, Wiesbaden, pp. 159-202, 2000.
- [KrLj99] Kristoffersen, Steinar; Ljungberg, Frederik: “Making place” to make IT work: empirical explorations of HCI for mobile CSCW, Proceedings of the international ACM SIGGROUP Conference on Supporting group work, Arizona, 1999.

- [KrLj98] Kristoffersen, Steinar; Ljungberg, Frederik: Mobile Informatics: Innovation of IT use in mobile settings, IRIS'21 Workshop Report. Gothenburg, 1998. – URL http://www.viktoria.se/groups/mi2/results/papers/MI_workshop_IRIS21.pdf – (03.06.06).
- [Lech97] Lechler, Thomas: Erfolgsfaktoren des Projektmanagements, Frankfurt et al.: Lang, 1997.
- [LeBe03] Lee, Yan E.; Benbasat, Izak: Interface design for mobile commerce, Communications of the ACM, Vol. 46, No. 12, pp. 49-42, 2003.
- [MaOp83] Mambray, Peter; Oppermann Rainer: Beteiligung von Betroffenen bei der Entwicklung von Informationssystemen, Frankfurt: Campus Verlag, 1983.
- [MaRo88] Markus, M. Lynn; Robey, Daniel: Information Technology and Organizational Change: Causal Structure in Theory and Research, Management Science, 34, 5, pp. 583-598, 1988.
- [NaSS05] Fui-Hoon Nah, Fiona; Siau, Keng; Sheng, Hong: The VALUE of Mobile Applications: A Utility Company Study, Communications of the ACM, 48 , pp. 85-90, 2005.
- [Pflu02] Pflug, Volker: Mobile Business macht Geschäftsprozesse effizient. in Gora, Walter ; Röttger-Gerigk, S. (Hrsg.): Handbuch Mobile-Commerce – Technische Grundlagen, Marktchancen und Einsatzmöglichkeiten. Heidelberg: Springer-Verlag, pp.211-224, 2002.
- [PoTW03] Pousttchi, K.; Turowski, K.; Weizmann, M. (2003): Added Value-based Approach to Analyze Electronic Commerce and Mobile Commerce Business Models, in: *Andrade, R.; Gómez, J.; Rautenstrauch, C.; Rios, R. (Hrsg.): International Conference Management and Technology in the New Enterprise. La Habana, pp.414-423, 2003.*
- [Reis97a] Reiss, Michael: Change Management als Herausforderung, in: *Reiss, Michael; von Rosenstiel, Lutz; Lanz, Anette [Hrsg.] Change Management – Programme, Projekte und Prozesse. Stuttgart: Schäffer-Poeschel Verlag, 1997.*

- [Reis97b] Reiss, Michael: Instrumente der Implementierung, in: *Reiss, Michael; von Rosenstiel, Lutz; Lanz, Anette [Hrsg.] Change Management – Programme, Projekte und Prozesse*. Stuttgart: Schäffer-Poeschel Verlag, 1997.
- [Riem01] Riemenschneider, Frank: Implementierung integrierter Managementsysteme: Erfolgsfaktoren für die Unternehmenspraxis, Wiesbaden: Deutscher Universitätsverlag, 2001.
- [ScMc04] Scheepers, Helena; McKay, Judy: An empirical assessment of the business value derived from implementing mobile technology: A case study of two organisations, in: *T. Leino, T. Saarinen & S. Klein (Hrsg.), Proceedings of the 12th European Conference on Information Systems (ECIS2004)*, pp. 1-13. Turku, 2004.
- [ScHE05] Schnell, Rainer; Hill Paul B.; Esser, Elke: Methoden der empirischen Sozialforschung, 7. Auflage, München, Wien 2005.
- [SiSh03] Siau, Keng; Shen, Zixing: Mobile communications and mobile services, *International Journal of Mobile Communication*, Vol. 1 No.1/2, pp. 3-14, Nebraska, 2003.
- [StHa05] Stahlknecht, Peter: Einführung in die Wirtschaftsinformatik, 11. Auflage, Berlin, Heidelberg, New York: Springer Verlag, 2005.
- [VaHe02] Valiente, Pablo; van der Heijden, Hans: A method to identify opportunities for mobile business processes, SSE/EFI Working Paper Series in Business Administration, No 2002:14, Stockholm: Stockholm School of Economics, 2002.
- [Wang05] Wang, Yan: Emperical Case Study Report, Technische Universiteit Delft, Delft, 2005. URL: <http://www.tbm.tudelft.nl/webstaf/yanw/casestudy.pdf> (15.05.2006)
- [WaKM05] Wang, Yan; van de Kar, Els; Meijer, Geleyn: Designing Mobile Solutions for Mobile workers – Lessons Learned from a Case Study, Proceedings of the 7th international conference on Electronic commerce, Xi'an, 2005.
- [Yin03] Yin, Robert K.: Case Study Research: Design and Methods, Sage Publications, London, 2003.

Migration als Ansatz zur Gestaltung mobiler Services

Oliver Bohl, Shakib Manouchehri, Udo Winand

Forschungszentrum für Informationstechnik-Gestaltung (ITeG)

Universität Kassel
34125 Kassel

{bohl, manouchehri, winand}@wirtschaft.uni-kassel.de

Abstract

Die flächendeckende Verbreitung mobiler Endgeräte und Infrastrukturen, welche den Nutzern eine erhöhte Mobilität bei gleichzeitig mannigfaltigen Nutzungsmöglichkeiten erlauben, fördert die Suche nach Erfolg versprechenden mobilen Services. Beim Aufbau neuer Services erscheint die Nutzung existierender Ressourcen sowie des bestehenden Know-hows ratsam. In diesem Zusammenhang stellt Migration eine Methode zur Gestaltung mobiler Services dar. Der vorliegende Beitrag beleuchtet Spezifika der Migration als methodischen Ansatz zur Entwicklung entsprechender Angebote.

1 Grundlagen

Mit der weit reichenden Verbreitung mobiler Endgeräte und der flächendeckenden Verfügbarkeit leistungsstarker Infrastrukturen für die Mobilkommunikation geht eine Suche nach Erfolg versprechenden mobilen Services einher. Vor allem Telekommunikationsanbieter trachten nach so genannten Killerapplikationen/-services [Tele04; ChSk05]. Datenübertragungstechnologien wie WLAN, Bluetooth, GSM und UMTS ermöglichen die mobile Nutzung von Daten und eröffnen Möglichkeiten, Services überall zu nutzen und anzubieten [Saty02]. In Kombination mit mobilen Endgeräten, die dem Nutzer vielfältige Nutzungsmöglichkeiten erlauben, bieten sich Chancen zur Entwicklung mobiler Services (m-Services) mit innovativen Prozess-, Geschäfts- und Erlösmodellen. Dabei kann in Teilbereichen auf Bewährtes aufgesetzt werden, um Risiken der Neuentwicklung abzufedern [BFKS04]. So

können stationär genutzte Services um mobil realisierbare, zusätzliche Komponenten ergänzt werden, um Mehrwerte auf Anbieter- und Nutzerseite zu erzeugen.

Der Beitrag hat vor diesem Hintergrund zum Ziel, die Migration als Methode zur Erweiterung und Anreicherung bestehender Services um mobile Anwendungsformen vorzustellen. Dabei geht es darum, Migration als Instrument zum methodisch-fundierte Ausgleich zwischen den Aspekten Erneuerung und Bewahrung bei der Gestaltung mobiler Anwendungen zu positionieren. Der thematische Fokus liegt auf der methodischen Unterstützung des Innovationsschrittes von stationären zu mobil nutzbaren Services. Hierzu gilt es, eingangs die grundlegenden Begriffe Mobilität und mobile Services zu charakterisieren.

1.1 Mobilität

Der in diesem Beitrag verwendete Mobilitätsbegriff beinhaltet zwei Aspekte: die räumlich-physische und die virtuelle Mobilität. Die räumlich-physische Mobilität bezeichnet die Bewegung einer Person von einem Ort zu einem anderen. Die virtuelle Mobilität kennzeichnet hingegen die Erreichbarkeit von Informationen (und Personen) an jedem Ort, wobei auf räumlich-physische Mobilität verzichtet werden kann [ZoKJ02]. Im Folgenden wird unter Mobilität die Flexibilisierung von Anwendungssystemen und/oder Personen durch eine räumliche Entkoppelung unter Verwendung von Informations- und Kommunikationstechnologien (IKT) verstanden. In diesem Zusammenhang gewinnen Ansätze des mobilen, ubiquitären und pervasiven Computings an Bedeutung [Saty02]. Mobile Computing wird als Nutzung mobiler elektronischer Kommunikationstechniken in Verbindung mit mobilen Endgeräten verstanden, während Ubiquitous Computing als permanente Interaktion eines Nutzers mit für ihn unsichtbaren Rechnersystemen beschrieben wird [Weis93]. Pervasive Computing vereint diese beiden Ansätze: Der Begriff beschreibt die Verschmelzung der Informationstechnologie (IT) mit der Umgebung, indem IT in alltägliche Gebrauchsgegenstände integriert und für den Nutzer quasi unsichtbar wird. Der Nutzer muss somit nicht mehr direkt mit der Technik interagieren und erkennt diese eventuell nicht mehr als Computertechnologie [HBBB03]. Alle drei Computing-Ausprägungen unterstützen die beiden Mobilitäts-Dimensionen. Die Mobilitätszustände können sich dabei sowohl auf Anwendungssysteme beziehen, in diesem Fall also auf Daten, Funktionen, Prozesse und Aufbauorganisation, als auch auf Personen, in deren Fall die persönliche Kommunikation

fokussiert wird. Der Schwerpunkt dieses Beitrags liegt auf Services, welche sich Ansätzen des Mobile Computing bedienen; diese werden nachfolgend eingegrenzt.

1.2 Mobile Services

Für die Anbieter von Services erwächst die Notwendigkeit, Strategien hinsichtlich ihres unternehmerischen Handelns kontinuierlich den sich wandelnden Umfeldbedingungen anzupassen. Durch das Internet ergaben sich seit den 1990ern neue Formen IKT-gestützter Services, so genannte e-Services. Diese beinhalten in der Regel einen Mehrwert für den Nutzer und treffen teils auf sehr hohe Akzeptanz. Dabei sind die Beispiele für eine gelungene Invention von e-Services in unternehmerische wie in private Bereiche vielgestaltig. Angefangen von e-Shops als alternativen Vertriebsweg, über e-CRM-Systeme zur Pflege der Kundenbeziehungen bis hin zu e-Procurement-Systemen bietet die integrative Nutzung entsprechender Services explizite, wirtschaftlich messbare Vorteile für Unternehmen. Services wie eBay, Amazon und die Kommunikation via E-Mail sind weiterhin kaum noch aus dem Alltag zahlreicher Personen wegzudenken. Eine signifikante Erweiterung entsprechender Services ist die mobile Nutzung bisher „stationär“, d.h. mit einem klassischen Internetzugang, nutzbarer e-Services [ChSk05]. In der Literatur existieren erste Definitionen zu m-Services. Zumeist wird nach verschiedenen Möglichkeiten der Kategorisierung, z.B. nach den involvierten Partnern oder dem Anlass der Nutzung differenziert [Salk04; Vesa05]. m-Services können aufgrund ihrer Nutzung in drei Kategorien eingeteilt werden: Person-to-Person Messaging (Zwischenmenschliche Kommunikation), Content Services (Inhaltliche Angebote) und Mobile Data Access (Mobiler Datenzugriff) [Vesa05]. In der Regel werden unter m-Services mit mobilen Endgeräten nutzbare Dienstleistungsangebote gefasst. Daran orientiert sich der hier verwendete Servicebegriff und lehnt sich somit an betriebswirtschaftlichen Definitionen an [ScGK03].

Bei der Entwicklung mobiler Services spielt die Nutzung erweiterter Möglichkeiten der Mobilität eine wesentliche Rolle. m-Services ermöglichen einerseits die Erschließung neuer Anwendungsfelder, andererseits können eine Reihe von Mehrwert-stiftenden Potenzialen genutzt werden. Als solche lassen sich inhärente Möglichkeiten zur Unterstützung der nachfolgend erläuterten Kriterien nennen [Buse02; Vesa05]. Diese werden zu Teilen bereits durch e-Services unterstützt, erreichen in mobilen Einsatzszenarien jedoch eine neue Qualität. Ein erstes wesentliches Kriterium stellt die Lokalisierbarkeit dar, welche Unterstützungsmöglichkeiten durch die exakte Bestimmung des Aufenthaltsorts der Anwender

mittels Integration von GPS oder über den Funkbereich offeriert. Location based Services (LBS) nutzen dieses Potenzial. Die ebenfalls gesteigerte Ortsunabhängigkeit stellt ein zweites Kriterium dar. Sie unterstützt die Kommunikation, das Abrufen von Informationen und das Versenden von Daten unabhängig vom Aufenthaltsort der Anwender. Die Erreichbarkeit als drittes Kriterium gewährleistet in mobilen Szenarien die jederzeitige Erreichbarkeit der Nutzer unabhängig vom Ort. Als viertes Kriterium realisiert die Personalisierung das Beziehen von individualisierten Leistungen. Dieses Kriterium wird in mobilen Anwendungen durch die eindeutige Identifikation mittels der integrierten SIM-Karte und durch die persönliche Rufnummer gestärkt. Das fünfte Kriterium, die verbesserte Kontextspezifität, ermöglicht das Beziehen von Angeboten passend zur jeweiligen Umgebung oder Tätigkeit. Die erhöhte Bequemlichkeit als sechstes Kriterium realisiert im mobilen Kontext die einfache oder vereinfachte Bedienung und den schnelleren Zugang als bei herkömmlichen stationären Systemen. Die Kostengünstigkeit als siebtes Kriterium dient der potenziellen Zuschneidung der Services auf ihren Verwendungszweck. Als achtens Kriterium gewährleistet die Sicherheit eine eindeutige Identifikation des Besitzers, in diesem Fall insbesondere durch die SIM-Karte und bietet somit eine höhere Sicherheit als über stationäre PCs.

Trotz dieser ansehnlichen Potenziale besteht eine Diskrepanz zwischen der Nutzung von e- und von m-Services [ChSk05]. Diese lässt sich in Teilen durch das Umfeld erklären, welches dadurch gekennzeichnet ist, dass m-Services derzeit von Telekommunikations- sowie teilweise von e-Service-Anbietern bereitgestellt werden. Teils werden erfolgreiche e-Services ergänzend als m-Services angeboten, teilweise werden singuläre m-Services neu geschaffen. Zumeist ist die Situation dadurch gekennzeichnet, dass die mobilisierten e-Services nicht in ausreichendem Maße die genannten erweiterten Möglichkeiten von m-Services berücksichtigen oder nicht den Gewohnheiten und Anforderungen der Nutzer entsprechen. An dieser Stelle setzen Bestrebungen zur Migration von e- zu m-Services an. Sie unterstützen eine zielgerichtete, mehrwertorientierte Fortentwicklung existierender Services und werden im Folgenden skizziert.

2 Migrationsbegriff und -motivation

Veränderungsprozesse in Wirtschaft und Gesellschaft, die durch technologischen Wandel und/oder durch veränderte Bedürfnisse sowie strategische Ausrichtungen (Kunden-, Prozess-,

Wissensorientierung, Integration und Vernetzung) ausgelöst werden, erfordern ein spezielles Management von Wandlungsprozessen [Stae98]. Erfolgreiche Wandlungsprozesse sind nur möglich, wenn gesellschaftliche und technische Prozesse sowie Systeme aufeinander abgestimmt sind. Dies gilt aufgrund der dargestellten Ausgangssituation und den Defiziten derzeitiger Realisierungen auch bei der Entwicklung mobiler Services.

Im Folgenden gilt es den Begriff der Migration zu erläutern. Dieser unterliegt einer Bedeutungsvielfalt und wird in unterschiedlichen Zusammenhängen, meistens unspezifiziert, verwendet. In der Informatik bezieht er sich beispielsweise auf die Ablösung oder das Upgrade von einem Anwendungs- oder Software-System auf ein potenziell besseres [Simo92]. Es finden sich konzeptionelle Überlegungen zum Thema des Migrationsmanagements als Teilgebiet der Anwendungssystem (AWS)-Entwicklung und -Implementierung aus wirtschaftsinformatischer Perspektive: Dabei wird Migration als Managementaufgabe betrachtet, bei der neben technischen auch organisatorische Fragestellungen eine wesentliche Rolle spielen. Dömer definiert wie folgt: „Migration ist definiert als der Prozess der Umstellung eines computergestützten Informationssystems (IS) (=Altsystem/Quellsystem) auf ein anderes IS (=Neusystem/Zielsystem) im Rahmen einer eigenen Projektorganisation, wobei sich Alt- und Neusystem signifikant bezüglich der Informationsinfrastruktur i.e.S. (technische Komponenten: Hardware und/oder Software) unterscheiden, gleichzeitig aber elementare fachliche und/oder technische Komponenten (Hardware und/oder Software in beliebigen Repräsentationsformen und/oder Daten des Altsystems) übernommen werden.“ [Döme98] Der Begriff kennzeichnet damit den Übergang von bestehenden AWS zu Systemen mit erweiterten oder neuen Funktionalitäten bzw. mit innovativen Nutzungsmöglichkeiten oder veränderten Wirkungsgraden. Migration verknüpft eingeführte mit neuartigen, ggf. innovativen Systemkomponenten und Technologien zu einer funktional verbesserten und ökonomisch ausbalancierten Gesamtlösung. Diesem Artikel liegt ein einerseits weiter gefasstes und andererseits fokussierteres Verständnis zugrunde: Migration wird als anbieterseitig geplanter und kontrollierter Zustandswechsel von AWS und Kommunikationsformen definiert, der zu veränderten rechtlich, ökonomisch, technologisch, ökologisch und sozial bewerteten Mobilitätszuständen führt [BFKS04].

Verwandte Disziplinen des Migrationsmanagements sind das Innovations-, das Technologiemanagement, das Management des „Organisational and Technological Change“

sowie das Komplexitätsmanagement. Im Gegensatz zum Innovationsmanagement, welches zumindest in Teilbereichen auf „radical change“ setzt, ist Migration eine Methode der Erneuerung, die zugleich auch auf Bewahrung des Bewährten zielt. Ansätze des verwandten Changemanagements fokussieren hingegen, speziell im Kontext der IKT-Unterstützung ökonomischer und administrativer Prozesse, vorwiegend auf die Extreme „Wartung“ oder „Totalinnovation“. [Stae98] Wartung verzichtet auf die Implementierung notwendiger innovativer Elemente, Totalinnovation entwertet getätigte Investitionen, speziell solche in Qualifizierung und Prozessorganisation [Döme98]. Die nachfolgende Abbildung 1 verdeutlicht diese Differenzierung.

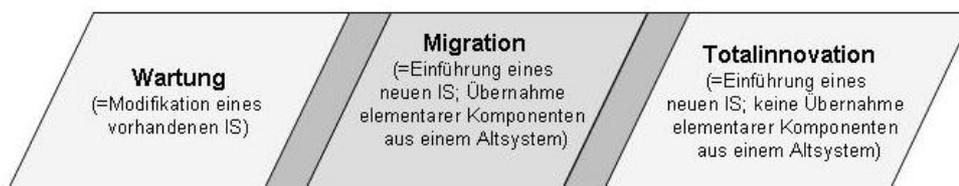


Abb. 1: Migrationsverständnis [Döme98]

Migration beschreibt somit einen dritten Weg zur Bewältigung geplanter technologischer und organisatorischer Wandlungsprozesse. Dies deutet einen wichtigen Aspekt der Migration an: Durch Migrationen können sich (partielle) Ablösungen von existierenden Anwendungssystemen durch Neuentwicklungen ergeben, es können sich jedoch gleichfalls Erweiterungen ergeben, welche zu großen Teilen auf Elementen des vorherigen Systems basieren.

3 Framework zur Migration von e- zu m-Services

Leitend für die Gestaltung von Migrationsprozessen ist das Ziel, zu migrierende Services ressourcenschonend und nutzerfreundlich zu implementieren. Dies beinhaltet, dass Investitionen in Hardware, Software und Netware möglichst weiter genutzt, eher ergänzt denn abgeschrieben werden [WWMB00]. Ebenso bedeutsam und ökonomisch gar brisanter ist es, Investitionen in Prozesse, Organisation und in das Know-how der Nutzer, weitgehend zu erhalten oder „schonend“ anzupassen. Dies kann den Erfolg des Übergangs nachhaltig stabilisieren, den Übergangsprozess beschleunigen und damit letztlich den wirtschaftlichen

Erfolg durch die Schaffung von Mehrwerten sichern. Dies verdeutlicht, dass bei der Ausgestaltung von Migrationen primär technisch ausgerichtete Betrachtungen zu kurz greifen [BFKS04]. Migrationen berühren gleichfalls organisatorische und wirtschaftliche Fragestellungen. Diesen Überlegungen entsprechend müssen Migrationen auf verschiedenen Ebenen betrachtet und durchgeführt werden. Abbildung 2 stellt die Ebenen der Migration in ihrer Abhängigkeit zueinander dar.

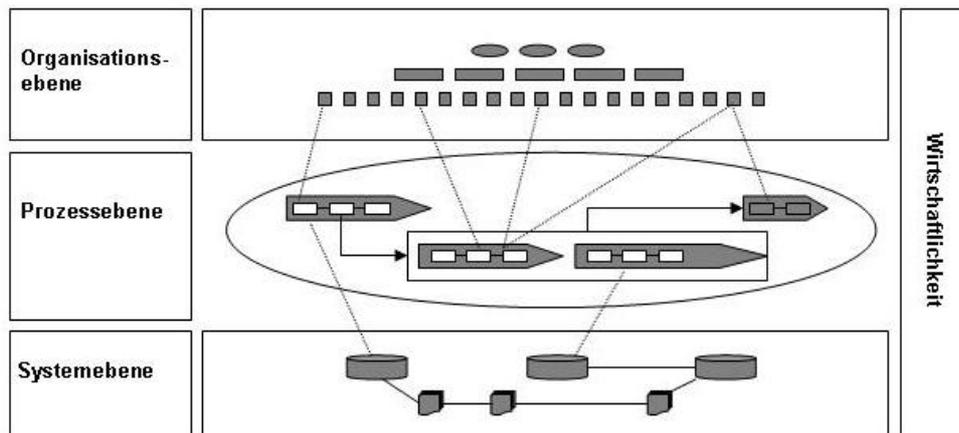


Abb. 2: Framework der Migration

Die Systemebene bildet die erste Ebene. In dieser existieren unterschiedliche Arten von Migrationen wie z.B. Daten-, Code-, System- oder Fachkonzeptmigrationen sowie Ansätze der Portierung [Döme98]. Für dieses Framework von besonderem Interesse sind komplexe Anwendungssystemmigrationen, die in der Regel von speziellen Migrationsarten begleitet werden. Neben den Systemen und Technologien werden in einer erweiterten Sicht auf die Migrationsthematik aus wirtschaftsinformatischer Sicht insbesondere Organisationsmodelle und Prozesse betrachtet, die bei der Migration von Systemen berücksichtigt sowie ggf. angepasst werden müssen. Hierbei sind zwei unterschiedliche Ansätze denkbar: Einerseits können neue Technologien eine effektivere und im Idealfall effizientere Organisation ermöglichen, andererseits können veränderte Rahmenbedingungen, die sich auf die Organisation oder die Prozesse auswirken, neue oder erweiterte Anwendungssysteme erfordern. Daher sind in ganzheitliche Migrationskonzepte nicht nur Technologien, sondern auch Kompetenzen, Organisations- und Geschäftsmodelle einzubeziehen. Die drei genannten Ebenen werden unter besonderer Berücksichtigung der Wirtschaftlichkeit betrachtet. In diesem Rahmen bewahrt Migration ökonomische Entwicklungen, die gesellschaftlichen Bedürfnisse

entsprechen, behält erfolgreiche Geschäftsprozesse und -modelle bei und bietet zugleich Chancen für deren Fortentwicklung. Nachfolgend werden die verschiedenen Ebenen detailliert erläutert. Die vorhergehend aufgezeigten Aspekte verdeutlichen, dass Entscheidungen für Migrationen sowie für konkrete Migrationsprojekte aus einer Gesamtperspektive getroffen werden sollten [Bues93; Simo92; Walk93]. Darauf aufbauend stellt sich die Frage, in welcher Form die dargestellten Aspekte bei der Gestaltung mobiler Services einsetzbar sind und wie die vorgestellten Elemente des Frameworks bei der Migration von e- zu m-Services wirken.

3.1.1 Organisationsebene

Betrachtet man die Migration auf Organisationsebene, ergeben sich zwei Bereiche, die relevant erscheinen: Die organisatorischen Aspekte der Migration sowie ihre Durchführung als Projekt. Bei der Untersuchung organisatorischer Aspekte kann einerseits eine statische Sicht und andererseits eine dynamische Sicht auf die Organisation gelegt werden. Dabei bezieht sich die statische Sicht auf die Organisation als solche und nicht darauf, was diese tut. Bei einer solchen Betrachtung wird eine systemorientierte Sichtweise eingenommen. Das System der Organisation gliedert sich in die Subsysteme Strategie, Technologie, Personal, Aufbau und Prozesse sowie Managementprozesse [Döme98]. Dabei wird die Entscheidung für eine Migration auf Managementebene positioniert und von der Strategie im Unternehmen und den Ressourcen abhängig gemacht [Berl04]. Legt man eine dynamische Sichtweise auf die Organisation stellt sich die Informationstechnologie als Impulsgeber und treibende Kraft des organisatorischen Wandels dar. Aus der Kombination der Sichtweisen resultieren Anlässe für die Migration, die nachfolgend klassifiziert werden.

Als zentrales Motiv zur Migration lässt sich der angestrebte Zustandswechsel identifizieren. Diesem können unterschiedliche Anlässe zugrunde liegen. Ziel ist jedoch stets ein verbesserter Gesamtzustand. Basierend auf den Anlässen kann eine Klassifikation anhand von zwei Dimensionen erfolgen. Dabei wird einerseits zwischen externen und internen Faktoren unterschieden, andererseits kann die Migration durch Problemdruck aufgrund funktionaler und technologischer Faktoren angestoßen werden. Migrationen können durch interne funktionale Ursachen angestoßen werden, vor allem da, wo Altsysteme die neuen geforderten Funktionalitäten nicht mehr erfüllen. Bei einer schlechten Wartungsqualität des Altsystems oder Kapazitätsüberlastungen wirken hingegen interne technologische Ursachen als Anlass zur Migration [Walk93]. Externe funktionale Ursachen begründen sich vor allem im Wettbewerb.

Hier wird eine Migration erforderlich, um dadurch wettbewerbsfähig zu bleiben. Externe technologische Ursachen dagegen basieren auf im Unternehmen eingesetzten externen Produkten, insbesondere wenn diese nicht weiter unterstützt oder vom Markt genommen werden. Bei diesen Anlässen sind vor allem Kosteneinsparpotentiale zu berücksichtigen, die mit dem Einsatz neuer Technologien zusammenhängen [Döme98]. Durch die Kombination der beiden Dimensionen entstehen die in der nachfolgenden Abbildung 3 dargestellten Impulsgruppen.

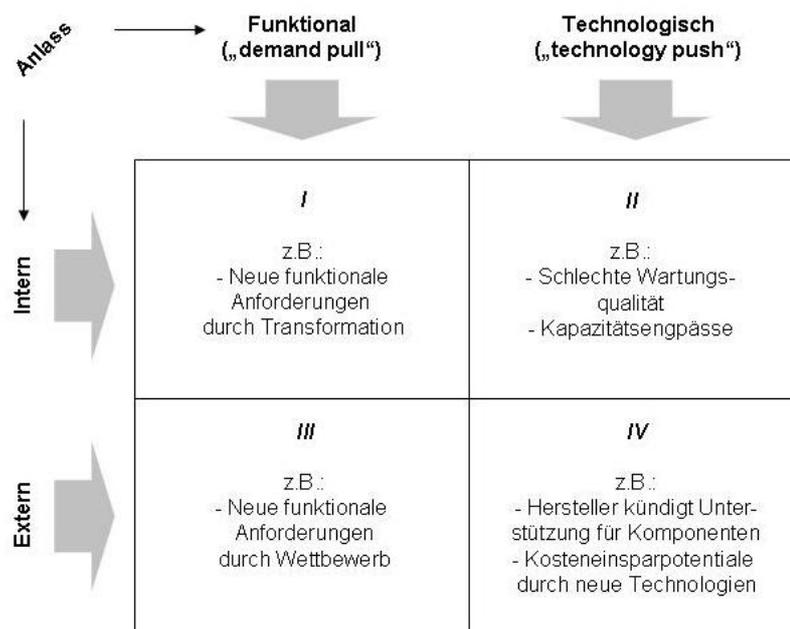


Abb. 3: Impulsgruppen der Migration [Döme98]

Migration sollte auf Organisationsebene weiterhin als Projekt betrachtet werden. Charakteristisch für Migrationen ist die große Komplexität im Vergleich zur Wartung, die unter anderem durch die Vielzahl von Beziehungen der aus dem Altsystem übernommenen Elemente induziert wird. In diesem Zusammenhang üben folgende Aktivitäten Einfluss auf das Gelingen von Migrationsprojekten aus: Initiierungsmanagement, Analyse der Rahmenbedingungen, fachlich-inhaltliches Projektmanagement und anfängliche Segmentierung des Altsystems sowie Bestimmung der Vorgehensmodelle, Festlegung der Migrationstypen, Qualitätssicherung, Projektkoordination und Projektmarketing in den darauf folgenden Stufen [Döme98].

Auch im Hinblick auf die Migration von e- zu m-Services ist der Bezug zum Migrationsanlass zu analysieren. Dabei tendiert die Einführung mobiler Services zumeist in zwei grundsätzliche

Richtungen zur Optimierung des Gesamtzustandes. Die beiden Extreme eines verbesserten Gesamtzustandes können dabei als Erweiterung/Ergänzung oder als radikaler Umbruch charakterisiert werden. Bei einer Entscheidung zugunsten der Erweiterung/Ergänzung geht es letztendlich um den Erhalt bewährter Elemente von e-Services und deren Nutzung in neuen, mobilen Kontexten; bei einem radikalen Umbruch ergibt sich ein verbesserter Gesamtzustand ohne die Nutzung bestehender Komponenten für das Angebot von e-Services. Bei einer Entscheidung zugunsten einer der Varianten und bei der Planung und Kontrolle von Zustandswechseln steht demnach neben der angestrebten Ökonomie die adäquate Nutzung existierender Systeme und deren zielgerichtete Anpassung und Erweiterung im Vordergrund.

3.1.2 Prozessebene

Die Anlässe der Migration und der organisatorische Wandel im Unternehmen können die Ablauforganisation des Unternehmens beeinflussen. Vor allem Wettbewerbsdruck und notwendige Effizienzsteigerung können Unternehmen zwingen, ihre Geschäftsabläufe zu optimieren. Auf Prozessebene kann die Migration somit vor allem unter Berücksichtigung der Veränderung der Prozesse im Unternehmen betrachtet werden. Dabei spielt die Nutzung moderner IKT eine bedeutende Rolle, mit deren Hilfe es möglich ist, Geschäftsprozesse neu zu gestalten [Berl04]. In diesem Zusammenhang beabsichtigen Unternehmen, die Potenziale neuer Technologien mit den Erfordernissen ihrer Geschäftsprozesse im Rahmen eines Business Process Reengineering (BPR) in Einklang zu bringen. Dabei zielt BPR auf die Umgestaltung der Geschäftsprozesse mit dem Ziel, substantielle Verbesserungen in der Aufbau- und Ablauforganisation des Unternehmens zu erreichen und im Hinblick auf die herrschenden Marktbedürfnisse zu optimieren [Snee99].

Auf Prozessebene offeriert die Migration von e- zu m-Services vielfältigen Potenzialen und stellt Unternehmen gleichzeitig vor neue Herausforderungen. So ist es notwendig, vorhandene oder potenzielle Mobilität im Rahmen der Wertschöpfungskette zu identifizieren und nutzbar zu machen [Berl04]. Dabei besteht das Potenzial bei einer Migration von e- zu m-Services nicht in der Ausstattung alter Prozesse mit neuen AWS oder Endgeräten, sondern in der Möglichkeit, neue Arbeitsweisen aufzubauen und Prozesse neu zu gestalten, vor allem um Medienbrüchen zu vermeiden. Die nachfolgende Abbildung stellt eine exemplarische Service-Prozesskette ohne eine systematische mobile Unterstützung einer Prozesskette nach einer systematischen mobilen Unterstützung gegenüber.

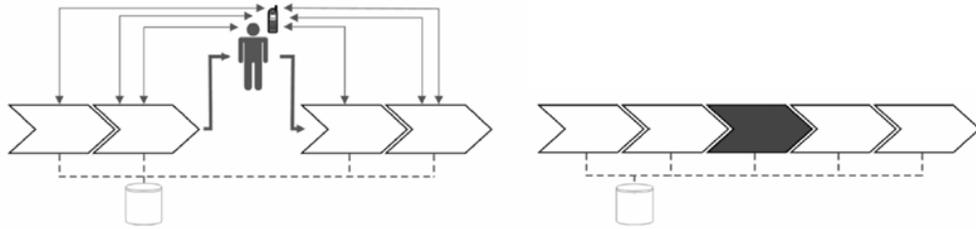


Abb. 4: Service-Prozesskette vor und nach einer mobilen Unterstützung [KhPW03]

Entscheidend für Migrationen auf Prozessebene sind die Mehrwerte, die mit ihrer Nutzung einhergehen. So lassen sich beispielsweise räumlich verteilt ablaufende Prozessschritte durch den Einsatz mobiler Technologien nahtlos in die unternehmerische Wertschöpfung integrieren. Bei einer zielgerichteten Steuerung der Geschäftsprozesse werden somit räumliche Barrieren überwunden und Informationen zu jeder Zeit verfügbar gemacht. Insbesondere kann durch die gezielte Nutzung bestehender, durch e-Systeme gestützter unternehmensinterner und -externer Prozesse eine effiziente und prozessadäquate Integration mobiler Services realisiert werden.

3.1.3 Systemebene

Auf Systemebene übernehmen IKT beim organisatorischen Wandel eine herausragende Bedeutung. Ihnen werden unterschiedliche Rollen zugewiesen. Sie fungieren als Enabler (Auslöser), als Faciliator (Erleichterer), als Supporter (Unterstützer) oder als Implementator (Umsetzer) [Geie99]. Wie bereits ausgeführt, existieren auf Systemebene unterschiedliche Arten von Migrationen, wie z.B. Code-, System- oder Fachkonzeptmigrationen. In Bezug auf die Migration von e- zu m-Services können Migrationen auf zwei elementaren Ebenen erfolgen: Während Softwaremigrationen in diesem Rahmen softwaretechnische Einführungen eines neuen Systems und Übernahmen elementarer Teile aus Altsystemen beschreiben, sind Datenmigrationen einmalige, auf konkrete betriebliche Anwendungsarchitekturen hin zu optimierende Prozesse.

Auf Systemebene ist zunächst die effektive Nutzung bestehender Datenbestände bei der systematischen Migration von e- zu m-Services zu nennen. Für die Datenmigration liegen bewährte Ablaufbeschreibungen und Instrumente vor [NeSe93]. Die Verantwortung für die Qualität und Vollständigkeit der zu migrierenden Datenbestände und das mit der Umstellung produktiver Systeme verbundene besondere Risiko verlangt es, den Prozess der Datenmigration im Sinne eines Projektes zu organisieren. Dies bedeutet die Festlegung von

Verantwortlichkeiten, die Bereitstellung kompetenter personeller sowie hinreichend dimensionierter sächlicher Ressourcen und eine sorgfältige Planung, Abwicklung und Kontrolle des Migrationsprojektes [Meie97]. Bezüglich softwareseitiger Migrationen können sich beim Übergang von e- zu m-Services positive Auswirkungen ergeben. Dabei ist im Hinblick auf mobile Services darauf hinzuweisen, dass diese zu großen Teilen auf Softwaresystemen oder -komponenten basieren. Hierauf aufbauend wird mit Migration die Weiterentwicklung und Pflege von Software bezeichnet, mit dem Ziel, diese in einer geänderten Umgebung nutzen zu können [IEEE91]. Ein wesentliches Ziel dabei ist die teilweise Erhaltung vorhandener technischer Komponenten, die eng mit der Vermeidung unnötiger Neuentwicklungen im Vordergrund steht. Eine Migration informationstechnischer Systeme erlaubt es, Innovationen nachzuvollziehen und die Produktivität zu steigern, ohne die Notwendigkeit, Nutzer in ein gänzlich neues System einzugewöhnen. Dabei lassen sich die fünf grundlegenden Migrationstypen zwei Bereichen zuordnen. Während der Investitionsschutzbereich die Alteration, Konversion und Kapselung umfasst und damit der Überführung von Komponenten bestehender Systeme dient, stellen die Nutzung von Standardlösungen und die Individualerstellung neuer Komponenten einen Innovationsbereich dar. Kombiniert ergeben sich durch diese Migrationsausprägungen unter Einbezug existierender e-Services innovative mobile Services.

3.1.4 Wirtschaftlichkeit

Bei Migrationsvorhaben muss zusätzlich zur Betrachtung der beschriebenen Ebenen eine explizite Betrachtung der Ökonomität erfolgen. Relevant bei entsprechenden Analysen sind Betrachtungen des Investitionsschutzes, der ökonomischen Nutzenpotenziale/Mehrwerte und der Kosten. Prinzipiell sollten Migrationen mit ökonomisch positiven Ergebnissen realisiert werden. Dies kann dadurch geschehen, dass der unternehmerische Wert eines Systems bewahrt wird und getätigte Investitionen nicht oder nur zu Teilen durch (redundante) Entwicklungen substituiert werden. Jedoch sind an dieser Stelle Analysen der Aufwendungen zwingend notwendig, welche bei Wiederverwendungen existierender Komponenten einerseits und bei Neuentwicklungen andererseits anfallen. In diesem Kontext sind zudem Risiken der Migration in ein Verhältnis mit denen einer Totalinnovation zu setzen. Darüber hinaus lassen sich positive ökonomische Effekte durch Systeme mit deutlich wahrnehmbaren Mehrwerten im Vergleich zu bislang genutzten, laufenden Systemen realisieren. Zusätzlich realisierbare Erlösströme oder verbesserte Prozessstrukturen sind prominente Beispiele hierfür. Weiterhin kann eine

Reduzierung der Betriebs- und Wartungskosten eines Altsystems als bestimmendes Motiv für Migrationen dienen. Dabei sind in diesem Fall nicht nur Kostenanalysen vorzunehmen, vielmehr müssen, einem geschäftsmodell-orientiertem Vorgehen folgend, zukunftsgerichtete Aspekte der (Weiter-) Entwicklung der Märkte wie auch die unternehmerische Wertschöpfung in die Betrachtungen einbezogen werden. Sneed stellt ein Modell (Abb. 5) vor, welches anhand der Kriterien „Technische Qualität“ und „Betriebswirtschaftliche Bedeutung“ Altsysteme als Kandidaten zur Entsorgung, Wartung, Neuentwicklung oder Migration klassifiziert [Snee99].

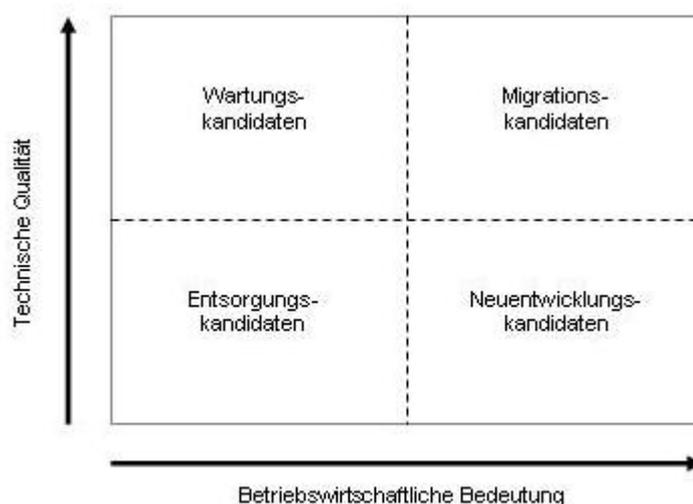


Abb. 5: Positionierung des Altsystems [Snee99]

Im Bereich der Wirtschaftlichkeit liegen umfassende Potenziale der Migration von e- zu m-Services. So können migrierte Services eine Reihe ökonomischer Nutzenpotenziale eröffnen, wie Kosteneinsparungs-, Zeiteinsparungs- und Qualitätspotenziale. Beispielsweise können Kosteneinsparungspotenziale durch bessere Ressourcennutzung, Mehrfachnutzung und höhere Preistransparenz erzielt werden. Ferner können mit gut aufeinander abgestimmten Services für den Endkunden eine Reihe von Zeiteinsparungspotenzialen durch den Einsatz intelligenter Optimierungssysteme erzielt werden. Auch sind vielfältige Qualitätspotenziale realisierbar, hierzu zählen beispielhaft Verbesserungen der Leistungsqualität bei der Nutzung akzeptierter Abrechnungssysteme.

Weitere Nutzenpotenziale für Unternehmen liegen in der Erschließung lukrativer Märkte durch das Angebot mobiler Services sowie in der Erweiterung ihrer Erlösmodelle. Unternehmen, die e-Services anbieten, welche derzeit (vor allem im WWW) primär werbefinanziert sind, können ihre Leistungen in mobilen Kontexten zusätzlich gegen Bezahlung anbieten [ChSk05]. Eine

wesentliche Voraussetzung für die breitenwirksame Durchsetzung innovativer m-Services ist, dass diese einen signifikanten Nutzen sowohl für Anbieter als auch für Nutzer schaffen. Der zu realisierende Nutzen muss für beide Parteien materiell und immateriell ausgeprägt sein.

Will ein Unternehmen etablierte e-Services auch als m-Services anbieten, ist es sinnvoll, verschiedene Services-Anbieter miteinander zu vernetzen. Dies verdeutlicht folgendes Zitat: „Für die Hersteller, Betreiber und Service Provider im mobilen Internet liegt also der Schlüssel zum Erfolg in der Fähigkeit, bei der Entwicklung und Bereitstellung von leicht nutzbaren Anwendungen für den Massenmarkt zusammenzuarbeiten. Die Gewinner des Marktes werden jene sein, denen es gelingt, Allianzen zwischen Content-Anbietern und Technologiefirmen zu formen, um für den Nutzer an seinem Aufenthaltsort wertvolle und nützliche Informationen bereitzustellen.“ [Stau01] Die Vernetzung einer Vielzahl solcher Service-Anbieter ist zunächst von hohen technischen und organisatorischen Voraussetzungen abhängig. So muss aus technischer Sicht eine Vielzahl von Systemen miteinander vernetzt werden und diese müssen zum Teil um neue m-Service-Schnittstellen und Standards erweitert werden. Wird ferner davon ausgegangen, dass Unternehmen in den letzten Jahren sehr viel in zugrunde liegende Anwendungssysteme investiert haben, muss darauf geachtet werden, dass dieser Entwicklungsprozess durch die Migration bestehender Systeme hin zu mobilen Systemen möglichst investitionsschonend gestaltet wird. Hierzu bedarf es einer methodischen Vorgehensweise, wie sie die Migration offeriert.

4 Vorteile der Migration als Methode

Migrationsstrategien zielen im dargestellten Kontext der Migration von e- zu m-Services darauf, Erhaltenswertes zu bewahren, indem erprobte Komponenten/Verfahren wieder verwendet und adaptiert werden. Sie dürfen nicht nur auf Techniksysteme (Anwendungen, Architekturen, Konzepte, Entwürfe, Schnittstellen) bezogen sein, sondern müssen sich auch auf organisatorische Aufbaustrukturen, personale Qualifikationen (Humankapital), Prozesse und Funktionen sowie auf Konfliktpräventionsstrategien und Konfliktlösungen erstrecken. Die Migration erfordert somit eine balancierte Gestaltung einer Vielzahl von Systemelementen. Eine Fokussierung auf einzelne Dimensionen ist ungeeignet, das Bündel positiver Migrationseffekte zu realisieren. Migration kann dabei

- die Akzeptanz und Motivation der vom Systemwandel Betroffenen stärken. Dadurch bewahren Wissen und Erfahrungen weitgehend ihren Wert.
- die Systementwicklung und -implementierung beschleunigen. Eine Teilmenge von Systemkomponenten wird nicht neu entwickelt, sondern „lediglich“ integriert.
- die Stabilität der migrierten Anwendungsunterstützung erhöhen und das Risiko des Scheiterns verringern. Die Nutzung durch die Betroffenen kann partiell auf gesicherten Erfahrungen aufsetzen und somit zur Vermeidung von Fehlverhalten beitragen.
- durch Wiederverwendung von vorhandenen Qualifikationen und Kompetenzen, von Strategien der Fehlervermeidung und von Prozesswissen die Chance bieten, Kosten zu senken und die Leistungsqualität zu steigern. Bestehende Geschäftsprozesse und -logiken können dabei zielgerichtet weiterverwendet werden.

Diese Migrationseffekte sind nicht nur bei Systemnutzern (Mitarbeitern, Kunden, Lieferanten etc.), sondern auch bei Systementwicklern (dank Mehrfach- und Wiederverwendung) und Systemintegratoren (z.B. bei internen und externen Vernetzungen von Anwendungssystemen und Prozessen) zu erwarten. In Bezug auf die Etablierung mobiler Services kann ein solches Vorgehen die Nutzung derselben durch das Aufgreifen bestehender, akzeptierter Systeme fördern und gleichfalls den Anbietern einen gesicherten Weg zur Realisierung offerieren.

Zusammenfassend ist zu konstatieren, dass die Erweiterung von etablierten, akzeptierten e-Services des täglichen Privat- und Berufslebens um mobile Komponenten Nutzer wie Anbieter vor massive Herausforderungen stellt. Diesen Wandel grundlegender Kommunikations- und Interaktionsformen gilt es seitens der Anbieter durch angemessene Planungs- und Kontrollmechanismen sowie -instrumente und unter besonderer Berücksichtigung der Nutzeranforderungen zu optimieren. In diesem Zusammenhang gelten sowohl technische und organisatorische als auch soziale Einflussfaktoren als beeinflussbare, aber auch beeinflussende Variablen im Prozess des Wandels. Ein umfassendes Migrationskonzept zu Gestaltung des Übergangs von e- zu m-Services ist entsprechend der Komplexität des Themas unabdingbar: einerseits, um vorhandene Investitionen zu schonen, andererseits, um bewährte Anwendungssysteme einer weiteren Nutzung zuzuführen. Ferner lassen sich somit Änderungen der zugrunde liegenden Anwendungssysteme und der dahinter verborgenen Gewohnheiten der Konsumenten, aber auch der Geschäftsprozesse auf Anbieterseite, prognostizieren.

Die dargestellten Ansätze unterstreichen, wie verschiedenartig Migrationsansätze je nach gewählter Perspektive ausgestaltet sein können. Dennoch folgen sie gemeinsamen Zielsetzungen und dienen kombiniert dem Aufbau von Migrationskompetenz. Dieser Aufbau und die Entwicklung von Migrationskompetenz bedeutet mehr als die Erlangung der Eigenschaften zur fachlichen und technischen Übertragung von Daten in neue Systeme. Eine Herausforderung besteht darin, bei der parallelen Anwendung verschiedener Ansätze Wechselwirkungen aufzudecken. Deren wechselseitige Betrachtung beherbergt Potenziale, die teils auch widerstreitenden Interessen in einem bestmöglichen Gesamtausgleich zu harmonisieren und so zu einer interdisziplinären Optimallösung zu gelangen. Dabei sind in die Bewertung von Migrationsprozessen mögliche Effekte einer Mehrfachverwendung von Konzepten, Infrastrukturen, Anwendungssystemen und (Medien-) Kompetenzen einzubeziehen. Die Ökonomität eines solchen Prozesses erscheint betrachtenswert und erfolgskritisch.

Insgesamt verdeutlicht dieser Beitrag, dass e-Services durch Migrationen zielgerichtet in m-Services überführt werden können. Die Ergebnisse lassen zudem erwarten, dass Migration als strategiegetriebener kontinuierlicher Verbesserungsprozess der IKT-Unterstützung in Wirtschaft, öffentlicher Verwaltung und im privaten Bereich zu einer Alternative für die Gestaltung technologischer und organisatorischer Innovationen avanciert. Die Anwendung verschiedener Ansätze zur Optimierung der Migrationsziele und der verbundenen Wandlungsprozesse ermöglicht erfolgreiche Überführungen bestehender in neuartige Services.

Literaturverzeichnis

- [Berl04] Berlecon Research: Prozesse optimieren mit Mobile Solutions. Basis Report. Berlin 2004. <http://www.berlecon.de>. Abruf am 2005-12-17.
- [BFKS04] Bohl, Oliver; Frankfurth, Angela; Kuhlenkamp, Andreas; Schellhase, Jörg; Winand, Udo: Migrationskompetenz im Kontext der Komplexität mobiler Systeme. In: Engeli, M.; Meißner, K. (Hrsg.): Virtuelle Organisationen und Neue Medien 2004. Eul, Köln/Lohmar 2004, S. 115-126.
- [Bues93] Bues, Manfred: Migration zu offenen Systemen. In: HMD - Theorie und Praxis der Wirtschaftsinformatik. Heft 172. dpunkt, Heidelberg 1993.

- [Buse02] Buse, Stephan: Der mobile Erfolg – Ergebnisse einer empirischen Untersuchung in ausgewählten Branchen. In: Keuper, F. (Hrsg.): Electronic-Business und Mobile- Business. Wiesbaden 2002, S. 91-116.
- [ChSk05] Chen, Lei-da; Skelton, Gordon: Mobile Commerce Application Development. Cybertech, Hershey, 2005.
- [Döme98] Dömer, Fabian: Migration von Informationssystemen – Erfolgsfaktoren für das Management. Deutscher Universitäts-Verlag, Wiesbaden 1998.
- [Geie99] Geier, Christoph: Optimierung der Informationstechnologie bei BPR-Projekten. Verlag Deutscher Universitätsverlag, Wiesbaden 1999.
- [HBBB03] Hilty, Lorenz M.; Behrendt, Siegrid; Binswanger, Mathias; Bruinink, Arne; Erdmann, Lorenz; Fröhlich, Jürg; Köhler, Andreas; Kuster, Niels; Som, Claudia; Würtenberger, Felix: Das Vorsorgeprinzip in der Informationsgesellschaft – Auswirkungen des Pervasive Computing auf Gesundheit und Umwelt. Studie des Zentrums für Technologiefolgen-Abschätzung. TA-SWISS, Bern 2003.
- [IEEE91] Institute of Electrical and Electronics Engineers: IEEE Standard Computer Dictionary. IEEE Standard Computer Glossaries. New York 1991.
- [KhPW03] Khodawandi, Darius; Pousttchi, Key; Winnewisser, Christian: Mobile Technologie braucht neue Geschäftsprozesse. Universität Augsburg. <http://www.wi-mobile.de>, Abruf am 2006-06-19.
- [Meie97] Meier, Andreas: Datenbankmigration – Wege aus dem Datenchaos. In: HMD Heft 194. dpunkt.verlag, Heidelberg 1997, S. 24-36.
- [NeSe93] Netze, Jost; Seelos, Hans-Jürgen: Szenarien und Strategien der Datenmigration. In: Wirtschaftsinformatik 35 (4/1993), S. 320-324.
- [Salk04] Salkintzis, Apostolis K.: The Evolution toward the Mobile Internet. In: Salkintzis, Apostolis K. (Ed.): Mobile Internet – Enabling Technologies and Services, CRC Press, Boca Raton, 2004.

- [Saty02] Satyanarayanan, Mahadev: A Catalyst for Mobile and Ubiquitous Computing. In: Pervasive Computing, January-March 2002, S. 2-5.
- [ScGK03] Scheer, August-Wilhelm; Griebel, Oliver; Klein, Ralf: Modellbasiertes Dienstleistungsmanagement. In: Bullinger, Hans-Jörg; Scheer, August-Wilhelm (Hrsg.): Service Engineering. Berlin, 2003, S. 19-49.
- [Simo92] Simon, Alan R.: Systems Migration. Van Nostrand Reinhold, New York 1992.
- [Snee99] Sneed, Harry M.: Objektorientierte Softwaremigration. Addison Wesley Longman Verlag, Bonn 1999.
- [Stae98] Staehle, Wolfgang H.: Management. Vahlen, München 1998.
- [Stau01] Staudt, Erwin: Die mobile Gesellschaft. In: Buhl, Hans Ulrich; Huther, Andreas; Reitwiesner, Bernd (Hrsg.): Information age economy – 5. Internationale Tagung Wirtschaftsinformatik 2001. Physica, Heidelberg 2001, S. 15-28.
- [Tele04] Telegance: Mobile Datendienste – Der Silberstreif am Mobilfunkhimmel. <http://www.telegance.de>, Abruf am 2004-05-07.
- [Vesa05] Vesa, Jarkko: Mobile Services in the Networked Economy. IRM, Hershey 2005.
- [Walk93] Walk, Christian: Aufgaben und Methoden eines Migrationszentrums. In: Wirtschaftsinformatik. Vieweg, Wiesbaden Heft 4/1993.
- [Weis93] Weiser, Mark: Some Computer Science Issues in Ubiquitous Computing. <http://www.ubiq.com/hypertext/weiser/UbiCACM.html>, 1993-03-23, Abruf am 2006-06-07.
- [WWMB00] Westarp, Falk von; Weitzel, Tim; Margaritis, Kosmas; Buxmann, Peter; König, Wolfgang: Entscheidungen über betriebliche Standardsoftware – Die Migration des SAP-Systems bei der LSG. SFB 403 Arbeitsbericht (00-08), Frankfurt 2000.
- [ZoKJ02] Zoche, Peter; Kimpeler, Simone; Joepgen, Markus: Virtuelle Mobilität: Ein Phänomen mit physischen Konsequenzen? Springer, Berlin 2002, S. 1-10.

Automotive Service Engineering

Systematische Entwicklung personalisierbarer und interaktiver mobiler Dienste für den Automobilsektor

Holger Hoffmann, Dr. Jan Marco Leimeister, Prof. Dr. Helmut Krcmar

Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
85748 Garching bei München
{holger.hoffmann,leimeister,krcmar}@in.tum.de

Abstract

Ziel dieses Beitrags ist die Konzeption und Erprobung eines Ansatzes zur systematischen Entwicklung mobiler Mehrwert-Dienste im Fahrzeug. Dies beinhaltet die Beschreibung eines *Design Frameworks* um den Entwicklungsprozess mobiler Dienste in den verschiedenen Entstehungsphasen adäquat unterstützen zu können. Anhand eines *Vorgehensmodelles*, das iterative Diensteentwicklung mit Prototyping kombiniert, wurde ein konkreter Beispieldienst nach dem Design Framework erstellt. Für diesen Dienst, MACS¹ MyNews, wurden *Diensteszenarios* aufgestellt, ein *Wertschöpfungsnetzwerk* aufgezeigt, *Technologien* zur Erbringung ausgewählt, eine *prototypische Implementierung* durchgeführt und der so entwickelte Dienst hinsichtlich des Einflusses auf die *Fahrsicherheit* untersucht. Die Evaluation von Framework und Vorgehensmodell anhand des MACS MyNews Dienstes erlaubt empirisch fundierte Aussagen über Eignung und Nutzen des entwickelten systematischen Entwicklungsansatzes.

1 Einleitung

Mobile Dienste im Automobilmarkt waren die letzten Jahre nicht sehr erfolgreich in Deutschland. Als Hauptgründe für deren Scheitern am Markt werden meist die folgenden

¹ Mobile Automotive Cooperative Services (MACS) ist ein vom BMBF gefördertes Forschungsprojekt (FKZ: 01 HW 0207); weitere Informationen finden Sie unter www.projektmacs.de

Gründe genannt: a) die Kommunikationskosten waren zu hoch [Fros03], b) die angebotenen Dienste trafen nicht wirklich die Bedürfnisse der Kunden [Fuhr01; Werd05] und c) waren mobile Dienste vor allem in der Entwicklung zu sehr auf Technologien fixiert [Werd05], die mangelhafte Berücksichtigung ökonomischer Aspekte machte es fast unmöglich langfristig erfolgreiche Dienste anzubieten.

Durch die Verfügbarkeit neuer breitbandiger digitaler Übertragungswege wie UMTS (Universal Mobile Telecommunications System) oder DAB (Digital Audio Broadcast, Digitalradio) und dem Preisverfall besonders im Mobilfunk kann das erste der Probleme eliminiert werden. Hieraus ergibt sich die zentrale Forschungsfrage die in diesem Beitrag adressiert wird: „Wie können innovative mobile Dienste für den Automobilsektor systematisch entwickelt und zur Serienreife getragen werden?“

2 Anforderungen an die systematische Entwicklung mobiler Dienste im Fahrzeug

Zur Ermittlung der Anforderungen an mobile Dienste im Fahrzeug sowie deren systematische Entwicklung wurden über zwei Jahre Expertengespräche bei einem bayerischen Automobilhersteller geführt und die Anforderungen entsprechend systematisiert.

Dienste müssen mehr die existierenden Probleme des Endnutzers betrachten: daraus folgt, dass Rahmenbedingungen für die Erbringung des Dienstes erhoben werden müssen, der Kundennutzen wird in Form von Use Cases abgebildet und Endnutzer werden aktiv in den Erstellungsprozess eingebunden.

Dienste sollen rentabel sein: aus diesem Grund muss die Zahlungsbereitschaft der Kunden abgefragt werden, ein Prototyp kann dies unterstützen, der Business Case des Dienstes mit Value Proposition muss stimmig sein und der Dienst muss in das Herstellerkonzept für Dienst vor Kunde integriert werden (können).

Dienstleistung(en) werden nicht alleine, sondern mit Partnern erbracht: Für jeden Dienst gibt es ein spezielles Anforderungsprofil der benötigten Partner, ein akzeptables Geschäftsmodell für jeden Partner muss gefunden werden.

Technologie soll die Dienstleistung unterstützen, keine technologiegetriebene Dienstentwicklung: Technologien müssen auf Rahmenbedingungen des Dienstes gemappt und

evaluiert werden können. Eine Dienstplattform erlaubt den einfachen Austausch von Einzelkomponenten.

Dienste müssen sicher während der Fahrt benutzbar sein: Beachtung der jeweiligen gesetzlichen Vorgaben je nach Art des Dienstes, realitätsnahe Prototypenevaluation im Fahrzeug muss möglich sein.

Um mobile Dienste zukünftig erfolgreicher entwickeln zu können werden im folgend dargestellten Design Framework diese ökonomischen und technologischen Anforderungen aufgegriffen und zusammengeführt. Ergänzt wird dieses Design Framework durch ein Vorgehensmodell zur effizienten und effektiven Anwendung des Frameworks.

3 Das MACS Design Framework

Aufbauend auf den vorhergehend identifizierten Anforderungen ist das MACS Design Framework in folgende Teilbereiche (bzw. Sichten) untergliedert (vgl. Abb. 1), die jeweils einen Schritt im Entstehungszyklus eines mobilen Dienstes beschreiben: Die *Diensteszenarien* umfassen den allgemeinen Aufbau und Rahmen in dem der Dienst zum Einsatz kommen soll, sowie Use Cases für den mobilen Dienst. Anhand der Diensteszenarien werden detaillierte Informationen, wie Nutzeranforderungen oder die Zahlungsbereitschaft (bspw. über Workshops mit Fokusgruppen [BoDö02]), für bestimmte Dienste möglichst frühzeitig ermittelt. Eine Analyse der Diensteszenarien nach benötigten Partnern zur Dienstleistungserbringung ergibt Informationen in welcher Form die einzelnen Partner im *Wertschöpfungsnetzwerk* interagieren können und wie das Wertschöpfungsnetz aufgebaut sein kann bzw. sein soll.

Eine der größten Herausforderungen für ein Framework im Automobilbereich ist der große Zeitverzug in den Lebenszyklen von Automobil und der im Auto verwendeten Software [Hart04]. Während die durchschnittliche Lebensdauer eines Autos rund zehn Jahre beträgt, lösen neue Technologien und Software einander alle zwei bis drei Jahre ab und machen es damit schwierig den Lebenszyklus der Automobilherstellung zu managen [Moha06]. Um eine tragfähige Auswahl an Technologien zu ermöglichen und die Erstellung einer Infrastruktur zur Kompensierung des „lifecycle mismatch“ zwischen Software und Automobil zu ermöglichen ist besonderes Augenmerk auf die Bereiche *Technologien* sowie *Prototyp und Plattform* zu legen. Hinzu kommt die notwendige Betrachtung der *Sicherheitsaspekte* zur Sicherstellung der sicheren Benutzbarkeit des Dienstes während der Fahrt. Der Ausgangspunkt für diese

Betrachtungen ist die Verknüpfung der Diensteszenarien mit den verfügbaren Technologien, gefolgt von der Prototypingphase zur Demonstration und Evaluation des Risiko- und Nutzenpotentials. Ist der Dienst für straßentauglich befunden folgt die Planung des Roll-Out des *Live Dienstes*.

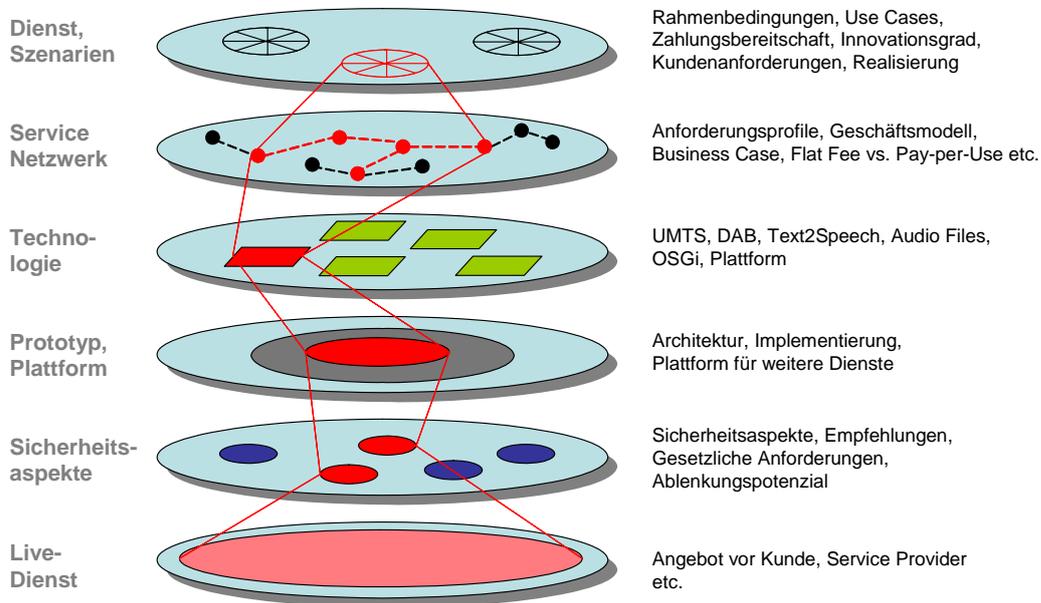


Abbildung 1: MACS Design Framework
(Quelle: eigene Darstellung)

Die Anwendung solch eines Frameworks zur systematischen, iterativen Entwicklung mobiler Dienste stellt die Entwickler allerdings vor einige Hürden: Strategien für das *Diensteszenario* müssen mit den verfügbaren *Technologien* in Übereinstimmung gebracht werden, das *Wertschöpfungsnetzwerk* muss die Strategien des Szenarios tragen, Beteiligte *unterschiedlicher Arbeitsgebiete* müssen zusammenarbeiten können, d.h. zumindest die Begrifflichkeiten der anderen verstehen und deren Entscheidungen nachvollziehen können und Funktionalität muss *iterativ aufgebaut* und kontinuierlich anhand bestimmter Faktoren *evaluiert* und *angepasst* werden.

Im weiteren Verlauf werden die Prozesse der Dienstleistungsentwicklung in den Mittelpunkt gestellt. Konkret bedeutet dies, dass die Stufen von den Diensteszenarien bis zur Evaluation der Sicherheitsaspekte weiter betrachtet werden, nicht jedoch die Planung des Live Dienstes und die Durchführung des eigentlichen Dienste Roll-Out.

4 Vorgehensmodell für die Anwendung des MACS Design Framework

Die Anwendung des MACS Design Frameworks beinhaltet einige Aspekte aus denen sich gezielt Anforderungen für ein Vorgehensmodell der Dienstleistungserstellung ermitteln lassen.

- Die neu konzipierten Dienste stellen Innovationen dar, Anforderungen lassen sich nicht alle im Vorfeld bestimmen und können sich während der Entwicklung auch ändern. Der Entwicklungsprozess muss also iterativ abgearbeitet werden können.
- Die mittels des Design Framework entwickelten Dienste sollen so gut wie möglich die Anforderungen der Kunden treffen, das bedingt ein hohes Maß an Nutzerintegration in den Entwicklungsprozess.
- Eine valide Sicherheitsevaluation für neu entwickelte mobile Dienste ist nur in der vorgesehenen Umgebung, d.h. im Auto sinnvoll. Das bedeutet, dass spätestens für die Sicherheitsevaluation ein im Fahrzeug testbarer Prototyp verfügbar sein muss.

Aus diesen Anforderungen ergibt sich entsprechend ein iterativer Entwicklungsprozess, der die Anforderungen der Nutzer sowie die Verkehrssicherheit des Dienstes durch Prototypenevaluationen einbeziehen kann. Die in der Informatik oft verwendeten linearen Entwicklungsmodelle, wie das Wasserfallmodell [Royc70], sind für diese Anforderungen also nicht anwendbar. Wesentlich geeigneter sind Prozessmodelle mit denen iterativ gearbeitet werden kann, z.B. das Spiralmodell [Boeh88]. Hier werden die Annahmen eines Prozessschrittes im Modell im folgenden Schritt evaluiert und liefern somit Eingaben für eine neue Iteration. Dies gilt sowohl für weitere Anforderungen während der Analysephase jedes Schrittes (d.h. Informationsfluss von oben nach unten) als auch für Evaluationsergebnisse für vorhergehende Schritte (d.h. Informationsfluss von unten nach oben).

Um den neuen Dienst hinsichtlich der Bedienbarkeit und der Straßentauglichkeit evaluieren zu können ist es nötig einen gut in die Fahrzeuginfrastruktur integrierten Prototypen zu erstellen und die Bedienschritte möglichst genau abzubilden. Auch hier ist die Anwendung der üblichen Standardmethoden des *vertikalen* oder *horizontalen Prototyping* (nach [Floy83]) ineffizient oder kaum zu gebrauchen. Ein horizontaler Ansatz verbietet die Integration des Prototypen in die Infrastruktur des Fahrzeuges, die sinnvolle Evaluation der mobilen Dienste in Ihrer späteren Umgebung ist damit unmöglich. Die komplette Integration durch einen vertikalen Ansatz bedeutet andererseits, dass die Benutzung des konkreten Dienstes nicht in seiner Gänze betrachtet werden kann. Damit ist wiederum eine Evaluation der Gefährdungspotentiale nur

bedingt möglich. Diese Probleme löst die Kombination beider Prototypingansätze zu einem *diagonalen Prototyping*: spezielle oder einzigartige Funktionalität des Dienstes, wie z.B. die Präsentation personalisierbarer Nachrichten oder die Eingabemöglichkeit via Sprachbefehl, wird im komplett implementiert (vertikaler Ansatz). Die meisten Funktionen davon in einer Form wie sie der Kunde im endgültigen Dienst erleben würde, aber nicht notwendiger Weise in derselben Form wie sie technisch im finalen Produkt umgesetzt würden (horizontaler Ansatz). Damit wird die Komplexität der Entwicklung in einer hoch proprietären embedded Umgebung (d.h. der Infrastruktur des Autos und seiner Infotainment-Einheit) entschärft. Es ist dabei jedoch immer noch möglich dem Endnutzer die wichtigen Innovationen in der in einem Fahrzeug erwarteten Form näherungsweise darzustellen.

Kombiniert man das iterativ gestaltete MACS Design Framework mit seinen abgegrenzten Phasen mit dem diagonalen Prototypingansatz ergibt sich das in Abbildung 2 dargestellte MACS Vorgehensmodell. Im Kern ist das Prozessmodell zur Entwicklung ein iterativer Prozess, abgeleitet aus den generischen Beschreibungen der Spiralmodelle von [Boeh88]. Hinzu kommen Elemente aus dem Prototyping, so wird im Vergleich zum ursprünglichen Spiralmodell mehr Wert auf die Ausarbeitung einzelner Szenarien, die Darstellung von (Zwischen-)Ergebnissen in Prototypen und der somit ermöglichten Einbeziehung der Nutzer, gelegt.

Jede der geplanten Iterationen beginnt mit einer Planungsphase für die Aktivitäten der jeweiligen Iteration. Darauf folgend werden die Anforderungen der aktuellen Phase erhoben, entweder anhand früherer Feldstudien und Experteninterviews (1. Iteration) oder durch die Einbeziehung von Endnutzern und Experten (2.-4. Iteration). Die in der Konstruktionsphase erstellten Komponenten/Systeme werden nach Abschluss der Phase evaluiert. Anhand der initialen Anforderungsanalyse und der aus der Evaluation extrahierten sozio-technischen & wirtschaftlichen Anforderungen ergeben sich somit die Eingaben für die folgende Iterationsphase.

Iterative Dienstentwicklung in Kombination ...

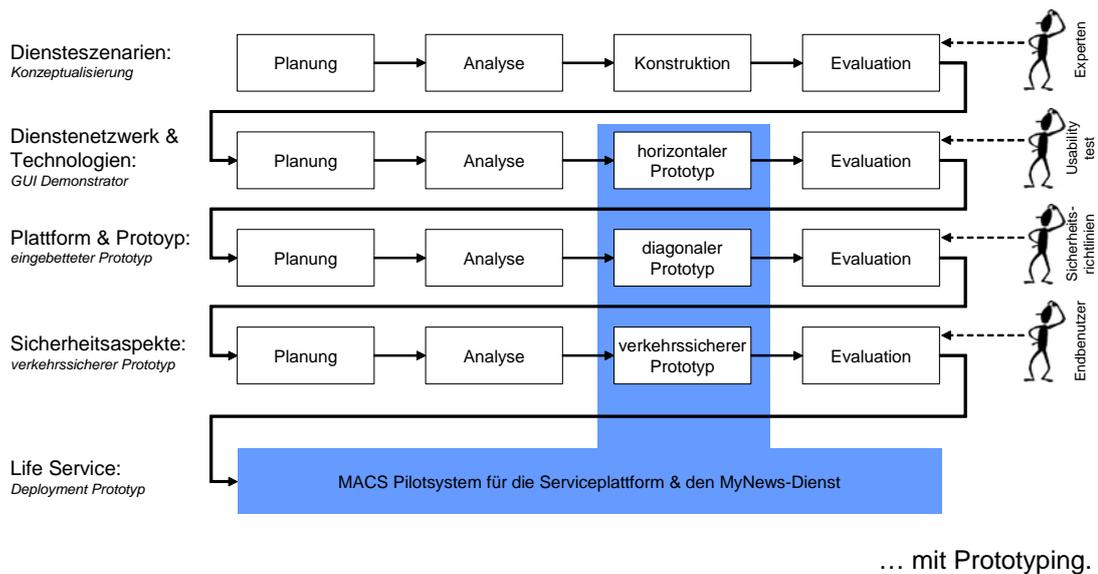


Abbildung 2: MACS Vorgehensmodell
(Quelle: eigene Darstellung)

5 Praktische Anwendung des MACS Design Framework nach dem Vorgehensmodell

Im folgenden Abschnitt wird die Anwendung des MACS Design Framework auf einen exemplarischen Dienst nach dem entsprechenden Vorgehensmodell beschrieben. Dabei werden jeweils insbesondere die Erkenntnisse der einzelnen Stufe auf dem Weg zu einem ausrollbaren mobilen Dienst für das Automobil herausgearbeitet. Die Planung des ausrollbaren Dienstes selbst ist nicht Gegenstand dieses Vorgehensmodelles. Erster Schritt dazu ist die Auswahl eines passenden Dienstes, der für eine Umsetzung Erfolg versprechend scheint. Als Basis der Auswahl dient ein „Forschungsradar“² von denkbaren personenbezogenen Diensten. Aus den insgesamt 40 Diensten des Forschungsradars wurde ein personalisierbarer Nachrichtendienst, MACS MyNews, aufgrund des Kundennutzen, der Bepreisbarkeit, des Innovationsgrades sowie der technischen Realisierbarkeit eines Prototypen als aussichtsreichste Idee bewertet und umgesetzt.

² Vgl. [Ehme02]. Hierzu sind unterschiedliche Vorschläge erarbeitet worden (vgl. bspw. Future Applications Lab (<http://www.viktoria.se/fal/>); Global Systems for Telematics (<http://www.gstproject.org/>)), das Thema ist ebenfalls laufend Gegenstand der SIGMobile (<http://www.sigmobile.org/>)

bestimmen, Gewichtungen der Kategorien vornehmen etc. Die Nachrichtensendung selbst beginnt „auf Knopfdruck“ und kann jederzeit pausiert werden, z.B. falls der Fahrer tanken muss. Es ist auch möglich mit MACS MyNews zu interagieren, also zwischen Nachrichten vor und zurück zu springen, bzw. sich eine Nachricht nochmals vorlesen zu lassen, falls man ein Detail der Nachricht verpasst hat.

5.2 MACS Wertschöpfungsnetz

Im zweiten Schritt der Entwicklung werden die benötigten Partner für ein Wertschöpfungsnetz spezifiziert und deren Verbindungen untereinander analysiert. Aus den vorher erstellten Diensteszenarien lassen sich mögliche Partner ermitteln. Im Falle von MACS MyNews wird für die Erstellung der Inhalte entsprechend ein Inhaltenanbieter für Nachrichteninhalte benötigt. Die Auswahl der Zielkunden erlaubt die Eingrenzung auf sinnvolle Partner bei den Automobilherstellern: Fahrer von hochpreisigen Limousinen sind eher an einem Live-Börsenticker interessiert als an mobilen Spielen. Durch die Klärung wie die Daten & der Dienst geliefert werden schließt sich die Lücke zwischen der Erstellung und dem Konsum der Inhalte. Es ist wichtig anzumerken, dass im Besonderen die erste und letzte Frage nicht sofort beantwortet werden können. In beiden Fällen ist es nötig die technologischen Möglichkeiten für den Dienst im folgenden Schritt auszuarbeiten um korrekte Antworten auf diese Fragen zu erhalten.

Das Wertschöpfungsnetz für MACS MyNews wird von der Dienstbeschreibung abgeleitet: die den Kunden präsentierten Nachrichten werden durch verschiedene Inhaltenanbieter (Content Provider) geliefert, ein Diensteanbieter (Service Provider) aggregiert die Daten (redundanzfrei) und bereitet sie redaktionell auf, bevor sie in das Auto übertragen und dargestellt werden. Der Hauptübertragungsweg zum Kunden ist Digitalradio (DAB), zusätzliche Mehrwertdienste, z.B. Detailinformationen oder Videos, können über das Mobiltelefon heruntergeladen werden. Die Anordnung der einzelnen Partner im MACS MyNews Wertschöpfungsnetzwerk ist sternförmig ausgelegt, mit dem Diensteanbieter im Zentrum. Die anderen Partner sind entsprechend ihrer Rollen während der Dienstbringung um den Diensteanbieter angeordnet.

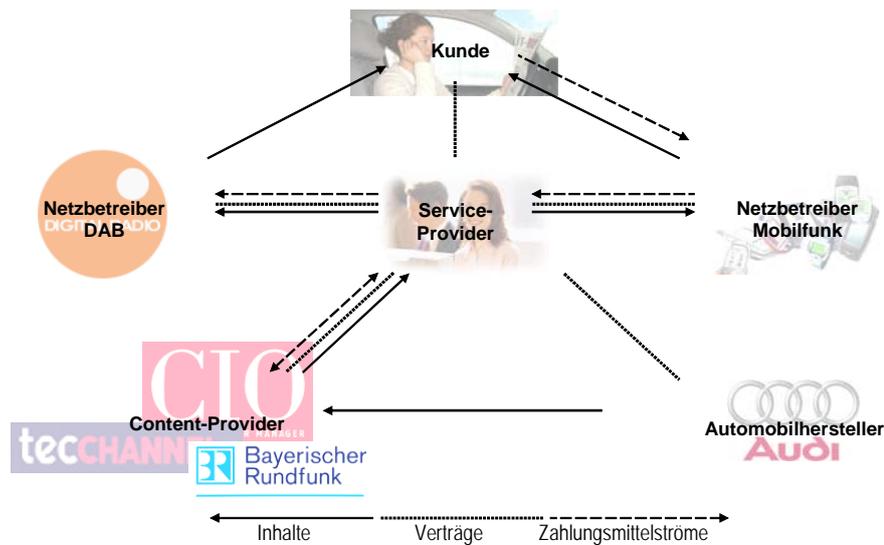


Abbildung 4: MACS MyNews Wertschöpfungsnetzwerk
(Quelle: eigene Darstellung)

5.3 MACS Technologien

Die logische Trennung der mobilen Dienste in die drei Abschnitte, Erstellung/Aggregation, Übertragung und Darstellung, wie sie im Wertschöpfungsnetz deutlich wird ist auch auf den Bereich der Technik anwendbar. So müssen für die *Erstellung/Aggregation* der Informationen passende Datenformate gefunden werden, welche zusammen mit dem Nutzungsszenario aus dem ersten Schritt beträchtlichen Einfluss auf die Art und Weise wie die Daten *übertragen* werden haben. Die im Fahrzeug verfügbare Menge und Art der Daten beeinflusst schließlich die Möglichkeiten der *Darstellung* im Fahrzeug.

Die Annahmen aus dem MACS MyNews Diensteszenario beeinflussen die Auswahl eines geeigneten Datenformates (Audio/Video oder Text) wie auch den Wahl des geeigneten Übertragungsweges. Die Kommunikationskosten der Übertragung per Mobilfunk werden vom Kunden bezahlt oder müssen vom Erbringer der Dienstleistung getragen werden, dadurch kann der Dienst allerdings unrentabel werden. Entsprechend dazu bietet sich als Übertragungsweg DAB an, das wiederum Fragen zu Datenformat, Übertragung und Darstellung nach sich zieht. In Anbetracht einer für Endnutzer kostenfreien MACS MyNews Version ist die logische Konsequenz die Übertragung über einen Rundfunkkanal, hier DAB. Audiodaten liefern zwar eine gute Sprachqualität, sind jedoch aufgrund ihrer Größe in der Übertragung sehr teuer und stellen deshalb keine Option für den konzipierten Dienst dar. Es bleibt die Übertragung der Informationen als Textdaten. Durch deren geringere Größe reduziert sich das

Übertragungsvolumen um einen Faktor von 30, die gesamten personalisierbaren Nachrichten können zeitnah in das Fahrzeug übertragen werden.

Als eine Art „Expertenevaluation“ kann hier gesehen werden, dass mit „Journaline“ das gleiche Problem der effizienten Datenübertragung von komplett unabhängiger Stelle auf genau dieselbe Art und Weise gelöst wurde³. Für zukünftige Versionen von MACS MyNews bietet es sich daher an auf dem jetzt verfügbaren Standard Journaline aufzubauen und dessen Verfügbarkeit und Verbreitung mit zu nutzen.

5.4 MACS Prototyp & Plattform

Eine gute Abstraktion der Abläufe in einem Framework für die Entwicklung einer Plattform zu finden ist eine schwierige Aufgabe, die oftmals mehrere Iterationsschritte umfasst [FaSJ99]. Um ein zuverlässiges Design in möglichst wenigen Iterationsschritten zu erhalten ist es nötig sowohl kontextspezifische Anforderungen zu erheben wie auch die aktuellen Best-Practices in einem möglichst ähnlichen Gebiet zu analysieren. Aus diesem Grund kommt eine Kombination zweier Analysemethoden zum Einsatz. Kontextspezifische Anforderungen in der Domäne (d.h. mobile Dienste im Fahrzeug) werden mittels der „domain analysis“ [ATMB99] erhoben und durch eine “best practice” Analyse [Boon99] der beiden am Markt zur Verfügung stehenden Architekturen, der Siemens “Top Level Architecture” (TLA) und AutoSAR [HSFB04], komplettiert.

Um eine strikt getrennte modulare Infrastruktur im Fahrzeug zu schaffen wird die Funktionalität des MACS Framework in „Basisdienste“ aufgeteilt, in denen Funktionen logisch gruppiert sind und die untereinander kommunizieren können. Als Laufzeitumgebung und als Kommunikationsplattform kommt das OSGi Framework [OSGi05] zum Einsatz. Jeder dieser Basisdienste steht entweder für ein Interface zur Infrastruktur des Automobils, stellt die sichere Benutzbarkeit (auch während der Fahrt) der mobilen Dienste sicher oder ist ein weiterer Architekturbaustein in OSGi [OSGi05]. Das OSGi Framework selbst stellt dabei wichtige und robuste Funktionen bereits zur Verfügung, u.a. das Lebenszyklusmanagement für Komponenten und das dynamische Finden von Funktionalität. Darüber hinaus ist es möglich das OSGi-basierte System aus der Ferne zu aktualisieren oder neue Komponenten hinzuzufügen [Pale02; Wong01].

³ Zink, A.: NewsService Journaline - Der Nachrichtenservice für den digitalen Rundfunk. In: Proceedings of the 17th International Scientific Conference Mittweida 2005. Mittweida, S. 20-23

Mit den vorgestellten Basisdiensten steht eine fahrzeugspezifische Programmierschnittstelle (API) zur Verfügung welche die Umsetzung neuer Dienste erleichtert. Dabei ist es nötig, dass von den Automobilherstellern (bzw. den Zulieferern) diese API entsprechend den offenen Spezifikationen umgesetzt wird um Entwicklern die eigenen Fahrzeuge als Zielplattform zugänglich zu machen.

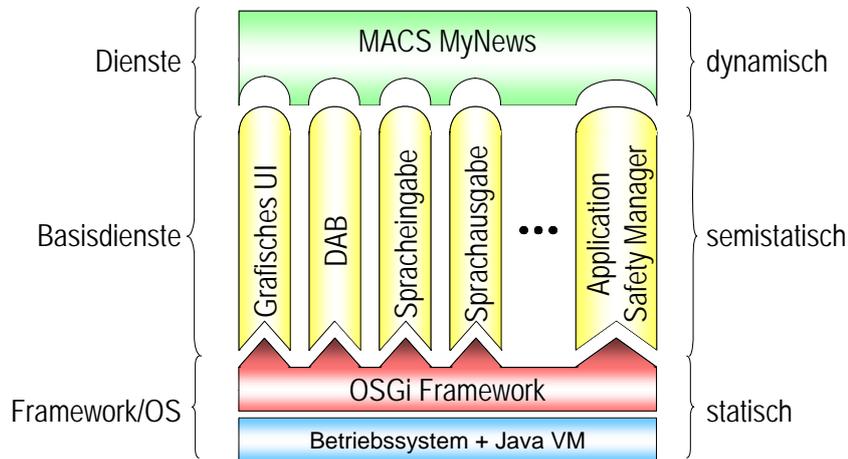


Abbildung 5: Architektur der MACS Dienstplattform
(Quelle: eigene Darstellung)

„User Interface“ Basisdienste ermöglichen die Ausgabe von Informationen auf einem grafischen Display wie auch die Ausgabe von Texten per Sprachsynthese (Text-To-Speech). Als Eingabegeräte können so die haptischen Bedienelemente des Praxispartners unterstützt werden, Sprachbedienung wurde als weitere Eingabemöglichkeit nachgerüstet. Die Auswahl der passenden Art und Menge an Information die während der Fahrt dargestellt werden kann, zusammen mit der Auswahl des passenden Ausgabemediums, wird von einem eigenständigen Basisdienst übernommen, dem „Application Safety Manager“. Der Datentransfer vom zu zum Fahrzeug wird über den „Universal Cellphone Adapter“ und den „Broadcast Adapter“ abgewickelt. Als eine der Grundlagen für den Aufbau dieser Basisdienste können die Spezifikationen der unidirektionalen Dienste (die lediglich Rundfunk nutzen) und bidirektionalen Dienste aus dem Kontext des DIAMOND Projektes [BHKP01; HKPR01] herangezogen werden.

5.5 MACS Sicherheitsaspekte

Da es sich bei den im Rahmen von MACS betrachteten Dienste um *mobile* Dienste handelt ist deren sichere Nutzbarkeit während der Fahrt von entscheidender Bedeutung. Um das Gefährdungspotential mobiler Dienste zu Bewerten hat das Institut für Arbeitswissenschaft der

TU Darmstadt eine Sammlung von Sicherheitsleitfäden für mobile Dienste zusammengestellt und als erstes den MACS MyNews Prototypen dagegen evaluiert. Eingang in diese Sammlung fanden sowohl Arbeiten der „state of the art“, z.B. [Beck96; Tije00], als auch die aktuelle Gesetzgebung für Dienste im Fahrzeug.

Um die Straßentauglichkeit mobiler Dienste zu gewährleisten müssen diese auch während der Fahrt sicher nutzbar sein und der lokalen Gesetzgebung gehorchen [Beck99]. Benutzerschnittstellen die während der Fahrt bedient werden können sind damit unabdingbar. Auf der MACS Plattform sind Benutzereingaben nicht an das haptische Gerät gebunden, bei dessen Benutzung der Fahrer eine Hand vom Lenkrad nehmen müsste und seine Blickrichtung stark ändert, sondern können auch per Sprache erfolgen [FäFä84]. Dasselbe gilt für Ausgabegeräte: neben der visuellen Schnittstelle zum Fahrer gibt es eine Text-to-Speech Engine für die Sprachsynthese von Texten über den Audiokanal. Der Fahrer kann sich somit auf die Straße konzentrieren und muss nicht auf Bedienteile oder Monitor sehen um den Dienst nutzen zu können.

Um das Risikopotenzial dieser Lösung und von MACS MyNews zu evaluieren, wurden mit verschiedenen Probanden Testfahrten durchgeführt, deren Aufbau mit der Arbeit von [WiNS98] vergleichbar ist. Die Verwendung von MACS MyNews wurde dabei mit verschiedenen Anwendungsszenarien geprüft. Auf der einen Seite variierte die Fahrsituation: die Fahrten fanden auf Landstrassen, Autobahnen und auf innerstädtischen Strassen statt. Auf der anderen Seite wechselte der Modus der Benutzerinteraktion: die Probanden bedienten den MyNews Dienst durch das haptische Bedienteil und durch Sprachbefehle. Als eine Vergleichsgröße des aktuellen Risikopotenzials wurde das Autoradio in den gleichen Situationen durch die haptische Schnittstelle bedient, Sprachbedienung für Autoradios stand nicht zur Verfügung. Nach jedem dieser Szenarien wurden der subjektive Stress und die Ablenkung durch Fragebögen evaluiert, die objektive Analyse der Szenarios erfolgte anhand von Videoaufzeichnungen der allgemeinen Verkehrssituation, der Mensch-Maschine-Schnittstelle (Eingabegeräte und Anzeige) und des Gesichtes des Fahrers.



Abbildung 6: Aufzeichnung eines Fahrversuches
(Quelle: TU Darmstadt; Inst. für Arbeitswissenschaft)

Erste Evaluationen der Antworten der Probanden haben ergeben, dass den Testfahrern die Nutzung des Dienstes während der Fahrt angenehm war und es keinen Unterschied in der Ablenkung zwischen dem gewohnten Autoradio und MACS MyNews gab. Aus den gegebenen Antworten ließen sich Änderungen an Details der Interfacegestaltung extrapolieren, diese wurden in einer weiteren Iterationsphase des Prototyping mit einbezogen.

6 Implikationen und Ausblick

Der so erstellte MACS MyNews Prototyp wurde Ende März 2006 auf der 6. Dienstleistungstagung des BMBF in Berlin vorgestellt und durch interessierte Fachbesucher, die die Möglichkeit hatten den Dienst im Fahrzeug zu testen, evaluiert. Die 51 Evaluationsteilnehmer beurteilen zu 76,5% (n=51) dass der Dienst als (sehr) interessante Erweiterung im Automobil angesehen wird. Bezogen auf den wahrgenommenen Nutzen beurteilten 64,7% (n=51) der Befragten den Dienst als (sehr) sinnvoll. Mit dem vom Prototypen dargestellten Funktionsumfang befanden 71,4% (n=49) Besucher als (sehr) gut. Die Benutzbarkeit / Usability wurde von 60% (n=50) der Teilnehmer als (sehr) intuitiv bezeichnet, 54,9% (n=51) bewerteten bereits die unoptimierte Form der Sprachausgabe als (sehr) gut. Hinsichtlich des Anschaffungspreises (inkl. Hardware und Verbindungskosten) wurden

verschiedene Preise zwischen 30€ und 480€ abgefragt, dabei ergab sich ein Ankerpreis von 230€.⁴

Zusammengefasst kann festgestellt werden, dass der Dienst im aktuellen Prototypenzustand nach der Evaluation als sehr viel versprechend angesehen werden kann. Daher wird der Dienst in die Neuproduktentwicklung bei einem Fahrzeughersteller überführt – der entsprechend dem MACS Framework letzten Gestaltungsebene.

Die praktische Anwendung des Design Framework anhand des Vorgehensmodelles bei der Entwicklung von MACS MyNews hat sich gezeigt, dass es möglich ist rasch zielgerichtetes Nutzerfeedback zu erhalten. Durch das iterative Vorgehen ergeben sich eine Reihe verschiedener Prototypen die jeweils den aktuellen Schritt im Design Framework betrachten. Das Design Framework für sich genommen trägt sehr gut dazu bei, eine gemeinsame Sicht auf den Dienst für alle Projektpartner zu ermöglichen. Dadurch, dass jeder einzelne Schritt gekapselt ist und konkrete Ergebnisse erarbeitet werden, wissen alle Beteiligten um den aktuellen Stand und können aktiv bei Entscheidungen diskutieren. Nachteil hierbei ist, dass nicht geplant ist einzelne Schritte auch parallel betrachten zu können, z.B. neue Technologien evaluieren und prototypisch umsetzen, während eine frühere Version auf ihre Sicherheitsaspekte hin überprüft wird. In einer weiteren Version des Design Frameworks sollen Prozesse die einander nicht direkt beeinflussen auch parallel bearbeitet werden können. Des Weiteren muss das Wissen auf Ebene der einzelnen Methoden in den Ebenen konkretisiert werden, idealerweise bekommen Entwickler zukünftig konkrete „Checklisten“ anhand derer sie ihren Entwicklungszyklus überprüfen können.

Die Erfahrungen aus der Umsetzung der unterschiedlichen Iterationen des Prototypen legen den Schluss nahe, dass die für den Entwicklungsprozess vorgestellte Plattform grundsätzlich geeignet ist, aber zusätzlich zu einer Rapid-Prototyping-Umgebung ausgebaut werden sollte. Damit ist es Dienstentwicklern rasch möglich Dienste schon in frühen Stadien im Fahrzeug zu zeigen was zu einer Verbesserung der Projektkoordination beiträgt und die Kommunikation zwischen verschiedenen Fachgruppen erleichtert und Kundenkliniken ermöglicht. Mit Hilfe dieses Werkzeuges zur Erstellung von Diensten lassen sich rasch neue Dienste aus dem Forschungsradar (Abbildung 3) umsetzen, da auf eine bestehende Infrastruktur zurückgegriffen werden kann und der Schwerpunkt der Entwicklungsarbeit auf neuen Komponenten liegen

⁴ Für einen ausführlichen Überblick zu den Evaluationsergebnisse vgl. auch Reichwald/Krcmar/Reindl (2007): Mobile Automotive Cooperative Services. Eul-Verlag (im Erscheinen)

kann. Aktuell werden bereits neue Dienste auf Basis von W-Lan aufgebaut und neuartige, avatarbasierte Mensch-Maschine Schnittstellen untersucht.

Literaturverzeichnis

- [ATMB99] *Aksit, M.; Tekinerdogan, B.; Marcelloni, F.; Bergmans, L.*: Deriving Frameworks from Domain Knowledge. In: *Fayad, M.E.; Schmidt, D.C.; Johnson, R.E. (Hrsg.): Building Application Frameworks - Object-Oriented Foundations of Framework Design.* John Wiley & Sons, New York 1999, S. 169-198.
- [Beck96] *Becker, S.*: Panel Discussion on Introduction of Intelligent Vehicles into Society: Technical, Mental and Legal Aspects. Mental Models, Expectable Consumer Behaviour and Consequences for System Design and Testing. In: *(Hrsg.): IEEE Intelligent Vehicles Symposium 1996.* S. 313-318.
- [Beck99] *Becker, S.*: Konzeptionelle und experimentelle Analyse von Nutzerbedürfnissen im Entwicklungsprozess. In: *Bundesanstalt für Straßenwesen (Hrsg.): Informations- und Assistenzsysteme im Auto benutzergerecht gestalten. Methoden für den Entwicklungsprozess.* Verlag für neue Wissenschaft, Bergisch Gladbach 1999, S. 64-72.
- [BHKP01] *Betram, G.; Hallier, J.; Koch, H.; Perrault, O.; Kuck, D.; Korte, O.; Twietmeyer, H.* (2001). *DIAMOND - Technical Specification for Bi-Directional Services.*
- [BoDö02] *Bortz, J.; Döring, N.*: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 3 Aufl., Springer Verlag, Berlin u.a. 2002.
- [Boeh88] *Boehm, B.W.*: A Spiral Model of Software Development and Enhancement. In: *IEEE Computer* 21 (1988) 5, S. 61-72.
- [Boon99] *Boone, J.*: Harvesting Design. In: *Fayad, M.E.; Schmidt, D.C.; Johnson, R.E. (Hrsg.): Building Application Frameworks - Object-Oriented Foundations of Framework Design.* John Wiley & Sons, New York 1999, S. 199-210.
- [Broc99] *Brockhoff, K.*: Produktpolitik. 4 Aufl., UTB für Wissenschaft, 1999.
- [Ehme02] *Ehmer, M.*: Mobile Dienste im Auto – Die Perspektive der Automobilhersteller. In: *Reichwald, R. (Hrsg.): Mobile Kommunikation: Wertschöpfung, Technologies, neue Dienste.* Gabler, Wiesbaden 2002, S. 459-472.

- [FäFä84] Färber, B.; Färber, B. (1984). *Sprachausgaben im Fahrzeug. Handbuch für Anwender*. Frankfurt am Main: Forschungsvereinigung Automobiltechnik e.V.
- [FaSJ99] *Fayad, M.E.; Schmidt, D.C.; Johnson, R.E.: Building Application Frameworks - Object-Oriented Foundations of Framework Design*. John Wiley & Sons, New York 1999.
- [Floy83] *Floyd, C.: A Systematic Look at Prototyping*. In: *Budde, R.; Kuhlenkamp, K.; Matthiassen, L.; Züllighofen, H. (Hrsg.): Approaches to Prototyping 1983*. Namur, S. 1-18.
- [Fros03] Frost & Sullivan (2003). *Customer Attitudes and Perceptions Towards Telematics in Passenger Vehicles Market*.
- [Fuhr01] *Fuhr, A.: Die Telematik ist tot - es lebe die rollende Schnittstelle*. In: (Hrsg.): Euroforum Jahrestagung "Telematik" 2001. Bonn.
- [Hart04] *Hartmann, J.: Wo viel Licht ist, ist starker Schatten - Softwareentwicklung in der Automobilindustrie*. In: *Automatisierungstechnik 52 (2004) 8*, S. 353-358.
- [HKPR01] *Hallier, J.; Kuck, D.; Perrault, O.; Rucine, P.; Twietmeyer, H.; Korte, O.; Capra, L.; Betram, G.; Fernier, M.; Schulz-Hess, T. (2001). DIAMOND - Technical Specification for Uni-Directional Services*.
- [HSFB04] *Heinecke, H.; Schnelle, K.-P.; Fennel, H.; Bortolazzi, J.; Lundh, L.; Leflour, J.; Maté, J.-L.; Nishikawa, K.; Scharnhorst, T.: AUTomotive Open System ARchitecture - An Industry-Wide Initiative to Manage the Complexity of Emerging Automotive E/E-Architectures*. In: (Hrsg.): *International Congress on Transportation Electronics 2004*. Detroit, S. 325-332.
- [Moha06] *Mohan, L.R.: Driving down the Fast Lane: Increasing Automotive Opportunities the EMS Provider Way*. <http://www.frost.com/prod/servlet/market-insight-print.pag?docid=67150588>, Abruf am: 29.04.
- [OSGi05] OSGi Alliance (2005). *About the OSGi Service Platform, Technical Whitepaper*.
- [Pale02] *Palenchar, J.: OSGi Networks Ready to Roll*. In: TWICE (2002), S.
- [Royc70] *Royce, W.W.: Managing the development of large software systems*. In: (Hrsg.): *International Conference on Software Engineering 1970*. Monterey, S. 328-338.
- [Stat05] Statistisches Bundesamt Deutschland (2005). *Leben und Arbeiten in Deutschland - Ergebnisse des Mikrozensus 2004*. Wiesbaden.
- [Tije00] *Tijerina, L.: Issues in the Evaluation of Drive Distraction Associated with In-vehicle Information and Telecommunication Systems*. <http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/3.PDF>, May 2000, Abruf am: 5.4.

- [Unit03] United States Census Bureau (2003). *2003 American Community Survey*. Washington D.C.
- [Werd05] *Werder, H.*: Verkehrstelematik als Element der Verkehrspolitik. In: (*Hrsg.*): its-ch 2005. Olten.
- [WiNS98] *Wikman, A.-S.; Nieminen, T.; Summala, H.*: Driving experience and time-sharing during in-car tasks on roads of different width. In: *Ergonomics* 41 (1998) 3, S. 358-372.
- [Wong01] *Wong, W.*: Open Services Gateway Initiative: OSGi Links Devices and Clients. In: *Electronic Design* (2001), S. 86.

Einführung in den Track

Outsourcing und IT-Governance

Prof. Dr. Armin Heinzl

Universität Mannheim

Prof. Dr. Gerhard F. Knolmayer

Universität Bern

Dr. Michael Heym

Navisco AG

Im Zuge der Globalisierung hat die IT nicht nur an Bedeutung für Unternehmen gewonnen - sie unterliegt auch selbst der Dynamik eines internationalen Wettbewerbsumfeldes. Für Unternehmen bietet sich in diesem Zusammenhang eine Vielzahl neuer Möglichkeiten der Erbringung ihrer IT-Leistungen, die es bei der Frage nach der Eigenerstellung oder dem Fremdbezug zu evaluieren gilt. Das Offshoring und das Business Process Outsourcing, aber auch die gebündelte Erbringung von IT-Leistungen in so genannten Shared Service Centers sind derzeit viel diskutierte Trends, die in diesem Track einer genaueren Betrachtung unterzogen werden sollen. Daneben ergeben sich auch neue Herausforderungen an und Konsequenzen für die Organisation, die Steuerung und das Management der IT. Diese sollen im Rahmen des vorliegenden Tracks behandelt werden.

Eine Typologie von Beziehungen im IT-Outsourcing: Ein konzeptioneller Ansatz

Stefanie Jahner, Tilo Böhmman, Helmut Kremer

Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
85748 Garching bei München
{jahner | boehmann | kremer} @in.tum.de

Abstract

Trotz der Vielzahl unterschiedlicher Varianten von IT-Outsourcing lassen sich in der Praxis zahlreiche gescheiterte Vorhaben dokumentieren. Die Forschung hat sich daher auf die Untersuchung des Managements von Kunden-Anbieter-Beziehungen im IT-Outsourcing fokussiert. Bislang konzentrieren sich die Erkenntnisse auf die Beschreibung genereller Beziehungsfaktoren. Unterschiedliche Kontexte und Erwartungen von Kunden und Anbietern werden dabei nicht berücksichtigt. Relevanz und Ausprägung unterschiedlicher Beziehungsfaktoren hängen jedoch stark vom Outsourcingt看 ab. Basierend auf einer Literaturanalyse und qualitativer Experteninterviews entwickelt der Beitrag eine Klassifikation unterschiedlicher Outsourcingbeziehungstypen. Ansatzpunkte für ein erfolgreiches Management der Outsourcingbeziehung, für die Weiterentwicklung und empirische Fundierung der Beziehungstypen schließen den Beitrag ab.

1 Einleitung und Relevanz der Forschung

Obwohl kein neues Phänomen, ist die Auslagerung der Informationstechnologie (IT) an externe Dienstleister immer noch ein relevantes und aktuelles Thema auf der Agenda von IT-Managern [LuKN06]. Relevanz, Ausprägungen und Vielfalt von IT-Outsourcing haben sich jedoch im Laufe der letzten Jahrzehnte stark gewandelt. Während in den 80er und 90er Jahren vorrangig großvolumige, langfristige Outsourcingverträge abgeschlossen wurden, tendieren Unternehmen heute aufgrund stark veränderlicher Anforderungen eines dynamischen Marktumfeldes zur sehr selektiven, vergleichsweise kurzfristigen Auslagerung ihrer IT, die oftmals im Rahmen einer

„Multi-Sourcing-Strategie“ auf mehrere IT-Dienstleister verteilt wird [CoYo06]. Gleichzeitig sind strategische Formen wie Geschäftstransformationen mit Unterstützung externer IT-Dienstleister ein aufkommender Trend. In diesem Kontext prägen Risikoübernahme-Modelle und kollaborative Entwicklung von IT-Services die Erwartungen der Kunden.

Trotz der Vielzahl der beschriebenen Varianten von IT-Outsourcing, die sich in den letzten Jahrzehnten entwickelt haben, erfüllen viele Outsourcingvorhaben nicht die Erwartungen der Kunden und scheitern oftmals vorzeitig [oV05]. Als Gründe hierfür werden einerseits die Unterschätzung der Komplexität von Outsourcingvorhaben, fehlende Erfahrung im Management von Outsourcing aber auch mangelndes Verständnis, dass die Steuerung der Partnerbeziehungen mehr Aufwand erfordert als geplant, angeführt [CoYo06]. In diesem Kontext hat sich ebenfalls gezeigt, dass ein detailliert spezifizierter Outsourcingvertrag Grenzen hat und das Management einer Outsourcingbeziehung jenseits von starren Vertragselementen zu berücksichtigen ist [Klep97].

Aus diesen praxisorientierten Problemstellungen hat sich die Diskussion in der Outsourcingforschung verstärkt auf das post-vertragliche Management und die Entwicklung einer kontinuierlichen Kunden-Anbieter-Beziehung im Outsourcing fokussiert [vgl. bspw. die Arbeiten von GoCh05; Kern97; KeWi00; KeWv02; KiCh03]. Obwohl zahlreiche Autoren die Relevanz der Betrachtung und Gestaltung einer solchen Kunden-Anbieter-Beziehung als Erfolgsfaktor im IT-Outsourcing betonen [LeKi99; LHCP00], wurden im Vergleich zu anderen Forschungssträngen im IT-Outsourcing bislang wenig Forschungsbemühungen in Richtung einer eingehenden Untersuchung von Determinanten, die eine Outsourcingbeziehung charakterisieren, unternommen [Kern97]. Arbeiten in diesem Bereich haben sich vorrangig auf die Aufarbeitung genereller Beziehungsfaktoren konzentriert, dabei aber nicht die unterschiedlichen Formen, Rahmenbedingungen und Erwartungen von Kunden und Anbietern im Outsourcing berücksichtigt [AISS04; GoCh05; KeWi00; Klep97]. Es lässt sich daher eine Forschungslücke in zweierlei Hinsicht identifizieren: Beziehungsfaktoren können nicht statisch und kontextfrei betrachtet werden. Vielmehr hängen ihre Relevanz und Ausprägungen essentiell von Form, Kontext und Erwartungen beider Outsourcingparteien ab. Bisherige Forschungsansätze unterscheiden zwar verschiedene Formen von Outsourcing, jedoch nicht zwischen unterschiedlichen Typen von Outsourcingbeziehungen, die diese Formen erfordern.

Der vorliegende Beitrag adressiert diese Forschungslücke, indem er eine Typologie unterschiedlicher IT-Outsourcingbeziehungen entwickelt. Die Identifikation von Archetypen ist ein zentra-

ler Ansatzpunkt, um unterschiedliche Erwartungen zu klassifizieren und eine entsprechend kontingente Gestaltung der daraus resultierenden Outsourcingbeziehung zu ermöglichen.

Als Ausgangspunkt wird hierbei unterstellt, dass beide Outsourcingparteien unterschiedliche Erwartungen an eine Outsourcingbeziehung haben. Aus den jeweiligen Erwartungen ergibt sich ein bestimmter Typ von Outsourcingbeziehung, der zu den Erwartungen passt. Eine gleichgerichtete bzw. komplementäre Ausprägung der Erwartungen und Ziele beider Parteien ist die Grundlage für die Gestaltung der Outsourcingbeziehung und determiniert Typ und Ausprägungen des Outsourcingvorhabens. Darauf aufbauend ist die passende Ausgestaltung der Art der Outsourcingbeziehung im Hinblick auf Strategie, Beziehungsstruktur, Steuerungsmechanismen und Interaktion von großer Bedeutung für den Erfolg des Outsourcingvorhabens. Die Prämisse des Beitrages ist, dass nur im Falle passender Typen von Outsourcingbeziehungen das Outsourcingvorhaben erfolgreich ist (siehe hierzu auch das übergeordnete Forschungsmodell in Abbildung 1).

Zunächst wird die relevante Literatur hinsichtlich bisher zusammengetragener, konstituierender Beziehungsfaktoren untersucht, die zu vier zentralen Kategorien von Beziehungsfaktoren verdichtet werden. Aus der Literatur wird ein Klassifikationsrahmen mit den in der Literatur identifizierten und zu Beziehungskategorien verdichteten Beziehungsfaktoren abgeleitet und aufgestellt. In einem nächsten Schritt werden Erwartungen von Kunden und Anbietern sowie Ausprägungen von Outsourcingbeziehungen in explorativen Experteninterviews empirisch untersucht. Die erhobenen Daten werden entlang den aus der Literatur identifizierten Beziehungskategorien in den Bezugsrahmen eingeordnet. Hieraus ergeben sich fünf Archetypen von Outsourcingbeziehungen, die entsprechend der Beziehungskategorien charakterisiert werden. Neben dem Vorschlag einer Typologie unterschiedlicher Beziehungen ist das übergeordnete Ziel dieses Beitrags ein tiefer gehendes Verständnis über die Notwendigkeit und Gestaltungsparameter kunden- und anbieterspezifischer Erwartungen an Outsourcingbeziehungen als Determinante für Outsourcingserfolg.

Diesem Beitrag liegt ein übergeordnetes Forschungsmodell zugrunde, das den Zusammenhang zwischen den Erwartungen von Kunde und Anbieter, den konstituierenden Gestaltungselementen einer Outsourcingbeziehung und dem Erfolg des Outsourcingvorhabens verdeutlicht (Abbildung 1). Obwohl das Forschungsmodell diesen Gesamtzusammenhang zeigt, liegt der Fokus dieses Beitrages auf der Entwicklung unterschiedlicher Typen von Outsourcingbeziehungen, die die Erwartungen beider Parteien abbilden. Die Wirkung von Erwartung und Typ der Outsour-

ingbeziehung auf den Erfolg der Outsourcingbeziehung wird hierbei unterstellt, ist jedoch erst anhand von Erfolgsmaßen wie bspw. Kundenbindung, Abwanderungsrate („Churn Rate“) oder Kundenzufriedenheit in einer nachfolgenden Forschungsarbeit zu überprüfen.

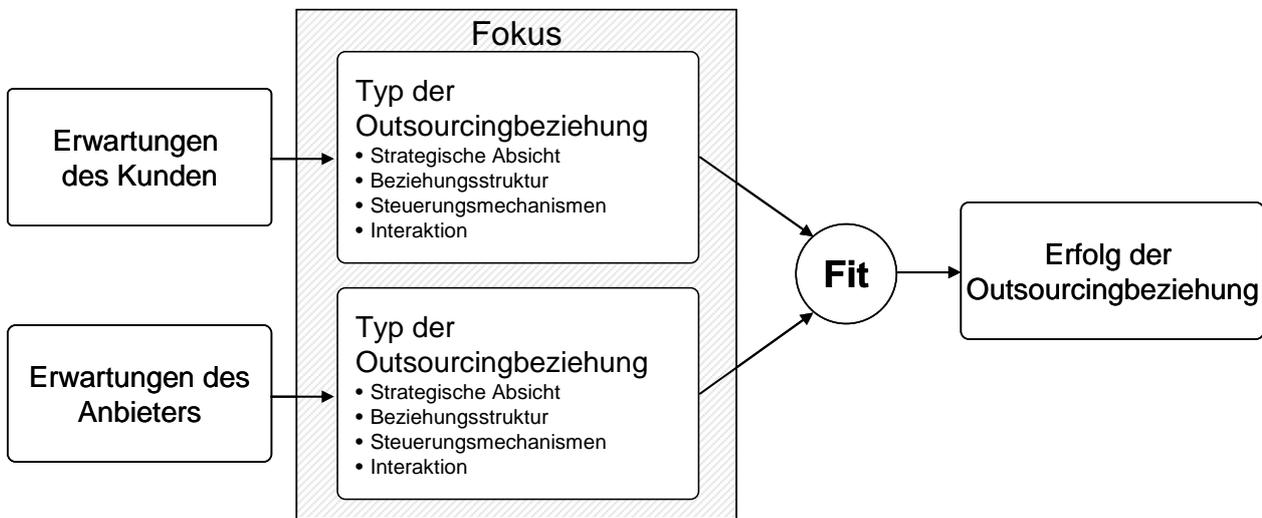


Abbildung 1: Gesamtrahmen: Forschungsmodell

2 Konstituierende Faktoren von IT-Outsourcingbeziehungen

Nach Konsynski und McFarlan [KoMc90] kann Outsourcing als eine Form von strategischer Informationspartnerschaft bzw. -beziehung angesehen werden, die den beteiligten Parteien gegenseitigen Zugang zu Information, Ressourcen, Kunden etc. eröffnet. Verbindet man dieses Beziehungsverständnis mit dem allgemeinen Auffassung von IT-Outsourcing als „Auslagerung von IT-Aktivitäten, -Prozesse, -Ressourcen oder -Services an einen externen Dienstleister“ [DGHJ04], gelangt man zu einer spezifischen Definition einer IT-Outsourcingbeziehung, die sich als *„eine kontinuierliche Verflechtung und Beziehung zwischen einem Outsourcingnehmer und -geber, die auf einer vertraglichen Übereinkunft basiert, eine oder mehrere übergreifende IT-Aktivitäten, -Prozesse und -Services anzubieten“* formulieren lässt, *„wobei Gewinn und Verlust beider Seiten zumindest in Teilen voneinander abhängen“* [GoCh05].

Eine für beide Seiten profitable Outsourcingbeziehung, bei der beide Parteien die gleichen Ziele verfolgen, kann jedoch nur unter bestimmten Rahmenbedingungen angenommen werden. So weisen Lacity und Hirschheim [LaHi93] kritisch darauf hin, dass eine Outsourcingbeziehung durchaus unterschiedliche Motive auf beiden Seiten aufgrund von asymmetrischer Ressourcen-

und Machtverteilung beinhalten kann. Solche Machtstrukturen müssen in der Beziehungsstruktur und bei der Wahl der Steuerungsmechanismen berücksichtigt werden (siehe Kap. 2.2).

Diese kritische Reflektion zeigt, dass Outsourcingbeziehungen ein facettenreiches und komplexes Phänomen sind. Aufbau und Ausprägung einer Outsourcingbeziehung hängen stark von den Erwartungen und Rahmenbedingungen von Kunde und Anbieter ab. In der Literatur lassen sich folglich zahlreiche Beziehungsfaktoren identifizieren, die eine Outsourcingbeziehung determinieren [AlSS04; GoCh05; KeWi00]. Da eine ausführliche Diskussion unterschiedlicher Beziehungsfaktoren außerhalb des Rahmens und Fokus dieses Beitrages liegt¹, gibt Tabelle 2 einen detaillierten Überblick über ausgewählte Beziehungsfaktoren. Zur Systematisierung werden diese Beziehungsfaktoren in Anlehnung an Kern und Willcocks [KeWi00] anhand von vier konstituierenden Beziehungskategorien im Folgenden kategorisiert.

2.1 Strategische Absicht und Kontext

Die Gestaltung einer Outsourcingbeziehung hängt maßgeblich von den strategischen Rahmenbedingungen beider Parteien ab. Kern und Willcocks [KeWi00] bezeichnen dies als kontextabhängige Dimension einer Outsourcingbeziehung. Hierunter fallen zum einen die Ziele und Erwartungen des Kunden gegenüber dem Outsourcingvorhaben (in ökonomischer, technischer und politischer Hinsicht) sowie das Verständnis von der Rolle des IT-Dienstleisters. Ziele und Erwartungen wiederum werden durch interne und externe wirtschaftliche Rahmenbedingungen und Geschäftsherausforderungen des Unternehmens determiniert [LaHi94]. In einfachen, meist sehr selektiven und gekapselten „commodity“-Outsourcingkonstellationen sind strategische Überlegungen und strategischer Einfluss vergleichsweise gering ausgeprägt. Der Fokus des Kunden liegt hier vielmehr auf einer effizienten IT oder dem Zugang zu einem Pool technologischer Ressourcen, als auf einer strategischen Geschäftstransformation oder Technologieführerschaft. Der Dienstleister wird hier in einer unterstützenden Rolle gesehen, die im Vergleich zu strategischen Allianzen kaum auf antizipative und proaktive Vorschläge zur Entwicklung komparativer Wettbewerbsvorteile ausgerichtet ist [KRNR03].

2.2 Beziehungsstruktur

Unter der Struktur einer Outsourcingbeziehung werden Aufbau und Gestaltung der Rahmenbedingungen innerhalb des Outsourcingarrangements subsumiert. Unterschiede in verschiedenen

¹ eine detaillierte Übersicht zu Beziehungsfaktoren im IT-Outsourcing findet sich bspw. bei Goles [GoCh05].

Outsourcingkonstellationen ergeben sich beispielsweise im Hinblick auf die Netzwerkstruktur, gegenseitige Abhängigkeiten, Fokus des Outsourcingobjektes oder Zeithorizont der Outsourcingbeziehung. Sambamurthy und Zmud [SaZm00] unterscheiden drei Arten von Beziehungsstrukturen: (1) strategische Strukturen, die relative stabile Beziehungen mit einer begrenzten Anzahl von Partner abbilden, (2) erweiterte Netzwerke, die entstehen, wenn Unternehmen auf einen Pool standardisierter IT-Fähigkeiten zurückgreifen, der sich aus mehreren externen Partnern zusammensetzt, und (3) virtuelle, lose Netzwerke mit einer großen Anzahl potentieller Partnern, auf die bei Bedarf zurückgegriffen werden kann. Je nach Fokus der Outsourcingbeziehung eignen sich unterschiedliche Strukturen, um die gesetzten Ziele zu erreichen.

Ein weiterer zentraler Aspekt, der bei der Gestaltung der Beziehungsstruktur in unterschiedlichen Outsourcingkontexten zu berücksichtigen ist, sind die Machtverhältnisse zwischen Anbieter und Kunde. Asymmetrien im Kräftespiel der Parteien resultieren aus den gegenseitigen Abhängigkeiten [East92] und werden dann kritisch, wenn der Einfluss einer Partei dominiert und beispielsweise zu opportunistischem Verhalten führt. In einer langfristigen, stabilen Outsourcingbeziehung, bei der große Teile der Informationstechnologie extern vergeben werden, dominiert meist der Anbieter, in selektiven Outsourcingverhältnissen ist das Machtverhältnis ausgeglichener zugunsten des Kunden, der die Abhängigkeiten durch das Outsourcingvolumen oder durch Verteilung auf mehrere Anbieter minimieren kann [Kern97]. Lacity und Willcocks [LaWi03] betonen, dass ausgeglichene Machtstrukturen eine zentrale Rolle spielen, wenn gleiche Ziele und Nutzen verfolgt werden. Allerdings können auch unterschiedliche Ziele, die beispielsweise komplementär zueinander sind, zu einer erfolgreichen Outsourcingbeziehung führen, je nach Dynamik und Interaktionsform im Outsourcingkontext (siehe Kap. 2.4).

2.3 Steuerungsmechanismen (Governance und Monitoring)

Verträge als formale Abkommen, zu denen sich beide Parteien verpflichten, sind ein grundlegendes Element jedes Outsourcingvorhabens und das primäre Instrument zur Steuerung von Outsourcing-Geschäftsbeziehungen. Detaillierte festgeschriebene Vereinbarungen im Hinblick auf Leistungsumfang, Qualitätsziele, und Kennzahlen (z.B. Service Level Agreements) liefern die Basis für eine spätere Ergebniskontrolle und schränken den Spielraum für opportunistisches Verhalten ein [BeSc05]. Obwohl Verträge im Rahmen der IT-Outsourcing-Governance ein vertrautes, ergebnisorientiertes Kontrollinstrument sind, sind sie nicht immer ein geeigneter Ansatz. Verträge verhandeln, aufsetzen, durchführen und kontrollieren ist ein aufwendiges und komplexes Unterfangen, oftmals sind Verträge unflexibel gestaltet und spiegeln nicht die not-

wendige Dynamik und Entwicklung einer Outsourcingbeziehung wider [JaBK06; LaWi03]. Vertragsanpassungen aufgrund veränderlicher Geschäftsherausforderungen können im Voraus oftmals nicht spezifiziert werden und sind dann mit erheblichen Kosten verbunden [KRNR03]. Folglich sind Verträge alleine keine zweckmäßigen Steuerungsmechanismen [Kern97]. Neben formalen Mechanismen werden daher oft auch informelle, auf gemeinsamen Erwartungen und Verhaltensnormen basierende Steuerungsformen gewählt. Darunter fallen ebenfalls sog. Beziehungsprotokolle, die im Laufe der Zeit durch kontinuierliche Zusammenarbeit zwischen Kunde und Dienstleister entstehen. Sie umfassen implizite Regeln, die zwischen den Beteiligten erwartete Verhaltensweisen und Einstellungen wie Vertrauen, Commitment, Fairness oder Flexibilität definieren. Solche Beziehungsprotokolle können sich einerseits, wenn auch zunächst auf Basis persönlicher Beziehungen entwickelt, auf der institutionellen Ebene auswirken. Die Outsourcingbeziehung kann sich dann von einer vertragsorientierten zu einer kooperativen, verhaltensorientierten Steuerungsform entwickeln [East92]. Andererseits besitzen sie bei entsprechender Ausprägung und in Abhängigkeit vom Typ der Outsourcingbeziehung eine tief greifende Kontroll- und Koordinationsfunktion [BeSc05]. Während die Erbringung selektiver Standardleistungen im IT-Outsourcing erfolgreich über formale, ergebnisorientierte Mechanismen wie Verträge gesteuert werden kann, sind insbes. bei strategischen Allianzen Mechanismen jenseits von starren Vertragselementen ein sinnvolles Instrument zur Steuerung der Beziehung [KRNR03].

2.4 Interaktion

Die Beziehungskategorie Interaktion zielt auf den wechselseitigen Austausch in Outsourcing-Geschäftsbeziehungen ab und beschreibt verhaltensorientierte Muster beider Parteien. Interaktion kann entlang von Dimensionen wie Wissenstransfer, Informationsweitergabe, Risiko-Nutzen-Teilung, Vertrauen oder die Rolle von kulturellen Ähnlichkeiten zwischen Outsourcingnehmer und -geber charakterisiert werden [KeWi00]. Viele Autoren betonen die Ausrichtung auf gemeinsame Ziele als essentielle Voraussetzung für eine erfolgreiche Outsourcingbeziehung [LHCP00]. Dies berücksichtigt jedoch nicht, dass systemimmanent die Absichten beider Parteien in einer Outsourcingbeziehung gegensätzlich gerichtet sind [LaWi03]. Lacity und Willcocks [LaWi03] betonen daher vielmehr die Notwendigkeit, Dynamik und Entwicklungspotenzial einer Outsourcingbeziehung als zentralen Erfolgsfaktor zu berücksichtigen. Dabei unterscheiden sie drei unterschiedliche Arten von Kunden-Anbieter-Interaktionen: vorläufige, kooperative und kollaborative Interaktionen. „Vorläufige“ Interaktionen kommen vor allem zu Beginn einer Outsourcingbeziehung, bspw. in Vertragsverhandlungen, zustande, wenn die Ziel-

funktionen der Beteiligten noch unbekannt und unsicher sind und jede Seite dazu tendiert, ihre Stärken besonders herauszustellen, die Schwächen hingegen zu verbergen. Kooperative Interaktionen treten bei komplementären Zielen auf, bspw. wenn der Kunde als Ziel einen Service in Anspruch nimmt, für den der Anbieter auf der anderen Seite Bezahlung fordert. Von kollaborativen Interaktionen schließlich spricht man, wenn beide Seiten gemeinsame Ziele verfolgen, wie bspw. die gemeinsame Entwicklung innovativer IT Services.

Zentrale Erkenntnis der Ausführungen von Lacity und Willcocks [LaWi03] ist, dass Kunde und Anbieter durchaus unterschiedliche Ziele verfolgen können, ohne den Erfolg der Outsourcing-Beziehung zu gefährden, solange die Machtverhältnisse auf beiden Seiten ausgeglichen sind und beide Seiten gleichwertige Ergebnisse erzielen können.

Ein weiterer Aspekt, der je nach Typ der Outsourcing-Beziehung unterschiedlich ausgeprägt ist und die Interaktion beider Parteien determiniert, ist die Erfahrung des Kunden. Meist wird beim Thema Erfahrung auf die Rolle des Anbieters und seine Fähigkeiten bei der Leistungserbringung abgezielt [KiCh03]. Für Aufbau und Management einer dauerhaften Interaktion in der Outsourcing-Geschäftsbeziehung spielt jedoch die Erfahrung des Kunden – gleichermaßen auf das auslagernde Unternehmen insgesamt wie auch auf die konkrete Person des IT-Verantwortlichen auf Kundenseite bezogen – mit Outsourcingprojekten eine zentrale Rolle.

3 Erwartungen an Outsourcingbeziehungen in der Praxis

3.1 Datenerhebung

Basierend auf den Erkenntnissen aus der Literatur über Beziehungsfaktoren und deren Ausprägungen wurde erste empirische Evidenz für die Existenz unterschiedlicher Typen von Outsourcingbeziehungen erhoben. Um die Vielfalt der Aspekte und Dynamik in Outsourcingbeziehungen zu erfassen und zu verstehen, wurde eine qualitativ-explorative Forschungsmethode gewählt. Entlang eines halbstandardisierten Interviewleitfadens wurden hierfür sieben Experteninterviews mit IT-Führungskräften, überwiegend Chief Information Officers (CIO), die Outsourcingkunden sind, sowie mit Account Managern von IT-Dienstleistern geführt (vgl. Tabelle 1). Trotz der kleinen Sample-Größe der befragten Experten dienen die durchgeführten Expertengespräche als erster, explorativer Indikator, um die konzeptionelle Idee der Unterscheidung unter-

schiedlicher Beziehungsarchetypen im IT-Outsourcing zu unterstützen und in Richtung einer späteren Generalisierbarkeit der Ergebnisse weiterzuentwickeln.

Inhaltlich zielten die Interviews insbesondere auf drei Bereiche ab: aktuelle Geschäftsherausforderungen und Rahmenbedingungen für die Outsourcingentscheidung, Erwartungen des Kunden an den Provider sowie Erfahrungen aus der aktuellen Outsourcingbeziehung.

Interview-partner (Position)	Branche	Vertragsdauer	Erfahrung Gesprächspartner mit Outsourcing	Outsourcing-bereiche	Erwartungen gegenüber IT Provider
Kunde A Chief Information Officer (CIO)	Telekommunikation	Aktueller Vertrag seit 2004, davor anderer Anbieter seit 1998, Laufzeit 5 Jahre	>15 Jahre Erfahrung als Anbieter und als Kunde	Infrastruktur (Desktop, Laptop, WAN, LAN, Messaging, Telefonie, Server&Storage)	<ul style="list-style-type: none"> ▪ Experten Know How bei Technologie u. Prozessen ▪ Kostensenkung: Skalierungseffekte, Synergien ▪ Individuelle Anpassung auf Kunden ▪ Proaktive Vorschläge zur Effizienz ▪ IT-Unterstützung neuer Produkte
Kunde B Chief Information Officer (CIO)	Medien / TV	Aktueller Vertrag seit 2005, Laufzeit 4 Jahre, davor 3-Jahresverträge seit 1996	>4 Jahre	Infrastruktur (Netzwerk, WAN, SAP, Datenbanken), Call Center Application Hosting	<ul style="list-style-type: none"> ▪ flexibel auf Nachfrageveränderung / schnelle Verfügbarkeit (geschäftskritisch, da Livebetrieb TV) ▪ Ausrichtung d. Systeme auf Lastspitzen aufgrund dynamischer Endkundenbedarfe ▪ Anbieter muss Geschäft und Prozesse des Kunden verstehen (dynamisches Geschäft)
Kunde C Head of IT	Bank / Finanzen	Seit 2001, dann Insourcing, aktueller Vertrag seit 2004, Laufzeit 3 Jahre	>5 Jahre	Aktuell: SAP, Desktop, On-Site-Support, Help Desk Früher zusätzlich: Netzwerke, RZ	<ul style="list-style-type: none"> ▪ Kosteneinsparungen ▪ Professional Expertise: Zurückgreifen auf Expertenpool, State-of-the-Art-Know How ▪ Fit des Deliverymodells Anbieter und Kunde ▪ Flexible Anpassung an Geschäftsentwicklung
Kunde D Head IT-Management	Chemie	Seit 1996, Laufzeit 7 Jahre, aktueller Vertrag seit 2003, Laufzeit 5 Jahre	>4 Jahre	Netzwerk, SAP, spezielle Applikationen	<ul style="list-style-type: none"> ▪ Branchen Know How des Anbieters: Verständnis Geschäft des Kunden ▪ transparentes, auf Kunde angepasstes Vertragswerk mit spezifizierten Bedingungen (SLA) ▪ Fit des Deliverymodells Anbieter und Kunde
Kunde E Chief Information Officer	Luft- und Raumfahrt	Seit 1999, Laufzeit zweimal 3 Jahre, aktueller Vertrag 5 Jahre	>4 Jahre	Netzwerke (WAN, LAN), RZ, Applikationen, Desktop (total outsourcing)	<ul style="list-style-type: none"> ▪ proaktive Verbesserungsvorschläge für Prozesse und Kosten ▪ schlanker Prozess beim Anbieter ▪ kurze Reaktionszeit auf Kundenanforderungen
Anbieter F Account Manager	Kunde: Bank / Finanzen	Aktueller Vertrag seit 2004, Laufzeit 10 Jahre	n/a	IT-Systeme, spezielle Bankapplikationen	<ul style="list-style-type: none"> ▪ Transfer des Geschäftsrisikos des Kunden (schwankende Endkundenanforderungen im Transaktionsbereich) auf Anbieter ▪ Flexibilität: Pay-on-production ▪ innovates Delivery- und Abrechnungsmodell
Anbieter G Account Manager	Kunde: Gesundheit / Medizin	Aktuelle Verträge zwischen 1 und 3 Jahren	n/a	RZ-Infrastruktur, Applikationen (ASP)	<ul style="list-style-type: none"> ▪ Anbieter-Expertise ▪ Risikoabwälzung: Risiko-Nutzenteilung ▪ Gemeinsame Entwicklung von innovativen Services im ASP Medical Bereich

Tabelle 1: Übersicht der geführten Experteninterviews

3.2 Zentrale Ergebnisse der Experteninterviews

Während die befragten Experten über alle Interviews hinweg die Relevanz des Beziehungsaspektes innerhalb eines Outsourcingvorhabens betonten, unterschieden sich die Erwartungen an eine solche Outsourcingbeziehung und die Fähigkeiten des Anbieters erheblich. Diese Unterschiede machten sich insbesondere bei der Relevanz und Ausprägung der in Kapitel 2 vorgestellten Beziehungskategorien bemerkbar. Bei der nachfolgenden Beschreibung der Interviewinhalte ist zu berücksichtigen, dass die gewonnenen Erkenntnisse aus den Expertengesprächen sich teilweise nicht immer eindeutig nur einer Beziehungskategorie zuordnen ließen, sondern aufgrund der Interdependenz des Faktors in mehreren Kategorien eine Rolle spielen. Bei der

nachfolgenden Darstellung werden z.T. wörtliche Zitate der Interviewpartner eingeflochten, um den Argumentationspunkt zu verdeutlichen bzw. zu untermauern.

3.2.1 *Strategischer Kontext*

Im Hinblick auf Rahmenbedingungen und aktuelle Geschäftsherausforderungen betonten alle Interviewpartner die Dynamik ihres Geschäftsumfeldes und die daraus entstehende Notwendigkeit von Flexibilität und effizienten Geschäftsprozessen. Die Einschätzungen der Experten in Bezug auf die daraus entstehende Motivation und Erwartungen an den Provider unterschieden sich jedoch erheblich. Während drei der Befragten insbesondere Kosteneinsparungen und IT-Effizienz als zentrale Erwartung an die Outsourcingbeziehung formulierten, stand bei zwei weiteren Interviewpartnern explizit der Wunsch nach proaktiven Verbesserungsvorschlägen von Prozessen, Kosten und zur Entwicklung komparativer Wettbewerbsvorteile im Vordergrund.

„Mir ist wichtig, dass der Anbieter nicht im Akquisitionsmodus, sondern im Verbesserungsmodus auf uns zukommt. Der Outsourcer ist zu stark im Reaktionsmodus und nicht proaktiv. Besser machen bedeutet Aufwand für den Anbieter, aber der Outsourcer ist nur dann erfolgreich, wenn ich als Kunde erfolgreich bin. Ich als Kunde bin nur erfolgreich, wenn meine Prozesse effizient laufen. Also erwarte ich Vorschläge vom Anbieter.“ (CIO, Kunde A)

Hierfür wurde dem Outsourcing-Provider auch die entsprechende strategische Rolle und Rechtstruktur zugewiesen. Noch einen Schritt weiter ging die Erwartung im Kunden-Anbieterverhältnis G. Hier zielte der Kunde (geschildert aus der Perspektive des Anbieters) auf die gemeinsame Serviceentwicklung eines innovativen ASP-Modells in der Gesundheitsbranche ab, verbunden mit einer Risiko-Nutzen-Teilung.

3.2.2 *Beziehungsstruktur*

Mit Blick auf die Entwicklung unterschiedlicher Formen und Varianten im Outsourcing lässt sich auch bei den Gesprächspartnern der aktuelle Trend der selektiven Multi-Sourcing-Strategie feststellen. Fast alle Befragten haben mehr einen Outsourcingpartner beauftragt. Auch bei der Laufzeit der Verträge lässt sich eine Entwicklung konstatieren. Während zu Beginn aufgrund der geringen Erfahrung mit der Outsourcingpraxis und dem speziellen Anbieter eher kurzfristige Verträge abgeschlossen wurden, stieg mit wachsender Outsourcingerfahrung die Laufzeit der Verträge an. Je strategischer Outsourcing angesehen wurde, umso weniger spezifisch wurden die Ergebnisse bzw. der genaue Output der Outsourcingbeziehung festgelegt. Demgegenüber

wurde jedoch die klare Festlegung der Zuständigkeiten zwischen Anbieter und Kunde betont und die Notwendigkeit des Verständnisses und Branchen Know Hows auf Anbieterseite.

„Wichtig ist eine funktionierende Kompetenzaufteilung, dass alle an einem Strang ziehen. Ein gutes Dienstleisterverhältnis [ist zentral] und dass unser Geschäft läuft, das muss unser Dienstleister verstehen.“ (CIO, Kunde B)

3.2.3 Steuerungsmechanismen

In Anlehnung an die in der Literatur identifizierte Unterscheidung von formalen, ergebnisorientierten und informellen, verhaltensorientierten Steuerungsmechanismen finden sich diese Formen auch bei den befragten Experten. Während alle Gesprächspartner die Notwendigkeit eines transparent und klar formulierten Regelwerks als Grundlage der Outsourcingbeziehung hervorheben, steuern zwei der Interviewpartner jedoch bewusst über verhaltensbasierte Mechanismen, da ihrer Ansicht nach der Vertrag alleine nicht zum Erfolg der Outsourcingbeziehung ausreicht, sondern vielmehr ein kooperatives Management der Outsourcingbeziehung notwendig ist. Dies sind gleichzeitig die bereits oben genannten Kunden, die vom Anbieter proaktive Verbesserungsvorschläge erwarten und ihm eine entsprechend strategische Rolle einräumen.

„Der Vertrag ist grundlegend und wichtig. Aber wenn ich den Vertrag herausholen muss, ist es meistens schon zu spät. Dann läuft an anderer Stelle etwas schief.“ (CIO, Kunde E).

3.2.4 Interaktion

Im Hinblick auf den wechselseitigen Austausch beider Outsourcingparteien fokussierten die Interviewpartner insbesondere zwei Aspekte: das Machtverhältnis zwischen Outsourcinggeber und -nehmer sowie das Deliverymodell des Anbieters mit Blick auf den Kunden. Bei beiden Punkten spielt die in der Literatur bereits angesprochene Rolle der Ausgeglichenheit und des „Fits“ zwischen Anbieter und Kunde eine zentrale Rolle. Auf gleicher Augenhöhe zu verhandeln, wurde von mehreren Interviewpartnern als wichtiger Erfolgsfaktor für die Outsourcingbeziehung genannt.

„Das Deliverymodell des Outsourcers hat nicht gepasst. Die Prozesse und Ressourcen, die uns der „große“ Anbieter übergestülpt hat, haben nicht zu uns als mittelständischem Kunden gepasst, der Aufwand von Seiten des Outsourcers stand in keiner Relation zu den Kosten bzw. der aufgewendeten Zeit. So hat sich der Deal für unseren Anbieter nicht gelohnt, entsprechend hat das Interesse bei ihm nachgelassen.“ (Head of IT, Kunde C)

4 Ein Bezugsrahmen von Archetypen in IT-Outsourcingbeziehungen

In der Literatur gibt es bereits erste Ansätze zur Systematisierung von IT-Outsourcingbeziehungen [vgl. bspw. KeWv02; KRNR03; NRRC96]. Nam et al. [NRRC96] bspw. schlagen ein zweidimensionales Rahmenwerk vor und beschreiben vier Outsourcingttypen entlang der Dimensionen „Ausmaß der Substitution durch den Anbieter“ und „Strategischer Einfluss der IT“. Kern et al. [KeWv02] klassifizieren ebenfalls vier Typen von Outsourcingbeziehungen entlang zweier anderer Dimensionen: „Strategische Absicht“ und „Technische Fähigkeiten“. Obwohl diese ersten Ansätze in Richtung einer Typologisierung zu einem systematischeren Verständnis von Outsourcingbeziehungen beitragen, sind sie zu einseitig, da sie nur zwei ausgewählte, keinesfalls umfassende Beziehungsdimensionen einbeziehen [NRRC96]. Bisherige Typologien sind nicht in der Lage, eine Outsourcingbeziehung umfassend zu charakterisieren oder Managementempfehlung zur Steuerung derselben abzuleiten. Aufgrund der gestiegenen Komplexität und Variantenvielfalt im IT-Outsourcing müssen daher weitere Beziehungsfaktoren in die Analyse von Outsourcingbeziehungen einbezogen werden, um eine ganzheitliche Typologie von Outsourcingbeziehungen entwickeln zu können.

Die in den Experteninterviews erhobenen Erwartungen, Erfahrungen und Ausprägungen von Outsourcingbeziehungen können in fünf verschiedene Archetypen eingeteilt und entlang der aus der Literatur erarbeiteten Beziehungsdimensionen unterschieden werden: (1) „Commodity Provider“, (2) „Technologie Exzellenz Partner“, (3) „Stabilitätspartner“, (4) „Partnerschaft für kollaborative Serviceentwicklung“ und (5) „Strategischer Allianz- und Innovationspartner“.

Der aktuell in der Praxis vorrangig auftretende Typ ist die „Commodity Provider“-Outsourcingbeziehung. Ziel einer solchen Beziehungskonstellation ist die Erreichung einer effizienten IT verbunden mit Kosteneinsparungen oder -variabilisierung durch den selektiven Einkauf externer Ressourcen. Der Provider wird hier als Anbieter von standardisierten Leistungen (commodity, utility) angesehen, die aus einem Ressourcenpool je nach Bedarf und meist auf einer kurzfristigen Basis bezogen werden können. Die strategische bzw. partnerschaftliche Komponente mit Fokus auf langfristigen komparativen Wettbewerbsvorteilen steht nicht im Vordergrund. Geeignete Maßnahmen zur Steuerung dieses Outsourcingtyps sind formale, ergebnisorientierte Mechanismen, d.h. ein stark strukturiertes und mit detaillierten Steuerungs- und Outputgrößen formuliertes Vertragswerk. Entsprechend gering ist meist die Rolle von verhaltensorientierten Mechanismen wie Vertrauen, Commitment, oder informeller Informationsaustausch ausgeprägt.

Ein weiterer Beziehungstyp ist der „*Technologie Exzellenz Partner*“, bei dem der Anbieter aufgrund seines umfassenden Branchen Know How oder seiner „best-in-class“ Fähigkeiten im Hinblick auf eine zukunftsorientierte und innovative Ausrichtung in der technologischen Entwicklung ausgewählt wird [KeWv02]. Eine solche Beziehung ist meist auf eine eher kurzfristige, projektbasierte Dauer ausgerichtet und basiert ebenfalls überwiegend auf ergebnisorientierten Steuerungsmechanismen.

Der dritte Beziehungstyp „*Stabilitätspartner*“ erfordert ein hohes Maß an Commitment, Vertrauen und informellem Informationsaustausch auf beiden Seiten, da hier meist ein signifikanter Anteil der IT des Kunden für eine längere Dauer extern vergeben wird. Dementsprechend spielen informale verhaltensorientierte Steuerungsgrößen eine wichtige Rolle. Verglichen mit den ersten beiden Typen sind hier die Transaktionskosten ebenso wie die Kosten, die ein Wechsel des Providers mit sich bringt, verhältnismäßig hoch, da die Verbindung und Abhängigkeiten beider Partner voneinander sehr stark ist. Wichtig bei diesem Beziehungstyp ist die Abgrenzung zu strategischen Allianzen: Auch wenn Stabilitätspartnerschaften meist langfristig und vom Outsourcinggegenstand her großvolumig ausgerichtet sind, sind es meist keine strategische Allianzen, bei denen Innovation und Geschäftstransformation im Fokus sind [KRNR03]. Vielmehr steht bei der Stabilitätspartnerschaft die Zuverlässigkeit des Anbieters bei der dauerhaften Erbringung von Basis IT-Leistungen und seine unterstützende (meist reaktive) Rolle im Vordergrund.

Die „*Partnerschaft für kollaborative Serviceentwicklung*“ ist ein Beziehungstyp, der sich erst langsam in der Praxis entwickelt. Er ist weder ausschließlich von den Eigenschaften einer kurzfristigen commodity-Beziehung geprägt, noch hat er eine rein strategische Ausrichtung. Vielmehr kann er als selektive Transformationspartnerschaft verstanden werden, bei der der Fokus auf der gemeinsamen Entwicklung von IT-Services liegt, bei der beide Partner gleichermaßen Know How und Ressourcen einbringen, sowie Risiko und Nutzen teilen. Dementsprechend ist der Grad der gegenseitigen Abhängigkeit sehr hoch, zumal die gemeinsame Serviceentwicklung auch strategische Elemente im Sinne der Erreichung komparativer Wettbewerbsvorteile und Wissenszuwachs enthält. Die Steuerung einer solchen Beziehungsform erfolgt über verhaltensorientierte Maßnahmen, im Vertrag können die zu entwickelnden Services aufgrund der Dynamik und Unsicherheit über das Ergebnis schwer festgeschrieben werden.

Der Beziehungstyp „*Strategischer Allianz- und Innovationspartner*“ wird oft auch als „transformationales Outsourcing“ bezeichnet und zielt – zunächst ähnlich wie bei der gemeinsamen Ser-

viceentwicklung – auf die gemeinsame Entwicklung und eine Risiko-Nutzen-Teilung innerhalb der Partnerschaft ab und wird erfolgreich über verhaltensorientierte Mechanismen gesteuert. Im Gegensatz zur gemeinsamen Serviceentwicklung, die sehr selektiv und fokussiert auf einen Bereich abzielt, sind strategische Allianzen von Dauer und Ausrichtung sehr viel umfassender und ganzheitlicher angelegt. Nicht selten steht eine Geschäftstransformation im Vordergrund, die langfristiges beidseitiges Commitment, Vertrauen und eine entsprechende Rechtsstruktur für den Anbieter beinhaltet und eine hohe Erfahrung beider Partner mit Outsourcing erfordert.

Entlang den aus der Literatur abgeleiteten Beziehungskategorien und den explorativen Erkenntnissen aus der Empirie können die vorgestellten Archetypen nun umfassender charakterisiert werden. Obwohl einige der Archetypen bereits vereinzelt in der Literatur Erwähnung finden, war es bislang nicht möglich, diese unterschiedlichen Typen genau anhand eines detaillierten Bezugsrahmens zu charakterisieren. Darüber hinaus konnte ein weiterer Beziehungstyp aus der Empirie identifiziert werden, der bislang in der Literatur keine Erwähnung fand: Partnerschaft für kollaborative Serviceentwicklung.

Tabelle 2 gibt einen Überblick über die identifizierten Archetypen und beschreibt deren unterschiedliche Ausprägungen entlang der entwickelten Beziehungskategorien.

Archetyp	Commodity Provider	Technologie Exzellenz Partner	Stabilitäts-partner	Partnerschaft für kollaborative Serviceentwicklung	Strategischer Allianz- und Innovationspartner	Beispielhafte Referenzen für Dimension
Strategische Absicht und Kontext						
Strategische Absicht und Motivation für externe Leistungsvergabe	IT Effizienz, Kostenreduktion, -variabilisierung	IT Effizienz, Zugang zu best-of-breed Technologie und Know How	Business Value, Performanz, Stabilität, Fokus auf Kernkompetenzen	Business Value / Innovation, gemeinsames Service Engineering / Wissenstransfer	Geschäfts-transformation, Innovation, Risikoübernahme	[KeWv02], [KRNR03] [AISS04]
Strategischer Einfluss auf die IT des Kunden	⬇	⬆	⬇	⬇	⬆	[NRRC96], KeWv02; KRNR03]
Erwartungen an Rolle des Service Providers	standardisierte Leistungen, Zugang Ressourcenpool	Lösungsanbieter "best in class"	Stabiler / langfristiger Lösungs-provider	Gemeinsamer Partner für Serviceentwicklung	Strategischer Allianzpartner	[KeWv02], [KRNR03] [AISS04]
Einbindung des Service Providers in aktive Planung der Kunden-IT (Erwartung des Kunden)	⬇	⬇	⬇	⬆	⬆	[KRNR03]
Ausmaß des komparativen Wettbewerbsvorteils / Innovationsopportunities durch ITO (Erwartung)	⬇	⬇	⬇	⬆	⬆	[KRNR03]

Archetyp	Commodity Provider	Technologie Exzellenz Partner	Stabilitäts-partner	Partnerschaft für kollaborative Serviceentwicklung	Strategischer Allianz- und Innovationspartner	Beispielhafte Referenzen für Dimension
Potenzial der Weiterentwicklung der Outsourcingbeziehung (Evolution)	⬇	↔	↔	⬆	⬆	[KRNR03]
Einfluss auf die langfristige Strategie des Kunden	⬇	⬇	↔	↔	⬆	[KRNR03]
Beziehungsstruktur						
Ausmaß der gegenseitigen Abhängigkeit Barrieren/Höhe der Wechselkosten	⬇	⬇/↔	⬆	⬆	⬆	[NRRC96], [LeKi99] [GoCh05]
Gleichrichtung der Ziele zwischen Kunde und Anbieter	⬇	↔	↔	⬆	⬆	[East92], [LaWi03]
Zeithorizont / Dauer der Beziehung Vertragslaufzeit	Kurzfristig projektbasiert o. kont.	Kurzfristig projektbasiert	Langfristig kontinuierlich	Mittelfristig projektbasiert o. kont.	Langfristig kontinuierlich	[LeKi99], [KRNR03]
Relationale Netzwerkarchitektur (strategisch, erweitert, virtuell)	Virtuell, erweitert	Virtuell, erweitert	Erweitert, strategisch	Erweitert, strategisch	strategisch	[SaZm00]
Konkretionsgrad von Output / Ergebnis der Beziehung	Gut spezifiziert/ spezifizierbar	Gut spezifiziert/ spezifizierbar	Gut spezifiziert, aber dynamisch	Schwer spezifizierbar, dynamisch	Schwer spezifizierbar, dynamisch	[KRNR03]
Ausmaß der Substitution durch Anbieter	⬇	⬇	⬆	↔	⬆	[NRRC96]
Rolle der Erfahrung mit Outsourcing	⬇	⬇ / ↔	↔	↔	⬆	[LeKi99]
Steuerungsmechanismen (Governance und Monitoring)						
Steuerungs- und Kontrollmechanismen	Formal, ergebnisorientiert	Formal, ergebnisorientiert	Formal und informal, ergebnisorientiert	Informal und formal, verhaltensorientiert	Informal und formal, verhaltensorientiert	[KRNR03], [East92]
Rolle und Level der vertraglichen Kontrolle	⬆	↔ / ⬆	⬆	⬇	⬇	[KRNR03]
Konkretisierungsgrad der vereinbarten Leistungen	⬆	⬆	↔/⬆	⬇	⬇	[BeSc05]
Interaktion						
Relevanz informeller Informationskanäle	⬇	↔	↔	⬆	⬆	[KRNR03]
Informationsübergabe und Wissenstransfer	⬇	⬆	↔	⬆	⬆	[LeKi99], [KRNR03]
Anreize zur Teilung von Risiko und Nutzen	⬇	⬆	↔/⬆	⬆	⬆	[KRNR03], [LeKi99]
Ausmaß der Freiheitsgrade / Entscheidungsautorität des Anbieters	⬇	↔	⬆	⬆	⬆	[KRNR03]
Koordinationskosten und Intensität	⬇	↔	⬆	⬆	⬆	[KRNR03]
Commitment von Anbieter und Kunde	⬇	↔	⬆	⬆	⬆	[LeKi99], [GoCh05]
Rolle von Vertrauen	⬇	⬇	↔	⬆	⬆	[Kern97, LeKi99, GoCh05]
Notwendigkeit kultureller Ähnlichkeiten	⬇	↔	⬆	⬆	⬆	[LeKi99]

Legende: ⬆= hoch / extrem wichtig, ↔= mittel / wichtig, ⬇= niedrig / kaum wichtig

Tabelle 2: Typologie von Beziehungen im IT-Outsourcing

5 Fazit und Ausblick auf zukünftige Forschung

Der Beitrag dieser Forschungsarbeit liegt in der Identifikation, Klassifikation und Beschreibung unterschiedlicher Typen von IT-Outsourcingbeziehungen. Aus den empirischen und theoretischen Vorarbeiten können erste Gestaltungsprinzipien und Ansatzpunkte für das erfolgreiche Management von unterschiedlichen Beziehungen im IT-Outsourcing abgeleitet werden. In der Literatur vereinzelt erwähnte Typen konnten anhand des entwickelten Bezugsrahmens präzise klassifiziert und beschrieben werden. Darüber hinaus wurde ein neuer Typ identifiziert, der bisher in der Literatur keine Berücksichtigung gefunden hat. Ein wichtiger Erfolgsfaktor einer für beide Seiten erfolgreichen Outsourcingbeziehung scheint die Spezifikation und der Abgleich der unterschiedlichen Erwartungen an die Outsourcingbeziehung zu sein. Erste Erkenntnisse aus den Experteninterviews haben gezeigt, dass viele Unternehmen ihre Beziehung sehr stark auf Kosteneffizienz ausrichten und eng am Vertragswerk steuern, gleichzeitig aber Risikoteilung und Geschäftstransformation wie in einer strategischen Partnerschaft erwarten.

Für zukünftige Forschungsarbeiten eröffnen sich insbesondere zwei Forschungsfelder: Zum einen bietet die Typologie einen ersten Schritt in Richtung eines tiefer gehenden Verständnisses unterschiedlicher Erwartungen beider Parteien an eine Outsourcingbeziehung und deren Ausprägungen. Zum zweiten bildet der Vorschlag eine Grundlage für zukünftige positivistische Forschung, die die Existenz und Ausprägung der vorgeschlagenen Typen anhand großzahliger Empirie substantiell untermauert. Spannend hierbei ist die Untersuchung der Tendenzen in der Entwicklung unterschiedlicher Typen. Zwar ist der Typ des Stabilitätspartners das aktuell dominierende Modell in Outsourcingbeziehungen, es ist jedoch zu untersuchen, ob sich Outsourcingbeziehungen tendenziell zu den Rändern, also zu sehr spezifischen Typen im Outsourcing entwickeln, da Kunden entweder mehr Proaktivität und Engagement von Seiten des Anbieters oder aber die Erbringung standardisierter, kostengünstiger IT-Leistungen erwarten.

Literatur

- [AISS04] Alborz, S.; Seddon, P.B.; Scheepers, R.: Impact of Configuration on IT Outsourcing Relationships. In: Proceedings of the Tenth Americas Conference on Information Systems 2004. New York, S. 3551-3560.

- [BeSc05] Behrens, S.; Schmitz, C.: Ein Bezugsrahmen für die Implementierung von IT-Outsourcing-Governance. In: HMD - Praxis der Wirtschaftsinformatik 245 (2005), S. 28-36.
- [CoYo06] Cohen, L.; Young, A.: Multisourcing: Moving beyond outsourcing to achieve growth and agility. Harvard Business School Press, Boston, Mass., USA 2006.
- [DGHJ04] Dibbern, J.; Goles, T.; Hirschheim, R.; Jayatilaka, B.: Information Systems Outsourcing: A Survey and Analysis of the Literature. In: The DATA BASE for Advances in Information Systems 35 (2004) 4, S. 6-102.
- [East92] Easton, G.: Industrial Networks. A Review. In: Axelson, B.; Easton, G. (Hrsg.): Industrial Networks: A New View of Reality. Routledge, London 1992, S. 1-34.
- [GoCh05] Goles, T.; Chin, W.W.: Information systems outsourcing relationship factors. In: SIGMIS Database 36 (2005) 4, S. 47-67.
- [JaBK06] Jahner, S.; Böhmman, T.; Krcmar, H.: Anticipating and considering customers' flexibility demands in IS outsourcing relationships. In: Proceedings of the 14th European Conference on Information Systems 2006. Göteborg, Sweden.
- [Kern97] Kern, T.: The gestalt of an information technology outsourcing relationship: an exploratory analysis. In: Proceedings of the 18th International Conference on Information Systems 1997. Atlanta, GA, S. 37-58.
- [KeWi00] Kern, T.; Willcocks, L.: Exploring information technology outsourcing relationships: theory and practice. In: The Journal of Strategic Information Systems 9 (2000) 4, S. 321-350.
- [KeWv02] Kern, T.; Willcocks, L.P.; van Heck, E.: The Winner's Curse in IT Outsourcing: Strategies for avoiding relational trauma. In: California Management Review 44 (2002) 2, S. 47-69.
- [KiCh03] Kim, S.; Chung, Y.-S.: Critical success factors for IS outsourcing implementation from an organizational relationship perspective. In: Journal of Computer Information Systems 43 (2003) 4, S. 81-90.
- [Klep97] Klepper, R.: The management of partnering development in IS outsourcing. In: Willcocks, L.P.; Lacity, M.C. (Hrsg.): Strategic sourcing of information systems: Perspectives and practices. Wiley, Chichester 1997, S. 305-326.

- [KoMc90] Konsynski, B.R.; McFarlan, E.W.: Information partnerships - Shared data, shared scale. In: Harvard Business Review 68 (1990) 5, S. 115-120.
- [KRNR03] Kishore, R.; Rao, H.; Nam, K.; Rajagopalan, S.; Chaudhury, A.: A Relationship Perspective on IT Outsourcing. In: Communications of the ACM 46 (2003) 12, S. 87-92.
- [LaHi93] Lacity, M.C.; Hirschheim, R.A.: The Information Systems Outsourcing Bandwagon. In: MIT Sloan Management Review 35 (1993) 1, S. 73-86.
- [LaHi94] Lacity, M.; Hirschheim, R.: Realizing outsourcing expectations: incredible expectations, credible outcomes. In: Information Systems Management 11 (1994) 4, S. 7-18.
- [LaWi03] Lacity, M.; Willcocks, L.: IT Sourcing Reflections: Lessons for Customers and Suppliers. In: Wirtschaftsinformatik 45 (2003), S. 115-125.
- [LeKi99] Lee, J.-N.; Kim, Y.-G.: Effect of Partnership Quality on IS Outsourcing Success. In: Journal of Management Information Systems 15 (1999) 4, S. 29-61.
- [LHCP00] Lee, J.-N.; Huynh, M.Q.; Chi-wai, K.R.; Pi, S.-M.: The Evolution of Outsourcing Research: What is the Next Issue? In: Proceedings of the 33rd Hawaii International Conference on System Sciences-Volume 7 2000.
- [LuKN06] Luftman, J.; Kempaiah, R.; Nash, E.: Key Issues for IT Executives 2005. In: MIS Quarterly Executive 5 (2006) 2, S. 27-45.
- [NRRC96] Nam, K.; Rajagopalan, S.; Rao, H.R.; Chaudhury, A.: A two-level investigation of information systems outsourcing. In: Communications of the ACM 39 (1996) 7, S. 36-44.
- [oV05] o.V.: Outsourcing verfehlt Zielvorgaben. <http://www.computerwoche.de/index.cfm?pageid=254&artid=74335&linktype=nl>, Abruf am: 30.06.2006.
- [SaZm00] Sambamurthy, V.; Zmud, R.W.: Research Commentary: The Organizing Logic for an Enterprise's IT Activities in the Digital Era--A Prognosis of Practice and a Call for Research. In: Information Systems Research 11 (2000) 2, S. 105-114.

Qualität von IT-Leistungen aus den Perspektiven von Anbietern und Nachfragern

Ergebnisse einer Umfrage in der Schweiz

Roger Grütter

Dipl.-Inform.
CH – 3006 Bern
rgruetter@rgruetter.ch

Prof. Dr. Gerhard Schwabe

Institut für Informatik
Universität Zürich
CH – 8050 Zürich
schwabe@ifi.unizh.ch

Felix-Robinson Aschoff

Dipl.-Psych.
Institut für Informatik
Universität Zürich
CH – 8050 Zürich
aschoff@ifi.unizh.ch

Abstract

Dieser Beitrag stellt die Ergebnisse einer Umfrage zum IT-Service-Management in der Schweiz vor. Er vergleicht, wie Anbieter und Nachfrager die Qualität der IT-Services beurteilen. Insgesamt beurteilen sowohl Anbieter als auch Nachfrager die Dienstleistungsqualität mit dem Prädikat „gut“. Allerdings ist in der Bewertung der Dienstleistungsqualität zwischen den Anbietern und den Nachfragern ein signifikanter Unterschied auszumachen. Werden Qualitätsfaktoren in die Auswertung einbezogen, die die Dienstleistungsqualität beeinflussen, zeigt sich, dass die Nachfrager ein viel klareres Verständnis haben, was für sie eine gute Dienstleistungsqualität ausmacht. Die Anbieter hingegen tappen bei der Zusammensetzung der Dienstleistungsqualität zum großen Teil im Dunkeln. Weiter zeigt die Auswertung, dass die operativen Leistungen der Anbieter gut bewertet werden. Demgegenüber stehen aber schlechte

Bewertungen, und somit ein Verbesserungsbedarf, gegenüber den Aktivitäten des höheren Managements, die im Zusammenhang mit dem Dienstleistungsmanagement stehen.

1 Einleitung

IT- Dienstleistungen werden zunehmend professionell erbracht. Während in den 80er und 90er Jahren die IT-Abteilungen ihre Dienstleistungen noch ad-hoc und mit einer stark variablen Qualität erbrachten, wird heute zunehmend auf eine planbare, konstant hohe Dienstleistungsqualität gesetzt. IT ist ein Teil des Geschäfts geworden [Grem04] und deshalb werden in den Unternehmen aus Technologieanbietern Dienstleister [Sall04]. Die Dienstleistungsanbieter sollten dabei kundenorientiert und in Prozessen denken [BöKr04]. Erste Studien beobachteten in den letzten Jahren einen erheblichen Fortschritt der IT- Servicequalität durch die zunehmende Verbreitung des ITIL-Frameworks [Offi03a, Offi03b, Dete04, Schm04]. Solide Untersuchungen zur IT-Service-Qualität basieren aber auf Fallstudien [HoZB04b], während quantitative Daten zur Servicequalität für den deutschsprachigen Raum¹ bisher fehlen.

Deshalb interessierte uns in Zusammenarbeit mit dem Schweizer Verband der Informatik-anbieter (SWICO), wie gut IT-Dienstleistungen heute wirklich sind und welche Faktoren für die Beurteilung von IT-Dienstleistungen ausschlaggebend sind. Da Qualität von Dienstleistungen im größeren Umfang nicht mit vertretbarem Aufwand objektiv messbar ist, setzten wir auf die Wahrnehmung der Dienstleistungsqualität durch Service- Anbieter und Kunden. Dabei interessieren uns nicht nur die absolute Höhe der Dienstleistungsqualität, sondern auch die Differenzen zwischen den Wahrnehmungen beider Seiten.

2 Verwandte Arbeiten

Die Kundenbeteiligung, das zeitliche Zusammenfallen von Produktion und Konsum sowie die Immaterialität machen es schwierig, die Qualität von IT-Dienstleistungen objektiv zu messen. Deshalb spielen subjektive Faktoren, insbesondere die Erwartungen des Kunden, eine bedeutende Rolle. Gängige Modelle zur Messung von allgemeiner Dienstleistungsqualität [Cors01; ZeBP88] und IT-Dienstleistungsqualität (SERVQUAL) [BöKr04, PaZB85] stellen

¹ Uns sind keine internationalen Daten bekannt. Da Dienstleistungen und das Qualitätsempfinden kulturbedingt sind, ist eine Beschränkung auf einen Kulturkreis sinnvoll.

deshalb die Erwartungen und die Wahrnehmungen der Kunden sowie die Wahrnehmung der Kundenerwartung durch die Serviceanbieter ins Zentrum. Die Diskussion zu IT-Service-management wird in den letzten Jahren zunehmend durch die Einführung von ITIL in Unternehmen geprägt [Offi03a; Offi03b; ZaHB05; Grem04; ViGü05, HoZB04a, HoZB04b]. ITIL ist eine strukturierte Sammlung von „Best practices“ für die Gestaltung von IT-Services. ITIL hat sich in den letzten Jahren zu einem Industriestandard entwickelt, aus dem Beratungs- und Softwarefirmen eigene Servicemanagement-Frameworks ableiten (z.B. MOF von Microsoft [PuQA03] oder HP ITSM Reference Model von Hewlett-Packard [Hewl03]). Laut einer Studie von Beratungsunternehmen [Dete04] haben sich bei deutschen Unternehmen die Transparenz der IT-Prozesse, die Qualität der IT-Prozesse und die Reaktionszeiten bei über der Hälfte der befragten Unternehmen aufgrund der Einführung von ITIL verbessert. Gemäß einer weiteren Studie konnten gerade 11% der befragten mit ITIL signifikante Kosteneinsparungen realisieren [Schm04]. Dieser Zahl stehen Kostenerhöhungen bei 9% der Unternehmen gegenüber. Immerhin konnte bei 39% eine geringe Kostenreduzierung erreicht werden.

3 Forschungsdesign und Datenerhebung

Im Folgenden werden die Forschungsfragen eingeführt und die Vorgehensweise begründet. Sodann wird ein Modell der Dienstleistungsqualität entwickelt, aus dem dann der Fragebogen abgeleitet wird. Ein Abschnitt zur Datenerhebung schließt dieses Kapitel.

Forschungsfragen

Wie in der Einleitung bereits beschrieben, soll diese Studie zwei Forschungsfragen zur Dienstleistungsqualität beantworten.

Forschungsfrage 1: Wie beurteilen Anbieter und Nachfrager die Qualität von IT-Leistungen und wo sind die größten Unterschiede?

Die Bewertung der Dienstleistungsqualität ist das zentrale Merkmal dieser Studie und bietet eine Kennzahl, wie gut oder schlecht die Kunden, beziehungsweise die Anbieter die Dienstleistungsqualität im Schweizer Markt einschätzen. Es gilt zu beachten, dass die Kunden die Qualität der Anbieter bewertet haben. Die Anbieter hingegen wurden gefragt, welche

Bewertungen sie von den Kunden erwarten. Diese Ausgestaltung der Frage hatte zum Ziel, eine ehrliche Selbsteinschätzung der Anbieter zu erreichen und eine Überschätzung der eigenen Leistung zu verhindern. Wenn von der Qualität der Anbieter die Rede ist, gilt es zu bedenken, dass bei den Anbietern nicht direkt die eigene Qualität bewertet wurde, sondern die Qualitätsbewertung, die diese von den Kunden erwarten. Mit der erwarteten Dienstleistungsqualität auf Anbieterseite ist also diejenige Bewertung gemeint, die die Anbieter von ihren Kunden, aufgrund der erbrachten Leistungen, erwarten. Um genauere Aussagen zu erreichen wird die erste Forschungsfrage im Folgenden weiter aufgefächert.

Forschungsfrage 1.1: Wie beurteilen die Kunden die Dienstleistungsqualität ihrer Anbieter?

Forschungsfrage 1.2: Welche Beurteilung der Dienstleistungsqualität erwarten die Anbieter von ihren Kunden?

Forschungsfrage 1.3: Gibt es einen Unterschied und wenn ja wo?

Wie aus der Literatur ersichtlich ist [BöKr04] bestehen zwischen Kunden und Anbietern Unterschiede in der Wahrnehmung der Dienstleistungsqualität. Die Kunden als Abnehmer der Leistungen werden nicht jede kleine Unzufriedenheit in der Leistungserstellung mitteilen, beziehungsweise die Anbieter können die erwartete Dienstleistungsqualität aufgrund verschiedener Gaps [Cors01] nicht korrekt einschätzen. Dadurch tendieren die Anbieter zu einer besseren Beurteilung der Dienstleistungsqualität als es die Kunden tun.

Im Expertenworkshop wurden neben der Dienstleistungsqualität verschiedene Faktoren identifiziert, die, nach Meinung der Experten, einen relevanten Einfluss auf die Dienstleistungsqualität ausüben. Dies führt uns zu der zweiten Forschungsfrage, mit der die wichtigsten Einflussfaktoren der Dienstleistungsqualität identifiziert werden sollen.

Forschungsfrage 2: Welchen Einfluss haben die verschiedenen Einflussfaktoren auf die Dienstleistungsqualität?

Dieser zweite Frageblock betrachtet nicht nur die Unterschiede zwischen Kunden und Anbietern, sondern versucht die hauptsächliche Zusammensetzung der Dienstleistungsqualität bei Kunden und Anbietern zu erklären.

Forschungsfrage 2.1: Welche Einflussfaktoren tragen bei den Kunden wie stark zur Erklärung der Dienstleistungsqualität bei?

Forschungsfrage 2.2: Welche Einflussfaktoren tragen bei den Anbietern wie stark zur Erklärung der Dienstleistungsqualität bei?

Forschungsfrage 2.3: In welchen Faktoren bestehen die größten Unterschiede?

Ein abschließender Vergleich der jeweiligen Einflussfaktoren mit den größten Unterschieden soll das Verständnis der unterschiedlichen Zusammensetzung der Dienstleistungsqualität bei Kunden und Anbieter verbessern und aufzeigen, in welchen Einflussfaktoren die beiden Parteien unterschiedliche Auffassungen über den Einfluss der Faktoren auf die Dienstleistungsqualität haben.

Vorgehensweise

Parasuraman, Zeithaml und Berry entwickelten in den 80er Jahren den SERVQUAL- Ansatz [PaZB88], um die Dienstleistungsqualität aus Sicht der Kunden zu messen. Der SERVQUAL- Ansatz besitzt eine theoretische Einbindung in das Gap-Modell [PaZB85] und bewertet die Dienstleistungsqualität anhand der fünf Dimensionen Umfeld/Materielles, Zuverlässigkeit, Entgegenkommen, Souveränität, Einfühlung.

Eine Verwendung des SERVQUAL-Modells (bzw. des Nachfolgers SERVPERF [CrTa94], der nach früher Kritik [Hens90] entwickelt wurde) war in der vorliegenden Studie aus praktischen Gründen nicht möglich und aus Sicht der Forschungsfragen auch nicht unbedingt sinnvoll.

Die Untersuchung der Servicequalität zum IT-Servicemanagement in der Schweiz war ursprünglich vom Interesse des SWICO getrieben, mit einem Qualitätslabel gute Servicequalität auszuzeichnen und damit über eine höhere Markttransparenz insgesamt das Service-Niveau zu heben. Zwar war der Zweitautor von Beginn an der Konzeption dieser Untersuchung beteiligt, aber die entscheidende Frage der Erstellung des Untersuchungsdesigns wurde an eine Arbeitsgruppe aus Experten aus der Praxis und Wissenschaftlern delegiert². Für die Arbeit dieser Arbeitsgruppe standen die Relevanz der Fragen im Vordergrund und die Verwendung eines in der Wissenschaft etablierten Modells nur im Hintergrund. Wir entschlossen uns dennoch aus drei Gründen, die Daten mit zu erheben und wissenschaftlich auszuwerten:

1. Als Verband hat der SWICO sehr viel besseren Zugang zu Unternehmen als eine Universität. Dies sorgt für eine vergleichsweise gute Stichprobe.

² An diesem Prozess waren Erst- und Zweitautor nur als Reviewer beteiligt.

2. Die Kompetenz der einbezogenen Experten aus der Praxis war beeindruckend, wurde in einem systematischen Prozess einbezogen und der Designprozess war auch ausreichend wissenschaftlich begleitet.

3. Die Forschungsfragen sind explorativ: Es ging darum zu entdecken, welche Faktoren wie wichtig sind und wie sie in einem Zusammenhang stehen und nicht darum, Hypothesen zu bestätigen. Für eine solche Zielsetzung ist eine offener Vorgehensweise angemessen, zumal das SERVQUAL-Modell nicht unumstritten ist [Hens90].

Unsere Forschungsfragen und damit auch der gewählte Ansatz unterscheiden sich in zwei entscheidenden Punkt von SERVQUAL:

1. Es ging ausdrücklich um Spezifika des IT-Servicemanagements. Dies erfordert in jedem Fall eine Anpassung des allgemeinen Modells.

2. Wir gehen (wie [CrTa94]) davon aus, dass ein Kunde die komplexe Größe Dienstleistungsqualität mit einer Note bewerten kann, nicht zuletzt weil er bei der Auswahl weiterer Leistungen immer reflektieren muss, zu welcher Qualität der Anbieter bisher die Leistungen erbracht hat. Zudem möchten wir nicht die Unterschiede zwischen erwarteter und wahrgenommener Dienstleistungsqualität beim Kunden untersuchen. Sondern in dieser Studie interessiert die Differenz zwischen der Bewertung der Dienstleistungsqualität durch den Kunden, und die vom Anbieter erwartete Bewertung durch den Kunden.

Entwicklung eines Modells der Dienstleistungsqualität

Aus den oben vorgestellten Gap-Modellen der Dienstleistungsqualität [Cors01; ZeBP88; BöKr04] wurde die Grundidee übernommen, Erwartungen und Wahrnehmungen von Kunden und Anbietern zum Ausgangspunkt der Untersuchung zu machen. Da der Fragebogen bei der Aufnahme unserer Arbeit bereits weit fortgeschritten war, stand bei der Ausarbeitung des Modells die Kompatibilität zum Fragebogen im Vordergrund. Bestimmte Mängel im Fragebogen, die durch die Praxisorientierung des Fragebogens entstanden sind, mussten in Kauf genommen werden. So wird auf der Kundensicht die wahrgenommene Dienstleistungsqualität gemessen. Auf der Anbieterseite wurde abgefragt, welche Bewertung die Anbieter von der Kundenseite erwarten. Wir nennen das „erwartete Dienstleistungsqualität“. Damit wurde versucht eine ehrlichere Antwort von den Anbietern zu bekommen, als wenn die Frage gelautet hätte: „Wie beurteilen sie Ihre erbrachte Dienstleistungsqualität?“. Außerdem wird diese Vorgehensweise der Beobachtung gerecht, dass letztendlich immer der Kunde über die Qualität einer Dienstleistung entscheidet.

Die eigentlichen Messgrößen für Dienstleistungsqualität wurden in mehreren Workshops des SWICO mit jeweils 10-12 Teilnehmern aus Wissenschaft³ und Praxis erarbeitet. Das für die Umfrage aufgestellte Dienstleistungsqualitätsmodell umfasst 14 Qualitätsfaktoren (vgl. Abb. 1). Zentrales Element ist die erwartete Dienstleistungsqualität bei den Anbietern und die wahrgenommene Dienstleistungsqualität bei den Kunden. Zwischen diesen beiden Qualitätsbeurteilungen besteht ein Unterschied, der die Differenz in der Bewertung der Dienstleistungsqualität zwischen beiden Seiten ausdrückt. Die Gesamtbeurteilung der Dienstleistungsqualität ist auf beiden Seiten von den 14 Qualitätsfaktoren abhängig.

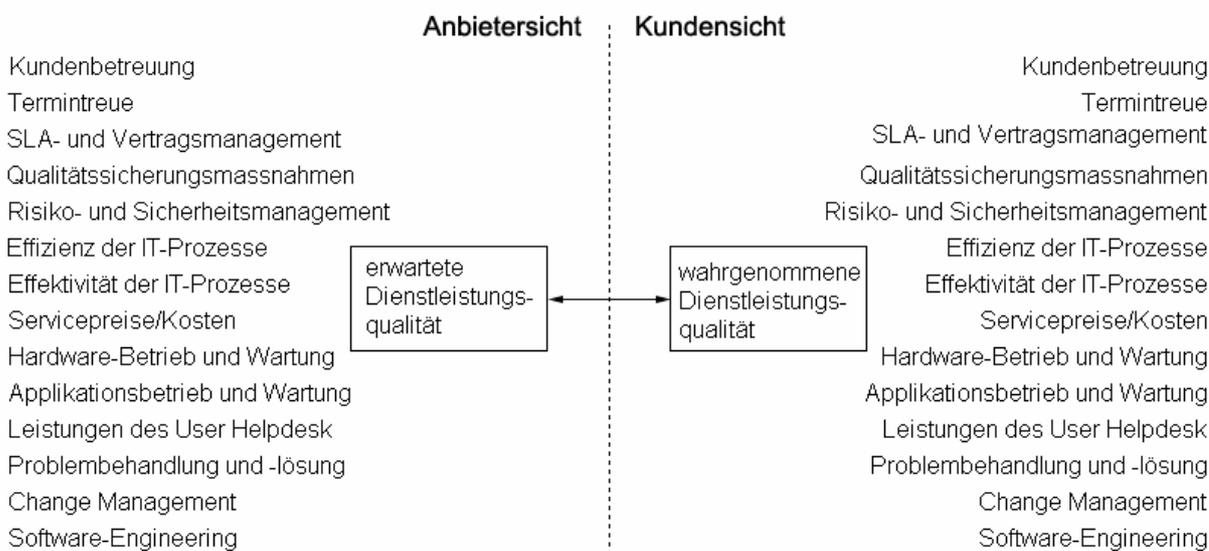


Abb. 1: Modell der Dienstleistungsqualität

Die 14 Qualitätsfaktoren, für welche ein Zusammenhang zur Dienstleistungsqualität vermutet wird, werden im Folgenden kurz vorgestellt. Das erste Merkmal „Kundenbetreuung“ soll allgemein zeigen, wie gut die Anbieter die Bedürfnisse der Kunden erkennen und darauf eingehen. „Termintreue“ ist für den Kunden ein wichtiger Aspekt bei der Bewertung der Dienstleistungsqualität. In ITIL werden die Leistungen oft in Form von Service Level Agreements (SLA) vereinbart, welche durch das Merkmal „SLA- und Vertragsmanagement“ bewertet werden. Das Merkmal „Qualitätssicherungsmaßnahmen“ ermöglicht eine Bewertung der Maßnahmen für eine standardisierte dauerhaft gesicherte Qualität der IT. Das „Risiko- und Sicherheitsmanagement“ zeigt, wie diese, in der Datenverarbeitung wichtige Leistung, ausgeführt wird. Die beiden Merkmale „Effizienz der IT-Prozesse“ und „Effektivität der IT-Prozesse“ bewerten die Produktivität, beziehungsweise die Wirksamkeit der vorhandenen IT-

³ An dieser Stelle möchten wir uns beim Kollegen Prof. Dr. Gerhard Knolmayer herzlich für die Mithilfe an der Fragebogenentwicklung bedanken.

Prozesse. Eine höhere Qualität ist oft mit höheren Kosten verbunden, daher wird beim Faktor „Servicepreise/ Kosten“ nach der Zufriedenheit bei den Kosten nachgefragt. „Hardware-Betrieb und Wartung“ ist eine standardisierte Leistung die von vielen Kunden konsumiert, beziehungsweise von vielen Anbietern angeboten wird. Softwareseitig wird im Faktor „Applikationsbetrieb und Wartung“ die Leistung bewertet, wie die vorhandene Software funktioniert und deren Wartung realisiert ist. Das User Helpdesk ist für den Kunden mit Problemen eine wichtige, oft einzige, Anlaufstelle und dementsprechend wichtig sind die „Leistungen des User Helpdesk“. Wie der Anbieter mit Problemen umgeht und in Lösungen transferiert, wird beim Faktor „Problembehandlung und -lösung“ bewertet. Das „Change Management“ ist eine weitere sehr wichtige Leistung eines IT-Serviceanbieters, bei der der korrekte Ablauf der Geschäftsprozesse direkt betroffen ist. Als letztes Merkmal der Qualitätsfaktoren wird im Faktor „Software-Engineerin“ die Erstellung neuer Software oder projektbezogene Leistungen bewertet.

In dem Fragebogen wurden Anbietern und Kunden identische Fragen zur Bewertung der einzelnen Qualitätsfaktoren sowie zur Bewertung der Gesamtqualität vorgelegt. Zur Bewertung der Zufriedenheit der einzelnen Merkmale wurde die Schweizer Notenskala verwendet, wobei die Note 1 unbrauchbar, die Note 6 sehr gut bedeutet. Da einzelne Unternehmen eine dem Qualitätsfaktor zugrunde liegende Dienstleistung gar nicht in Anspruch nehmen und damit nicht bewerten konnten, wurde eine zusätzliche Kategorie geschaffen, deren Werte in der Auswertung ausgeschlossen werden. Der Fragebogen enthielt noch eine Reihe weiterer Fragen, die für diesen Artikel nicht von Belang sind.

Datenerhebung

Mit der Erhebung der Daten wurde Anfangs Februar 2005 begonnen. Zum einen wurde der Fragebogen als Webumfrage verfügbar gemacht. Nur eine kleine Minderheit nutzte diesen Kanal. Weitere Daten wurden bei Workshops des SWICO zum Thema IT-Service-Management, über telefonische Direktansprache zufällig ausgewählter Unternehmen erhoben. Bis Mitte März 2005, dem Ende der Umfrage, konnten von den Kunden 95 Fragebögen und von den Anbietern 121 Fragebögen gesammelt werden. Von den Anbietern mussten fünf Fragebögen aufgrund Doppelerfassungen oder eines ungenügenden Ausfüllungsgrades von der Auswertung ausgeschlossen werden. Bei den Kunden mussten vier Fragebögen ausgeschlossen werden. Insgesamt lag die Zahl der ausgelassenen Antworten für die folgenden Auswertungen unter 12%.

Eine Generalisierbarkeit der Resultate ist nicht grundsätzlich gegeben, wenn die Grundgesamtheit aus allen Unternehmen, beziehungsweise allen Unternehmen die IT Leistungen anbieten, besteht, da die Mikro-Unternehmen in dieser Umfrage zu wenig stark berücksichtigt wurden [Bund01a]. Betrachtet man jedoch nicht die prozentuale Verteilung der natürlichen Klasse Unternehmensgröße [Bund01b], sondern die wirtschaftliche Relevanz der Unternehmen, können, nach Meinung der Autoren, die Ergebnisse auf die Grundgesamtheit der kleinen bis großen Unternehmen generalisiert werden. Auf jeden Fall können durch generalisierende Aussagen, Tendenzen im Schweizer Markt erkannt werden, die vor allem die „kleinen“ bis „großen“ (also nicht die sehr kleinen) Unternehmen betreffen.

4 Ergebnisse

Im Folgenden werden die Ergebnisse der Umfrage vorgestellt. Zuerst wird die Gesamtbeurteilung der Dienstleistungsqualität von Anbietern und Kunden vorgestellt und auf Unterschiede eingegangen. Der zweite Abschnitt behandelt daraufhin die durchschnittlichen Beurteilungen der 14 Qualitätsfaktoren durch Anbieter und Kunden. Der dritte Abschnitt analysiert die Bewertungsunterschiede beider Gruppen und im abschließenden vierten Abschnitt wird der Einfluss der Qualitätsfaktoren auf die Gesamtbeurteilung der Dienstleistungsqualität ermittelt.

Gesamturteil Dienstleistungsqualität: Unterschiede zwischen Anbietern und Kunden

Die Einschätzungen für das Fragebogen-Item „Generelle Beurteilung der Dienstleistungsqualität“ wurden getrennt für Anbieter und Kunden ausgewertet. Die Einschätzung der Anbieter auf der 6-Punkte-Likert Skala erreichte einen Mittelwert von $M=4,99$ ($SD=0,519$) während der Mittelwert der Einschätzungen der Kunden bei $M=4,71$ ($SD=0,694$) lag. Die statistische Signifikanz dieses Unterschieds wurde mit einem t-Test für unabhängige Stichproben geprüft. Der Unterschied erwies sich als hochsignifikant mit $t(203) = 3,36$; $p < 0,01$. Eine Voraussetzung des t-Test, nämlich die Homogenität der Varianzen, ist für unseren Datensatz nicht gegeben. Es kann jedoch davon ausgegangen werden, dass der Test bei dieser großen Stichprobe robust gegenüber dieser Verletzung reagiert (vgl. [RFHN04]) Der t-Test mit korrigierten Freiheitsgraden wird ebenfalls hochsignifikant mit $t(157,528) = 3,22$; $p < 0,01$. Die Effektgröße

beläuft sich auf $d = 0,462$. Für den Unterschied zwischen Anbietern und Kunden ergibt sich somit gemäß den Konventionen von [Cohe88] ein mittlerer Effekt.

In Bezug auf Forschungsfrage 1 lässt sich somit sagen, dass die Anbieter die generelle Dienstleistungsqualität auf der vorgegeben Skala als „gut“ einschätzen, während die Einschätzung der Kunden statistisch signifikant ca. $\frac{1}{4}$ Notenpunkt niedriger liegt.

Beurteilung der Qualitätsfaktoren

Die Urteile für die 14 Qualitätsfaktoren wurden getrennt für Anbieter und Kunden ermittelt. Die Mittelwerte aller 14 Items sind in Abbildung 2 abgetragen, wobei die Qualitätsfaktoren absteigend nach Kundeneinschätzungen sortiert sind.

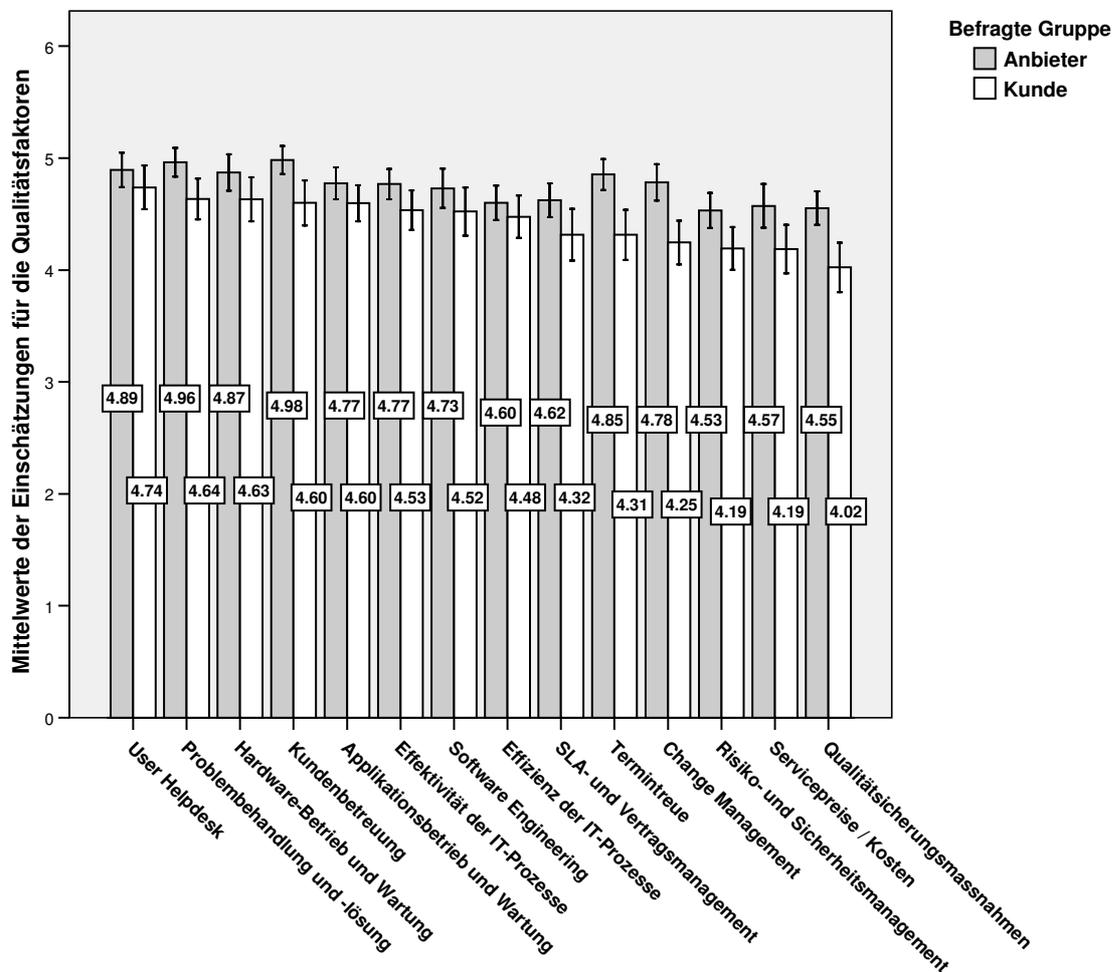


Abb. 2: Mittlere Einschätzung der 14 Qualitätsfaktoren von Anbietern und Kunden (absteigend sortiert nach der Höhe der mittleren Urteile durch die Kunden; Fehlerbalken: 95% Vertrauensintervall)

Die Einschätzungen der Anbieter schwanken zwischen $M = 4,98$ und $M = 4,53$ bei einem Gesamtmittelwert von $M = 4,77$ ($SD = 0,788$). Die höchsten Bewertungen erhielten die Faktoren „Kundenbetreuung“ ($M = 4,98$), „Problembehandlung und -lösung“ ($M = 4,96$) sowie „Leistungen des User Helpdesk“ ($M = 4,89$). Am schlechtesten beurteilt wurden die Items „Risiko und Sicherheitsmanagement“ ($M = 4,53$), „Qualitätssicherungsmaßnahmen“ ($M = 4,55$) sowie die „Effizienz der IT Prozesse“ ($M = 4,60$).

Die Einschätzungen der Kunden schwanken zwischen $M = 4,74$ und $M = 4,02$ bei einem Gesamtmittelwert von $M = 4,44$ ($SD=0,913$). Die höchsten Bewertungen erhielten die Faktoren „Leistungen des User Helpdesk“ ($M = 4,74$), „Problembehandlung und -lösung“ ($M = 4,64$) sowie „Hardware-Betrieb und Wartung“ ($M = 4,63$). Am niedrigsten eingestuft wurden die Items „Qualitätssicherungsmaßnahmen“ ($M = 4,02$), „Servicepreis / Kosten“ ($M = 4,19$) sowie „Risiko- und Sicherheitsmanagement“ ($M = 4,19$).

Unterschiede zwischen Anbietern und Kunden

Bei den Differenzen zwischen den Urteilen zwischen den Anbietern und den Kunden für die 14 Qualitätsfaktoren haben sich die größten Differenzen ergeben für die Items „Termintreue“ ($\text{Diff.}=0,54$), „Qualitätssicherungsmaßnahmen“ ($\text{Diff.}=0,53$) sowie „Change Management“ ($\text{Diff.}=0,53$). Diese drei Faktoren weisen alle ein ähnlich hohes Niveau auf. Die nächst kleinere Differenz weist nur noch eine Differenz von $\text{Diff.}=0,38$ („Kundenbetreuung“) auf.

Die größte Einigkeit herrschte bezüglich der Items „Effizienz der IT Prozesse (Produktivität)“ ($\text{Diff.}=0,12$) sowie „Leistungen des User Helpdesk“ ($\text{Diff.}= 0,15$). Die Differenzen zwischen Anbietern und Kunden für die 14 Qualitätsfaktoren sind in Abbildung 3 abgetragen. Um zu testen, ob die Anbieter die Servicequalität über alle 14 Qualitätsfaktoren hinweg signifikant höher einschätzen als die Kunden wurde eine 2×14 Varianzanalyse mit den beiden Faktoren Personengruppe (Anbieter / Kunde) und Frageitems (1 - 14) berechnet. Der Faktor Personengruppe erwies sich als signifikant ($SS=20,44$; $df=1$; $MS=20,443$; $F=5,977$ $p<.05$). Die Effektstärke dieses Faktors erwies sich mit $\omega^2=0,30$ als hoch (vgl. [Cohen88]).

Bezüglich Forschungsfrage 1.3 lässt sich demnach festhalten, dass es über alle 14 Qualitätsfaktoren hinweg eine signifikant höhere Einschätzung der Dienstleistungsqualität seitens der Anbieter im Vergleich zu den Kunden gibt. Es gibt aber deutlich unterschiedliche Diskrepanzen bei den 14 Qualitätsfaktoren.

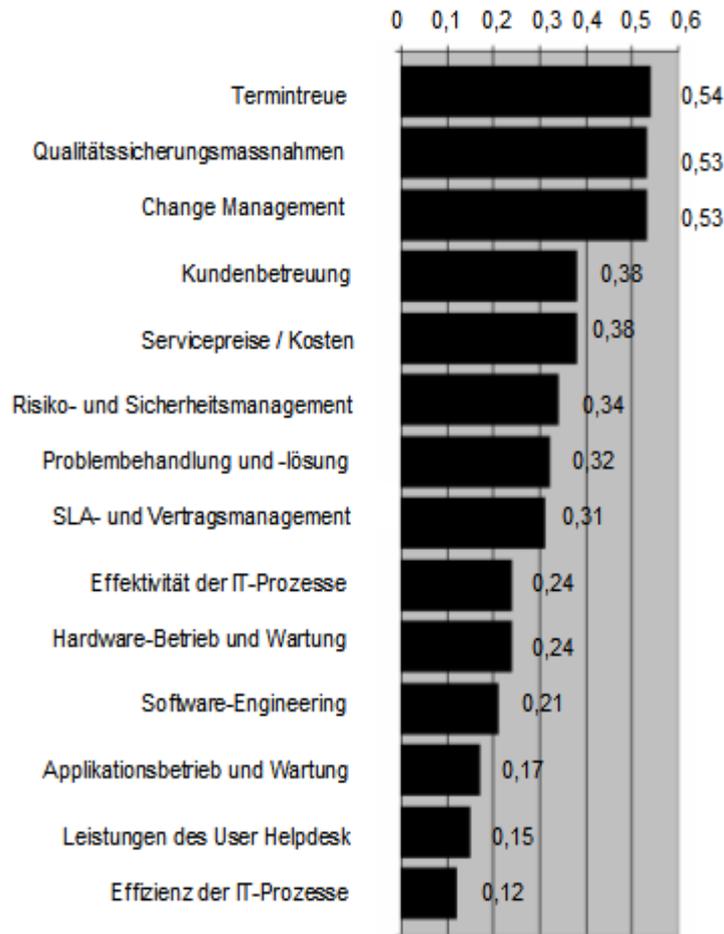


Abb. 3: Differenzen zwischen den Einschätzungen der Anbieter und der Kunden für die 14 Qualitätsfaktoren

Einfluss der Qualitätsfaktoren auf die Gesamteinschätzung der Dienstleistungsqualität

Um den Einfluss der eingeschätzten Qualitätsfaktoren auf die Gesamteinschätzung der Dienstleistungsqualität zu ermitteln, wurde für Anbieter und Kunden getrennt eine Lineare Regression mit paarweisem Fallausschluss berechnet. Die 14 Einzelfaktoren fungierten als Prädiktor und das Item „Generelle Beurteilung der Servicequalität“ wurde als abhängige Variable des Regressionsmodells deklariert. Werden alle 14 Einzelfaktoren in das Regressionsmodell für die Anbieter aufgenommen, ergibt sich ein $R^2=0,405$. Für die Kunden ergibt sich ein entsprechendes $R^2=0,561$.

Tab. 1 zeigt die standardisierten Beta-Koeffizienten für die Qualitätsfaktoren für Anbieter sowie für Kunden.

Für die Anbieter ergibt sich ein statistisch signifikanter Zusammenhang in diesem Regressionsmodell zwischen dem Item „Effektivität der IT-Prozesse“ und der generellen

Einschätzung der Dienstleistungsqualität. Für die Kunden erweist sich der Zusammenhang zwischen dem Item „Kundenbetreuung“ und dem Gesamturteil als statistisch signifikant.

	Anbieter	Kunden
	R ² =0,405	R ² =0,561
	Standardisierter Koeffizient Beta	Standardisierter Koeffizient Beta
Kundenbetreuung	0,164	0,382 **
Termintreue	0,170	0,099
SLA- und Vertragsmanagement.	0,108	0,127
Qualitätssicherungsmassnahmen	0,088	0,043
Risiko- und Sicherheitsmanagement	-0,077	0,029
Effizienz der IT-Prozesse	0,098	0,140
Effektivität der IT-Prozesse	0,381 **	-0,028
Servicepreise / Kosten	-0,028	0,177
Hardware-Betrieb und Wartung	0,088	-0,019
Leistungen des User Helpdesk	-0,150	0,144
Problembehandlung und -lösung	-0,034	-0,075
Change Management	-0,022	-0,067
Applikationsbetrieb und Wartung	-0,019	0,112
Software-Engineering	0,003	-0,030

Sign.: * <0,1 ** <0,05 ***<0,01

Tab. 1: Standardisierte Beta-Koeffizienten für die 14 Einzelfaktoren für Anbieter und Kunden (abhängige Variable: „Generelle Beurteilung der Servicequalität“)

Tabelle 1 zeigt auf, dass es sich nicht als sinnvoll erweist, alle Variablen in das Regressionsmodell aufzunehmen. Eine Reihe von Fragen weisen keinen statistisch signifikanten Zusammenhang mit der abhängigen Variablen auf und es kann zudem davon ausgegangen werden, dass ein gewisser Grad an Multikollinearität vorliegt (im Modell der Anbieter liegen acht Konditionsindizes über 30, im Model der Kunden liegen sechs Indizes über 30).

Mittels Ausschlussverfahren (Backward-Deletion) (vgl. [Bort05]) wurde Variable um Variable aus dem Regressionsmodell entfernt mit dem Kriterium, dass nur solche Variablen verbleiben, die mindestens eine Signifikanz von $\alpha = 0,10$ aufweisen. Daraus ergibt sich für die Anbieter ein Modell mit den beiden Variablen „Termintreue“ sowie „Effektivität der IT-Prozesse“ welches ein $R^2 = 0,341$ aufweist. Für die Kunden ergibt sich ein Regressionsmodell von $R^2 = 0,519$, das folgende Einzelfrageitems als Prädiktoren enthält: „Kundenbetreuung“, „SLA- und Vertragsmanagement“ sowie „ Servicepreise / Kosten“. Tabelle 2 zeigt die Beta-Koeffizienten des Modells für die Anbieter sowie des entsprechenden Modells für die Kunden.

Bezüglich Forschungsfrage 2 lässt sich festhalten, dass bei den Anbietern zwei Faktoren ausreichen, um einen großen Teil der Erklärungskraft des Gesamtmodells zu erreichen. Bei den Kunden sind es drei Faktoren. Für die Anbieter weist neben dem Faktor der „Termintreue“ der hochsignifikante Faktor der „Effektivität der IT-Prozesse“ die stärkste Korrelation mit der

Gesamtbeurteilung auf. Aus Sicht der Kunden ergibt sich hier der hochsignifikante Faktor „Kundenbetreuung“ als Faktor mit größtem Beta-Gewicht. Hinzu kommen die Faktoren „SLA- und Vertragsmanagement“ sowie der Faktor „Servicepreise / Kosten“. Zudem fällt auf, dass das Modell der Kunden einen höheren Teil der Varianz aufklärt als das Modell der Anbieter.

	Anbieter	Kunden
	R ² =0,341	R ² =0,519
	Standardisierter Koeffizient Beta	Standardisierter Koeffizient Beta
Kundenbetreuung	-	0,494 ***
Termintreue	0,249**	-
Effektivität der IT-Prozesse	0,439 ***	-
SLA- und Vertragsmanagement	-	0,206 *
Servicepreise / Kosten	-	0,204 *
Sign.: * <0,1 ** <0,05 ***<0,01		
– nicht in das Modell aufgenommen		

Tab. 2: Regressionsmodell nach Ausschlussverfahren für Anbieter und Kunden

5 Interpretation der Ergebnisse

Schlechte Servicequalität wird von einzelnen Praktikern als ein Hauptgrund dafür genannt, IT-Outsourcing-Geschäfte zu überdenken. Die vorliegende Untersuchung kann diese negative Einschätzung für die Schweiz nicht bestätigen. Die Anbieter schätzen die generelle Servicequalität als „gut“ ein (M=4,99) während die Kunde sie im Durchschnitt ¼ Note niedriger und damit noch fast gut einschätzen. In der Einschätzung gibt es keinen sehr großen, aber einen statistisch eindeutig signifikanten Unterschied in der Wahrnehmung zwischen Anbieter und Kunden. Die Anbieter überschätzen die Zufriedenheiten der Kunden. Dieser Umstand zeigt sich auch bei allen 14 Qualitätsfaktoren durchgehend. Es gibt keinen Qualitätsfaktor, der von den Kunden höher bewertet wird als von den Anbietern. Bei den drei Items „Termintreue“, „Qualitätssicherungsmaßnahmen“ und „Change Management“ beträgt diese Überschätzung immerhin ½ Notenpunkt. Alle drei Punkte gehören auch zu den Punkten, die von den Kunden mit am schlechtesten bewertet wurden. Hier ist also ein großer Handlungsbedarf bei den Anbietern. Die große Bewertungsdifferenz deutet darauf hin, dass den Anbietern dieser Handlungsbedarf noch nicht bewusst geworden ist. Für die weitere Analyse der Ergebnisse lohnt es sich, die einzelnen Faktoren zu gruppieren: Ein Gruppe von Faktoren beschreibt operatives IT Service Management, die in ITIL dem Service Support zuzuordnen wären. Unter diese Faktoren können „Leistungen des User Helpdesk“, „Problembehandlung

und -lösung“, „Hardware-Betrieb und Wartung“, „Applikationsbetrieb und Wartung“, „Software-Engineering“ und „Change Management“ zusammengefasst werden. Mit den allermeisten dieser Faktoren sind die Kunden vergleichsweise zufrieden und die Einschätzungen von Kunden und Anbietern gehen nicht weit auseinander. Nur bei dem (wohl anspruchsvollsten) Thema Change und Releasemanagement sind die Kunden weniger zufrieden und die Anbieter haben den Handlungsdruck noch nicht so wahrgenommen. Eine zweite Gruppe von Qualitätsfaktoren beschreibt (höhere) Managementaktivitäten, die in ITIL eher dem Service Delivery und dem allgemeinen Management zuzuordnen wären. Hierzu gehört „Kundenbetreuung“, „SLA- und Vertragsmanagement“, „Risiko- und Sicherheitsmanagement“ und „Qualitätssicherungsmaßnahmen“. In diesem Bereich sind die Kunden vergleichsweise unzufrieden. Den Anbietern ist diese Unzufriedenheit nicht bewusst. Eine dritte Gruppe von Qualitätsfaktoren beschreibt nichtfunktionale Anforderungen. Hierzu gehören Effizienz und Effektivität der IT-Prozesse, Termintreue sowie Servicepreis/Kosten. Diese Faktoren sind eher im mittleren Bereich zu finden. Es fällt dabei auf, wie sehr die Anbieter Ihre Termintreue überschätzen und wie unzufrieden die Kunden mit den Preisen sind, sind doch niedrigere Preise ein klassischer Anreiz zum Outsourcing. Dies führt zu der ersten Schlussfolgerung unserer Studie:

Schlussfolgerung 1: Während die Anbieter im operativen IT-Servicemanagement gute Leistungen anbieten, besteht noch Handlungsbedarf bei den höheren Managementaktivitäten. Dieser Handlungsbedarf ist von den Anbietern noch nicht erkannt.

Bei den Regressionsgleichungen klärt das Modell der Anbieter weniger Varianz auf (41%) als das Modell der Kunden (56%). Dies deutet auf eine klarere Wissensstruktur seitens der Kunden hin. Die Kunden wissen mit einem hohen Maß an Zuverlässigkeit welche der Qualitätsfaktoren die Dienstleistungsqualität für sie beeinflusst. Sie nutzen sowohl für die generelle Beurteilung der Dienstleistungsqualität als auch für die einzelnen Qualitätsfaktoren die Likert-Skala stärker aus, was unter anderem zu dem Effekt der erhöhten Varianzaufklärung beiträgt. Mit drei Items („Kundenbetreuung“, „SLA- und Vertragsmanagement“, „Servicepreise / Kosten“) klärt das Regressionsmodell der Kunden über 50% der Varianz der Gesamtbeurteilung auf. Es fällt auf, dass diese Faktoren alle direkt an der Kundenschnittstelle zu finden sind, d.h. es sind Faktoren, deren Erfüllung das verantwortliche Management der Kunden unmittelbar erfährt.

Schlussfolgerung 2: Das Gesamtqualitätsempfinden der Kunden wird derzeit stark durch drei Faktoren an der Schnittstelle zu dessen verantwortlichen Management geprägt.

Dieses Ergebnis ist aus Sicht der Kunden logisch, da sich dessen verantwortliches Management um interne Qualitätsfaktoren eigentlich nicht kümmern möchte und an der operativen Nutzerschnittstelle Zufriedenheit herrscht. Bei den eigentlichen kundengerichteten Prozessen („Kundenbetreuung“, „Problembehandlung und –lösung“) leisten die Anbieter heute schon gute Arbeit; bei den Nebenbedingungen dieser Aktivitäten, den Servicepreisen und Kosten und der Termintreue besteht aber aus Sicht der Kunden noch erheblicher Handlungsbedarf.

Die Anbieter hingegen haben offenbar eine weniger genaue Vorstellung, wie sich die Zusammenhänge für die Kunden darstellen. Selbst bei einer Regression mit allen 14 Faktoren werden hier nur knapp 40% Varianzaufklärung erreicht. Dieser Effekt ist sicher zum Teil auch darauf zurückzuführen, dass sie sich die Anbieter die Erwartungen der Kunden vorstellen müssen, während diese eine konkrete Situation beurteilen.

Schlussfolgerung 3: Die Anbieter haben kein klares Bild davon, was genau das Qualitätsempfinden der Kunden prägt.

Während für die Anbieter die „Effektivität der IT-Prozesse“ das höchste Beta-Gewicht aufweist, gefolgt von dem Item „Termintreue“, stehen für die Kunden die „Kundenbetreuung“ sowie das „SLA- und Vertragsmanagement“ und die „Servicepreise / Kosten“ im Vordergrund. Für sie scheint eine möglicherweise etwas abstrakte Effektivität der IT Prozesse nicht so entscheidend zu sein (eventuell auch weil ihnen der direkte Vergleich fehlt). Was sie jedoch unmittelbar merken und was sich somit stark auf ihre Gesamtbeurteilung auswirkt ist zum einen das Eingehen auf ihre Bedürfnisse als Kunden und zum anderen die entstehenden Kosten.

Die Faktoren ergeben eine gute erste Schätzung der Zusammenhänge, die allerdings in weiteren Studien und an weiteren Stichproben auf ihre Stabilität hin untersucht werden sollten.

Literaturverzeichnis

- [Bort05] Bortz, Jürgen: Statistik. 6. Auflage. Springer Medizin Verlag, Heidelberg 2005
- [BöKr04] Böhmann, Tilo; Krcmar, Helmut: Grundlagen und Entwicklungstrends im IT-Servicemanagement. In: Meier, Andreas (Hrsg.): HMD - Praxis der Wirtschaftsinformatik 237. Dpunkt Verlag, Heidelberg 2004, S. 7–21.

- [Bund01a] Bundesamt für Statistik: Taschenstatistik Betriebszählung 2001. <http://www.bfs.admin.ch/bfs/portal/de/index.html>, Abruf am 2005-03-22.
- [Bund01b] Bundesamt für Statistik: Taschenstatistik der Schweiz 2004. <http://bfs.admin.ch/bfs/portal/de/index.html>, Abruf am 2005-03-22.
- [Cohe88] Cohen, J. Statistical power analysis for the behavioral sciences. Hillsdale, New York: Erlbaum
- [Cors01] Corsten, Hans: Dienstleistungsmanagement. 4. Aufl., Oldenbourg Wissenschaftsverlag, Oldenbourg 2001.
- [CrTa94] Cronin, Joseph; Taylor, Steven: SERVPERF versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality, In: Journal of Marketing 58 (1994) 1, S. 125-131.
- [Dete04] Detecon&Diebold Consultants: IT Service Management – Trends und Perspektiven in der IT Infrastructure Library (ITIL) in Deutschland. http://www.detecon.com/de/publikationen/studienbuecher_detail.php?pub_id=101&sid=d491ca6a4ab648e4dc04a1a5907efbe7, 2004-04-15, Abruf am 2005-02-01
- [Grem04] Grembergen Van, Wim: Strategies for Information Technology Governance. Idea Group Publishing, Hershey 2004.
- [Hall02] Haller, Sabine: Dienstleistungsmanagement. 2. Aufl., Gabler Verlag, Wiesbaden 2002.
- [Hens90] Hentschel, Bert: Die Messung wahrgenommener Dienstleistungsqualität mit SERVQUAL, In: Marketing ZFP 4 (1990), S. 230-240
- [Hewl03] Hewlet-Packard: The HP IT Service Management (ITSM) Reference Model. ftp://ftp.hp.com/pub/services/itsm/info/itsm_rmwp.pdf, 2003, Abruf am 2006-07-18
- [HoZB04a] Hochstein, Axel; Zarnekow, Rüdiger; Brenner, Walter: ITIL als Common-Practice-Referenzmodell für das IT-Service-Management. In: WIRTSCHAFTSINFORMATIK 46 (2004) 5, S. 382–389.
- [HoZB04b] Hochstein, Axel; Zarnekow, Rüdiger; Brenner, Walter: Evaluation of Service-Oriented IT-Management in Practice. In: Proc. Of the 2005 IEEE International Conference on Service System and Service Management, China 2005.

- [Offi03a] Office of Government Commerce: Service Delivery. CD-ROM Version, The Information Management Company, 2003.
- [Offi03b] Office of Government Commerce: Service Support. CD-ROM Version, The Information Management Company, 2003.
- [PaZB88] Parasuraman A., Zeithaml, Valerie; Berry, Leonard: SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perception of Service Quality, In: Journal of Retailing 64 (1988) 1, S.12-40.
- [PaZB85] Parasuraman, A.; Zeithaml, Valerie; Berry, Leonard: A Conceptual Model of Service Quality and Its Implications for Future Research, In: Journal of Marketing 49 (1985) 4, S. 41-50.
- [PuQA03] Pultorak, Dave; Quagliariello, Pete; Akker, Rolf: Das MOF- Taschenbuch. <http://download.microsoft.com/download/9/c/7/9c7605ee-21ab-4819-965a-21e5bc983087/MOFPocketguideGercleansample.pdf>, 2003, Abruf am 2006-07-18
- [RFHN04] Rasch, B., Frieze, M., Hofmann, W., Naumann E.: Quantitative Methoden. Band 1. Springer, Berlin 2005.
- [Sall04] Sallé, Mathias: IT Service Management and IT Governance: Review, Comparative Analysis and their Impact on Utility Computing. <http://www.hpl.hp.com/techreports/2004/HPL-2004-98.pdf>, 2004-06-02, Abruf am 2005-12-04.
- [Schm04] Schmidt, Rainer: IT-Service Management – Aktueller Stand und Perspektiven für die Zukunft. Hochschule Aalen, http://www.itsmf.de/upload/events/Auswertung_ITIL-Studie.pdf, Abruf am 2005-12-02.
- [ViGü05] Victor, Frank; Günther, Holger: Optimiertes IT-Management mit ITIL. 2. Aufl., Frier. Vieweg & Sohn Verlag, Wiesbaden 2005.
- [ZaHB05] Zarnekow, Rüdiger; Hochstein, Axel; Brenner, Walter: Serviceorientiertes IT-Management. 1. Aufl., Springer, Berlin 2005.
- [ZeBP88] Zeithaml, V. A.; Berry, L. L.; Parasuraman, A.: Communication and control processes in the delivery of service quality. In: Journal of Marketing 52 (1988) 2, S. 35–48.

The Interplay of Outsourcing Risks and Benefits

A Study of Business Process Outsourcing in the German Banking Industry

Kim Wüllenweber

E-Finance Lab

Johann Wolfgang Goethe University

60054 Frankfurt am Main

wuellenweber@wiwi.uni-frankfurt.de

Abstract

What is the influence of outsourcing risks on outsourcing benefits? Although outsourcing literature reveals findings on outsourcing risks and outsourcing benefits, their interplay has hardly been analyzed. Using data from 218 German banking managers within the context of Business Process Outsourcing (BPO), it turns out that financial and performance related risks and benefits heavily interplay. In addition, strategic considerations such as the concentration on core business and the ability to react to market changes have substantial impact on the perception of financial and performance risks and benefits. The findings of this paper can be used to analyze effective risk mitigation instruments and to design risk measurement models incorporating risk and benefit correlations.

1 Introduction

The examination of outsourcing — the purchase of a good or service that was previously provided internally [LaHi93] — has been a domain of IS research for several years now. When considering outsourcing, most of the academic discussions have addressed the questions of “why”, “what”, “which” and “how” to outsource as well as “the outcome of outsourcing” [DGHJ04]. This research provides findings to connect research on “how to outsource” and “the outcome of outsourcing” by analyzing the interplay of outsourcing risks and outsourcing benefits. Neither outsourcing risk [AuPR98] nor outsourcing benefit [LeMK04] can be described by a single measure. They are multi-dimensional constructs which heavily depend on

each other [AuPR98]. To assure a successful outsourcing arrangement, their interplay has to be analyzed to implement effective risk mitigation instruments and to design sophisticated risk measurement models using both risk and benefit correlations. Based on the outsourcer's perspective, we therefore aim at answering the following research questions:

- What is the influence of outsourcing risks on outsourcing benefits?

The research question addresses the concern that all outsourcing risks are financially driven and do only negatively impact the achievement of cost savings on the benefit's side [LaWi98]. But if also quality and strategic related risks affect cost savings and also negatively influence the strategic flexibility, outsourcers should account for these interplays in their outsourcing decision and also in the risk management activities during the delivery phase.

To approach this research question, a research model incorporating both risks and benefits has been developed and tested subsequently in an empirical study conducted on BPO in the German banking industry. The structure of the paper is as follows: In section 2, related research on outsourcing risk and benefit are presented to provide a theoretical base for the research model. Subsequently, in section 3 the research model comprising outsourcing risk and benefit measures is introduced and information provided on data collection and sample characteristics. The empirical results as well as their limitations are outlined in section 4. This paper closes with a summary of the results and provides some implications for further research (section 5).

2 Related Research

In this section, the building blocks of the research model are discussed. As primary research objective the concept of outsourcing benefits are first introduced. As these benefits can be periled by several undesirable outcomes, the concept of outsourcing risk is subsequently discussed to provide a theoretical base for the relation of risks and benefits. As this paper addresses risks and benefits in the context of BPO, current research findings on risks and benefits of BPO are outlined.

2.1 Outsourcing benefits

Multiple researches have been conducted to analyze outsourcing benefits [DGHJ04]. There are many reasons why companies choose to outsource. Academic effort has concentrated on factors

that influence the sourcing decision. Theories like Transaction Cost Economics [Coas37] and the Resource-Based View (RBV) [Barn86] explain the outsourcing decision from a theoretical point of view and have been tested thoroughly in empirical studies. Fundamentally, three drivers for outsourcing have been identified: economic, strategic and technological reasons [LoVe92]. The most important *economic* driver is anticipated cost reduction, as several researchers report. Companies are eager to reduce their overall expenditures (personnel, hard- and software) and discuss any sourcing scenario (incl. off-shoring arrangements) to find out which is the most cost effective one [AnSt98; LaWi98]. On a *strategic* level the demand for more flexible processes and IT systems which support the business in times of rapid change and global reach is often quoted [QuHi94]. This desire often results in the concept of focussing on core competencies through divestment of non-core areas [Quin99]. In times of tight budgets, companies need to allocate capital in the most efficient way and source out those activities which not provide substantial value to their core business. *Technologically*, companies expect to improve their operations through the access to new technology and skilled people by sourcing to a service provider specialized in the respective area [LaWi98]. Thus, quality improvement is often a vital reason to source out.

As this research aims at analyzing the interplay of risk and benefits, outsourcing benefits are viewed as potential benefits.

2.2 Outsourcing risks

Several researchers have addressed the importance of outsourcing risk research [DGHJ04; AuBR05]. However, a consistent definition of outsourcing risk has not emerged. In rational decision theory, the concept of risk reflects the variation in the distribution of possible outcomes, their likelihood and their subjective values [Knig21]. The theory suggests that decision-makers deal with decisions under uncertainty in a rational way, i.e., by computing different alternatives and selecting the option that best suits their personal risk-return profile, which is generally risk-averse [Yate92]. However, empirical studies indicate that this theoretical view is not consistent with how managers deal with risky choices in reality: several studies have shown that managers follow a less precise calculus, not using accurate probability calculations [MaSh98]. Instead managers follow a magnitude of undesired outcome concept neglecting the outcome probability. Therefore, Perceived Risk Theory (PRT) [Bauer67] will be used to analyze manager's risk perception. PRT analyzes the risk a person subjectively associates with the consequences of a decision. By doing this, it is possible to analyze individual risks using

only one measure instead of two (loss severity and loss probability) - thus overcoming the problem of how to combine these parameters.

Cunningham [Cunn67] distinguishes perceived risk into several dimensions called risk facets that are described in the following table and which are based on [FePa03].

No.	Risk facet	Description
1	Performance risk	The possibility of not performing as expected and therefore failing to deliver the desired benefits.
2	Financial risk	The risk to pay more money than initially anticipated.
3	Safety risk	The risk to damage the consumer's health or threaten his life due to the acquisition of the product.
4	Security (privacy) risk	The risk of misuse of private information, e.g. transaction data.
5	Opportunity / time risk	Loss of time when making a bad purchasing decision by wasting time researching and making the purchase decision.
6	Psychological risk	The risk that the selection or performance of the producer will have a negative effect on the consumer's peace of mind or self-perception.
7	Social risk	The potential loss of status in one's social group as a result of adopting a product or service.

Table 1: Perceived Risk Facets

As this research concentrates on manager's perception towards outsourcing in an organisational context, risk facets no. 4 – 7 are excluded. In addition, strategic risk replaces safety risk [Cunn67] as in this research context there is no threat to the life and health of the manager involved, but rather its equivalent for the organizational unit the manager is responsible for.

2.3 Business Process Outsourcing

BPO seems to be one of the largest areas of growth in the outsourcing market [Gart04]. Within this paper, BPO is defined as the delegation of one or more entire business processes to third party providers, including the software and hardware that support those processes [HaMe00]. A business process is defined as a “set of logically related tasks performed to achieve a defined business outcome” [Dave05]. Thus, BPO is the combination of application development/maintenance outsourcing, IT infrastructure outsourcing and the outsourcing of business activities which are not IT supported like business process re-design.

While practitioners and researchers alike agree on the importance of understanding the benefits and risks of BPO, this issue has so far only been addressed by a few authors [GeFr05], and there is a serious lack of theoretical and empirical knowledge on the impact and antecedents of BPO benefits and risk. One elementary question thus is whether the benefits and risks of ITO and BPO are basically the same, or if they differ in structure and magnitude and thus each justify a research domain on their own. The only available research on this issue is a paper by Gewalt and Franke [GeFr05] showing that out of 15 initially investigated risks, 11 are higher in BPO

than in ITO, 2 are regarded as identical and 2 are considered lower in BPO than in ITO. One additional risk previously ignored, i.e., "misuse of trust," is identified as specific to BPO. This serves a methodological base of this paper as some findings from ITO can be transferred to the context of BPO. However, it hardly provides insights on the relation of BPO benefits and risks.

3 Research Model and Methodology

Within this section, the hypotheses of the research model are introduced. In addition, some facts and insights on the methodology used are given.

3.1 Hypotheses and Research Model

Taking the considerations of section 2, the risk measures analyzed here are performance, financial and strategic risks. The achievement of cost advantages, quality improvements and the focus on core competencies will be used as benefit measures (see section 2.2). The proposition is that these measures interplay, the following provides hypotheses on their relations.

Performance risk becomes evident when the outsourcer does not receive the service as expected. This risk is closely related to service debasement as described by [AuPR98]. Service debasement occurs, if the service provider is either not capable or not willing to execute the process as expected by the outsourcer. Not achieving the desired quality standard implies that possible quality improvement will hardly be accomplished. As already outlined, quality improvement has been recognized by several researchers as an important outsourcing reason [DGHJ04]. For outsourcing business processes banks expect benefits of better service quality [ECB04]. This might be achieved by faster execution and/or lower error/failure rates [BCLW04]. Moreover, banks expect the service provider to continuously improve the process to achieve further efficiency [ReBa01]. However, if the service provider does not even provide the minimum performance level, quality improvements will hardly be achieved.

Hypothesis 1: Performance risk negatively influences the achievement of quality improvements.

Performance risk might also have financial effects. As outlined by Lacity and Willcocks [LaWi98] quality related issues often have cost related impacts. If the desired quality level is not achieved, the outsourcer can either increase monitoring activities and call for penalties (if

possible) or allocate internal resources to compensate for quality losses [WiLa99]. As a result, there are unexpected management costs or costly contractual amendments [AuPR98]. In other words, the outsourcer has to pay more than he expected which is used here financial risk (see section 2.1).

Hypothesis 2: Performance risk is positively related to financial risk.

Outsourcing business processes inhibits the potential to exploit vendor's superior knowledge and capabilities with the process [WHFL04]. However, the prerequisite to gain these advantages is a sustained and reliable process quality. In addition, the outsourcer will be limited in its strategic flexibility. New products or services of the outsourcer which affect the way the outsourced process is executed can hardly be introduced, since even 'the old' products are incorrectly processed. Thus, the outsourcer becomes dependent on the service provider and burdens a loss of innovation capabilities [WHFL04]. This negatively influences outsourcer's strategic flexibility, i.e. increases strategic risk.

Hypothesis 3: Performance risk is positively related to strategic risk.

The presence of financial risk might be substantially affected by strategic risk [LaWi98]. Although outsourcing may hinder the bank to quickly react to market changes or implement strategic decisions, the bank might be able to simply allocate additional internal or external staff to overcome these deficiencies. In both cases, the bank has to pay more for the process than expected resulting in rising financial risk.

Hypothesis 4: Strategic risk is positively related to financial risk.

If there are severe strategic risks, the outsourcer has to pay more attention to the outsourced process. By increasing internal process know-how and improving the relationship with the vendor the outsourcer might be able to reduce the risk of lock-in and incentives the vendor to exploit his superior capabilities. By doing this, the outsourcer will have a hard time to focus on his core business and enhance his core competencies [LaWi98; DGHJ04]. Managers and operative staff will not be free to create value for the firm [Earl96]. Instead they will be busy to compensate for the loss of strategic flexibility.

Hypothesis 5: Strategic risk negatively influences the focus on core competencies.

If the outsourcer heavily depends on the service provider and burdens therefore a loss of innovative capacity, he hardly contributes to the achievement of quality improvements [Earl96; WHFL04]. Achieving quality improvements means to analyze process inefficiencies and to identify operational ways to accelerate process execution and/or to lower error rates. This particularly becomes evident, if the outsourcer has not retained sufficient process experts which could help the service provider to improve the process.

Hypothesis 6: Strategic risk negatively influences the achievement of quality improvements.

The achievement of quality improvements implicates that the outsourcer does not have to pay special attention to quality issues. Thus, outsourcer's management and operative staff are free to improve core business [AuBR05].

Hypothesis 7: The achievement of quality improvements is positively associated with the focus on core competencies.

As operational excellence occurs, the service provider does not have to allocate resources to these processes. More transactions can be executed; error handling activities can be reduced. As a result, process costs of the service provider decrease which can be – at least partially – passed on to the client (outsourcer) [LaHi93; LaWi98].

Hypothesis 8: The achievement of quality improvements is positively associated with the achievement of cost savings.

Finally and most prevailingly, the achievement of cost savings is negatively affected by the occurrence of financial risk [AuPR98; HiLa00]. Additional, unexpected costs might consume anticipated cost savings. Even if fix prices are negotiated which represent a drastic reduction of process costs, extra management costs or hidden costs by the service provider (e.g. pricing new services exceptionally) might compensate these benefits.

Hypothesis 9: Financial risk negatively influences the achievement of cost savings.

3.2 Methodology and sample characteristics

The research model depicted in the previous section has been operationalized and transferred into a structural equation model (SEM). The SEM will be analyzed using Partial Least Squares (PLS) method [Chin98]. In contrast to covariance-based approaches (e.g. LISREL or AMOS), PLS has minimal requirements on measurement scales, sample size, and residual distribution

[Chin98]. It is particularly suitable if a more explorative analysis is preferred. As there is no strong theoretical foundation on the actual impact of outsourcing risk measures on outsourcing benefit measures, an explorative approach seems to be appropriate. Furthermore, one construct/latent variable (LV) has been operationalized in formative mode and PLS is the only algorithm that allows to both applying formative and reflective indicators (for the distinction of formative and reflective indicators cp. e.g. [JaMP87]).

Each LV in the research model is represented by a set of indicators, which were measured on a fully anchored 7-point Likert scale, ranging from “strongly agree” to “strongly disagree” for the LV quality improvements (QI), cost savings (CS) and focus on core competencies (CC). Performance risk (PR), financial risk (FR) and strategic risk (SR) were measured using scales from “very high” (risks) to “very low” (risks). Whenever possible, existing measures from prior empirical studies were adopted. The questionnaire was pre-tested independently with managers from different banks which were not included in the final sample. Based on the insights acquired in these pre-tests, the questionnaire was modified and finalized.

For this research the 200 largest banks in Germany were chosen, based on their total assets as reported in the balance sheet of the year 2003. The cumulated balance sheets of the 200 largest banks account for ca. 90% of the cumulated balance sheet of the German banking market.

To assess the interplay of risks and benefits, the managers in charge of four different banking processes were selected as units of analysis, which are generally not regarded as areas of core competence for banks: back office/settlement processes for transactions in securities, consumer credits, domestic payments and foreign exchange/money market. All 200 top banks were contacted by phone to personally identify the managers responsible for the business processes mentioned above. To increase the response rate and ensure that only the managers fill out the questionnaire themselves, the managers were identified and contacted personally. Some banks did not have all four business processes, therefore only 593 questionnaires (instead of the maximum number of 800) were sent out.

In total, 218 analyzable questionnaires from 126 banks were returned. This equals a response rate of 36.8% among managers and 63% of the banks. Taking the bank responses, the cumulated assets of the responses accounted for roughly 80% of the total cumulated German banking balance sheet. The response rate amongst large banks (assets > EUR 20bn) was exceptionally high (79.6%). The distribution of the banking groups (private banks, savings banks, cooperative banks, other banks) is representative.

4 Empirical Results

This section comprises the results from the PLS analysis, the discussion of key findings as well as limitation considerations. For analysis of the data, the software pls-graph, version 3.0, developed by Wynne Chin, has been used.

4.1 Model validation

4.1.1 Measurement model specification

All manifest variables used in the model have been derived from other studies. The manifest variables for measuring the latent variables (LV) are given in the table below. The LV CC has been operationalized in formative mode; all others are in reflective mode.

Indicator	LV	Mean	SD	Related Research
a16: The service provider will not provide the promised service.	PR	3.61	1.42	[Earl96; AuBR98]
a17: The service provider will not perform the process to the desired quality (speed and accuracy) and quantity.		4.32	1.48	
a18: The service provider will agree on more beforehand than he will actually deliver during the outsourcing venture.		4.59	1.56	
a19: The originally calculated business case will not include all the actual costs?	FR	4.76	1.42	[Earl96; LaWi98; HiLa00]
a20: Unanticipated costs will emerge that reduce the calculated cost savings?		4.74	1.42	
a21: The anticipated cost savings will not be achieved?		4.62	1.42	
a28: The bank loses its ability to react flexibly to changes in the market.	SR	4.34	1.62	[Earl96; WHFL04]
a29: The bank loses its ability to improve its position in the market by means of internal optimization procedures?		4.15	1.62	
a30: The bank loses know-how that is required to remain competitive in future markets?		4.68	1.58	
a34: Our bank can carry out the process internally at lower cost than an external service provider.	CS	3.63	1.60	[AnSt98; LaWi98; HiLa00]
a35: Our internal production costs are higher than the price		4.33	1.53	
a36: Outsourcing lowers the costs that arise from executing a business process.		4.40	1.44	
a37: Overall, I believe that outsourcing is an appropriate measure to lower the costs within this business process.		4.44	1.51	
a44: An external service provider has the potential to perform the process at a higher quality than our bank.	QI	3.74	1.48	[LaWi98]
a45: An external service provider is able to perform the process faster and/or at a higher accuracy than our bank.		3.68	1.49	
a46: By outsourcing the quality of executing the process will be improved.		3.53	1.48	
a93: By outsourcing the bank's processes can be better directed towards its core business.	CC	4.50	1.46	[GrCT96]
a94: By outsourcing the bank's IT systems can be better directed towards its core business.		4.13	1.49	

Table 2: Indicators

4.1.2 Formative measurement model

In the research model, “focus on core competencies” has been operationalized in formative mode since the indicators meet the criteria put forward in [JaMP87] for formative measurement models. According to the findings of both [DiWi01] and [Chin98], there are five critical issues determining the quality of the formative measurement model: (1) content specification, (2)

indicator specification, (3) indicator reliability, (4) indicator collinearity and (5) external validity.

Content specification consists of defining the scope of the latent constructs to be measured. The construct “focus on core competencies” was precisely defined and their domain intensively discussed, ensuring the proper specification of the applicable content of the construct.

Indicator specification comprises the identification and definition of indicators which constitute the construct. The indicators used in this model were identified by intensive literature review and have been validated through several pre-tests with senior bank managers who were knowledgeable about the topic of this research.

Indicator reliability analyzes the importance of each individual indicator that forms the construct. Two quantitative arguments have to be accounted for: (1) the sign of the indicator needs to be correct as hypothesized and (2) the weighting of the indicator should be at least 0.2 as proposed by [Chin98]. The model tested shows correct signs, sufficient weights (a93: 0.757; a94: 0.323) and are significant at the 0.05 level.

Because formative measurement models are based on linear equation systems, substantial *indicator collinearity* would affect the stability of indicator coefficients. Neither the analysis of correlations of indicators nor the calculation of variance inflation factors (all indicators fall far below the threshold of 10 [KIKM88]) necessitated the rejection of any indicators used. Therefore, all indicators were retained as no redundancy was identified.

External validity aims at ensuring that all indicators which form a construct are actually included in the model. Following [DiWi01], external validity can be analyzed by creating a phantom construct which is measured using reflective indicators. If the formatively measured construct strongly and significantly correlates with the reflective measured construct, external validity is given. The correlations of constructs within the tested model were all strong and significant at the 0.01 level. Thus, it is shown that the formative indicators used in this study actually form their respective constructs.

4.1.3 *Reflective measurement model*

The quality of the reflective measurement model is determined by (1) convergent validity, (2) construct reliability and (3) discriminant validity [BaYi88].

Convergent validity is analyzed by indicator reliability and construct reliability. In the model tested, all loadings are significant at the 0.001 level and above the recommended 0.7 parameter

value (significance tests were conducted using the bootstrap routine with 500 samples [Chin98]; results see appendix).

Construct reliability was tested using two indices: (1) the composite reliability (CR) and (2) the average variance extracted (AVE). Estimated indices were above the recommended thresholds [BaYi88] of 0.6 for CR and 0.5 for AVE (see appendix).

Discriminant validity of the construct items can be analyzed by looking at the cross-loadings. As for the reflective indicators, the loadings of each indicator are higher for their respective constructs than for any other construct. Furthermore, the square root of the AVE for each construct is higher than correlations between constructs. Therefore, the indicators of different constructs are not related to each other and discriminant validity of the latent variables is high.

As outlined in section 2, risk is closely related to benefit as it is variation in the distribution of possible outcomes. In particular, H1 and H9 might be tautological. Thus, indicator collinearity across constructs has to be analyzed. However, neither the analysis of correlations of indicators nor the calculation of variance inflation factors (all indicators fall far below the threshold of 10 [KlKM88]) supports this concern.

4.1.4 Structural model

After reviewing the measurement model, the explanatory power of the structural model is evaluated. The explanatory power is examined by looking at the squared multiple correlations (R^2) of the dependent variables. Encouragingly 49.2% ($R^2=0.492$) of the variation in financial risk is explained by performance and strategic risk. The R^2 value for cost advantages ($R^2=0.341$), quality improvements ($R^2=0.305$), focus on core competencies ($R^2=0.335$) and strategic risk ($R^2=0.289$) can be interpreted as moderate explanatory power according to Chin [Chin98](1998).

Predictive power is tested by examining the magnitude of the standardized parameter estimates between constructs together with the corresponding t-values that indicate the level of significance. All path coefficients exceed the recommended 0.2 level. The exception is the impact of strategic risk on focus on core competencies (H5). Additionally, as bootstrapping reveals, all path coefficients except H4 are highly significant (at the 0.001 level). Analysis of the overall effect size (f^2) reveals that all constructs have moderate effect except H5. However, small f^2 scores do not necessarily imply an unimportant effect [Coh88]. Thus, all hypotheses have been proven to be correct except H5. The following figure shows the structural model findings.

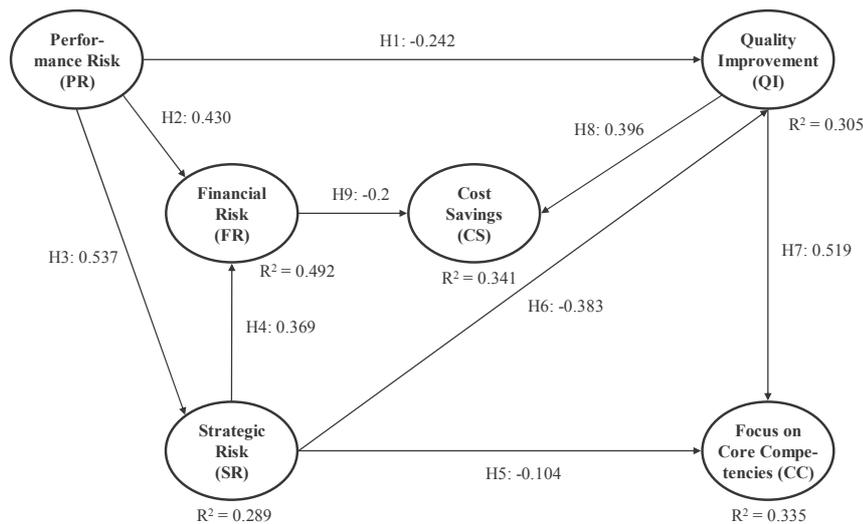


Figure 1: Research Model Results

4.2 Key findings

This study is the first quantitative analysis on the interplay of BPO risks and benefits. The results are very encouraging, and the high response rate indicates the importance of the topic and the interest of practitioners in the results of this research.

The data analysis reveals that almost all hypotheses hold as expected, showing significant loadings. The R^2 values of the dependent variables are satisfactorily high, indicating the high degree of explanatory power of its predecessors.

Quality improvements are negatively affected by performance and strategic risks (significant path coefficients (H1 and H6) as well as satisfactory high R^2). The influence of performance risk (H1) indicates that outsourcers are often confronted with basic quality issue making the achievement of high quality standards unreachable. This might be an indication that the BPO market is still immature and adopters are often confronted with more basic problems. This has been recognized for IT outsourcing a decade ago [LaHi93] and might now be applicable for the relatively new phenomenon BPO. This includes also the presumption that the service provider is not willing to increase quality due to opportunistic behaviour [AuPR98].

The influence of strategic risk (H6) suggests that there is a principal negative relation between outsourcing efficiency and flexibility gains. This contradiction has not been analyzed in outsourcing context before.

Financial risk can thoroughly be explained by performance and strategic risk (H2 and H4) as their explanatory power is exceptionally high (0.492) and path coefficients are high and

significant. The impact of performance risk on financial risk indicates that the outsourcer often has to either compensate vendor's underperformance internally entailing additional costs or is confronted with extra charges by the service provider although he expected the services to be included (see similar findings by [HiLa00]).

The influence of strategic risk shows that strategic considerations can significantly affect the financials of an outsourcing deal. If new products are launched or the existing portfolio is re-arranged, the service provider has to change his operations to fully accomplish outsourcer's requirements. As this is "extra work", the service provider will claim for "extra charge". Previous findings have revealed that "extra charges" occurs due to opportunistic behaviour of the service provider or inexperience of the service provider and/or the outsourcer [Earl96; AuPR98]. The relation between strategic and financial issues has hardly been addressed in outsourcing research before (see [LaWi98] on a similar topic).

Strategic risks can be partially explained by performance risks (H3). Process errors and/or long execution times imply operational difficulties impeding the bank to quickly react to market changes or to implement new strategies. As decreasing flexibility has even been recognized as an important reason to backsource [HiLa00], the role of performance problems become particularly interesting. In addition, the outsourced operations analyzed here are near bank's core business. Thus, quality problems have a direct and substantial influence on bank's market success.

The focus on core competencies depends substantially on the achievement of quality improvements (H7), but is only slightly affected by strategic risk (H5). This indicates that the bank will only be able to focus on its core business, if objectives – such as quality improvements – are achieved. Thus, allocation of staff to core business activities depends on the degree of operational excellence. This relation has hardly been recognized in outsourcing context before.

On the other side, the presence of strategic risk does not hinder the bank to focus on his core business. Even if the bank is not able to react quickly to market changes, the bank is still able to intensify activities to improve its core business. One reason might be that the service provider has to deal with consequences of strategic risk whereas the focus of core competencies can only be achieved internally.

The achievement of cost savings is significantly influenced by the achievement of quality improvements and the presence of financial risk (H8 and H9). The impact of quality improvements extends findings of Lacity and Willcocks [LaWi98] to BPO context. In addition, it shows that managers are confident that quality improvements achieved by the service provider will be passed over to the bank and imply further cost reductions.

Contrastingly, financial risks negatively affect the achievement of cost savings. Although this contributes to empirical studies reporting the failure of outsourcing ventures due to cost escalations, it does not imply that cost savings are compensated by losses due to the occurrence of financial risk.

4.3 Limitations

This study has focused on the German banking industry. Therefore, the results are only indicative for other industries and countries. As there are different national and industrial outsourcing regulations, the importance and awareness of risks differ. When switching to other countries, cultural differences entail additional risks and a change of risk severity [Dibb03].

5 Conclusion and further research

This research is the first quantitative study analyzing the interplay of outsourcing risks and benefits. It has been shown that not only performance and financial related risks and benefits strongly interrelate. But also the vital impact of strategic risks and benefits (focusing on core competencies) on financial and performance risks and benefits bring new light in outsourcing success research: researchers can use these findings to better analyze the consequences of risk mitigation instruments, such as the outsourcing contract. In addition, they are able to develop risk measurement models incorporating risk and benefit correlations. Further research could illuminate how the interplay of risks and benefits change over time – when risks turn to losses and/or benefits are achieved.

Practitioners can use the findings to better understand the existence and development of overall risk and benefits formation. This enables them to construct effective risk mitigation instruments and to design precise risk measurement models in the design phase. Firms in the decision phase benefit from the results by incorporating the perspective of quality and strategic related risks

and benefits and their interplay with financially driven objectives. As BPO is still a relative new phenomenon, these insights may be of special interest for the market.

References

- [AnSt98] Ang, S. and D. Straub (1998). "Production and Transaction Economies and IS Outsourcing: A Study of the U.S. Banking Industry." *MIS Quarterly* 4: 535-552.
- [AuBR05] Aubert, B. A., B. Bahli and S. Rivard (2005). "A Framework for Information Technology Outsourcing Risk Management." *The DATA BASE for Advances in Information Systems* 36(4).
- [AuPR98] Aubert, B. A., M. Patry and S. Rivard (1998). *Assessing the Risk of IT Outsourcing*. 31st Hawaii International Conference on System Sciences.
- [BaYi88] Bagozzi, R. P. and Y. Yi (1988). "On the Evaluation of Structural Equation Models." *Journal of the Academy of Marketing Science* 16: 74-94.
- [Barn86] Barney, J. B. (1986). "Organizational Culture: Can it Be a Source of Sustained Competitive Advantage?" *Academy of Management Review* 11: 656-665.
- [Baue67] Bauer, R. (1967). *Consumer Behavior as Risk Taking. Risk Taking and Information Handling in Consumer Behavior*. D. F. Cox. Cambridge, MA, USA, Harvard University Press: 21-33.
- [BCLW04] Bucu, M. J., R. N. Chang, L. Z. Luan, C. Ward, J. L. Wolf and P. S. Yu (2004). "Utility computing SLA management based upon business objectives." *IBM Systems Journal* 43(1): 159-178.
- [Chin98] Chin, W. W. (1998). *The Partial Least Squares Approach to Structural Equation Modeling. Modern methods for business research*. G. A. Marcoulides. London, Lawrence Erlbaum Associates: 295-336.
- [Coas37] Coase, R. H. (1937). "The Nature of the Firm." *Economica (New Series)* 4(16): 386-405.

- [Cohe88] Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New Jersey, Lawrence Erlbaum.
- [Cunn67] Cunningham, S. M. (1967). *The Major Dimensions of Perceived Risk. Risk Taking and Information Handling in Consumer Behaviour*. D. F. Cox. Boston, MA, USA, Harvard University Press: 82-108.
- [Dave05] Davenport, T. (2005). "The coming commoditization of processes." *Harvard Business Review* June: 100-108.
- [DiWi01] Diamantopoulos, A. and H. M. Winklhofer (2001). "Index Construction with Formative Indicators: An Alternative to Scale Development." *Journal of Marketing Research* 38(May): 269–77.
- [Dibb03] Dibbern, J. (2003). *The Sourcing of Application Software Development and Maintenance*. Berlin, Springer.
- [DGHJ04] Dibbern, J., T. Goles, R. Hirschheim and B. Jayatilaka (2004). "Information Systems Outsourcing: A Survey and Analysis of the Literature." *The DATA BASE for Advances in Information Systems* 35(4): 6-102.
- [Earl96] Earl, M. J. (1996). "The Risks of Outsourcing IT." *Sloan Management Review* Spring: 26-32.
- [ECB04] ECB (2004). *Report on EU banking structure*. Frankfurt, Germany.
- [FePa03] Featherman, M. S. and P. A. Pavlou (2003). "Predicting e-Services Adoption: A Perceived Risk Facets Perspective." *International Journal of Human-Computer Studies* 59: 451-474.
- [Gart04] Gartner (2004). *Outsourcing Market View, What the Future Holds*. Gartner Dataquest.
- [GeFr05] Gewalt, H. and J. Franke (2005). *A Comparison of the Risks in Information Technology Outsourcing and Business Process Outsourcing*. Proceedings of the Eleventh Americas Conference on Information Systems, Omaha, NE, USA.

- [GrCT9] Grover, V., M. J. Cheon and J. T. C. Teng (1996). "The Effect of Service Quality and Partnership on the Outsourcing of Information Systems Functions." *Journal of Management Information Systems* 12(4): 89-116.
- [HaMe00] Halvey, J. K. and B. M. Melby (2000). *Business Process Outsourcing - Process, Strategies and Contracts*. New York, John Wiley & Sons.
- [HiLa00] Hirschheim, R. A. and M. C. Lacity (2000). "The Myths and Realities of Information Technology Insourcing." *Communications of the ACM* 43(2): 99 - 107.
- [JaMP03] Jarvis, C. B., S. B. MacKenzie and P. M. Podsakoff (2003). "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research." *Journal of Consumer Research* 30(2): 199-218.
- [KIKM88] Kleinbaum, D. G., L. Kupper and K. E. Muller (1988). *Applied regression analysis and other multivariable methods*. Boston, PWS-Kent.
- [Knig21] Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Chicago.
- [LaHi93] Lacity, M. C. and R. Hirschheim (1993). *Information Systems Outsourcing: Myths, Metaphors, and Reality*. New York, John Wiley and Sons.
- [LaWi98] Lacity, M. C. and L. P. Willcocks (1998). "An Empirical Investigation of Information Technology Sourcing Practices: Lessons from Experience." *MIS Quarterly* September: 363-408.
- [LeMK04] Lee, J. N., S. Miranda and Y.-M. Kim (2004). "IT Outsourcing Strategies: Universalistic, Contingency, and Configurational Explanations of Success." *Information Systems Research* 15(2): 110-131.
- [LoVe93] Loh, L. and N. Venkatraman (1992). "Determinants of Information Technology Outsourcing: A Cross-Sectional Analysis." *Journal of Management Information Systems* 9(1): 7 - 24.
- [MaSh87] March, J. G. and Z. Shapira (1987). "Managerial Perspectives on Risk and Risk Taking." *Management Science* 33(11): 1404-1418.

- [Quin99] Quinn, J. B. (1999). "Strategic Outsourcing: Leveraging Knowledge Capabilities." *Sloan Management Review* 40(4): 9-22.
- [QuHi94] Quinn, J. B. and F. G. Hilmer (1994). "Strategic Outsourcing." *Sloan Management Review* 35(4): 43-55.
- [ReBa01] Rebouillon, J. and S. Bauer (2001). *Optimierung der Wertschöpfungskette durch Outsourcing. Management der Wertschöpfungsketten in Banken.* L. P. Marighetti, R. Jasny, A. Herrmann and F. Huber. Wiesbaden, Gabler.
- [WHFL04] Willcocks, L., J. Hindle, D. F. Feeny and M. C. Lacity (2004). "IT and Business Process Outsourcing: The Knowledge Potential." *Information Systems Management Summer*: 7-15.
- [WiLa99] Willcocks, L. P. and M. C. Lacity (1999). "IT outsourcing in insurance services: risk creative contracting and business advantage." *Information Systems Journal* 9: 163-180.
- [Yate92] Yates, J. F. (1992). *Risk-taking behavior.* Chichester, UK, Wiley.

Appendix

The following table shows the loadings of reflective indicators. All items are significant at the 0.001 level. Only indicator a34 is reverse coded.

Construct	Item	Loading	CR	AVE
Performance risk	a16	0.879	0.936	0.829
	a17	0.932		
	a18	0.918		
Financial risk	a19	0.942	0.951	0.865
	a20	0.948		
	a21	0.899		
Strategic risk	a28	0.928	0.923	0.801
	a29	0.917		
	a30	0.836		
Cost savings	a34*	-0.841	0.923	0.749
	a35	0.859		
	a36	0.876		
	a37	0.884		
Quality improvements	a44	0.930	0.963	0.897
	a45	0.950		
	a46	0.959		

Table 3: Constructs and Loadings

Einführung in den Track

Product Life Cycle Management in Unternehmenssoftwaresystemen

Prof. Dr. Stefan Kirn

Universität Hohenheim

Dr. C. Pohl

SAP AG

Prof. Dr. Peter Loos

Universität Saarbrücken

Die Gestaltung und das Controlling des Produktlebenszyklus gehört seit langem zu den Kernaufgaben des Produktmanagement. Um ein durchgängiges Management des Produktlebenszyklus zu verwirklichen, ist es notwendig, Produktdaten in jeder Produktlebensphase zu erheben und zu verarbeiten. Neue Technologien wie RFID und Mobile Computing versprechen hier ein großes Potenzial, da mit der Erfassung von Produktdaten unmittelbar am Entstehungsort eine Konvergenz der betrieblichen Informationssysteme ermöglicht wird. Die zunehmende Ausstattung mechanischer Bauteile mit Prozessoren und Embedded Software verändert nicht nur die Nutzungs- und Wartungseigenschaften von Produkten, sondern führen auch dazu, dass "smarte" Produkte drei voneinander weitgehend unabhängigen Lebenszyklen unterworfen sind: den Lebenszyklen des eigentlichen Bauteils als Trägersystem, der Hardware und der Embedded Software. Die daraus entstehenden Herausforderungen an die Unternehmenssoftwaresysteme und Ansätze zu ihrer Lösung sind bis heute weitgehend ungeklärt.

Programmkomitee:

Dr. Konstantina Geramani, TWT Neuhausen

Prof. Dr. Manfred Grauer, Universität Siegen

Prof. Dr. Norbert Gronau, Universität Potsdam

Jun.-Prof. Dr. Axel Hahn, Universität Oldenburg

Prof. Dr. Stefan Kirm, Universität Hohenheim (Sprecher)

Prof. Dr. Peter Loos, Universität Saarbrücken

Dr. C. Pohl, SAP AG

Prof. Dr.-Ing. B. Scholz-Reiter, Universität Bremen

Hybride Wertschöpfung im Maschinen- und Anlagenbau

Prozessorientierte Integration von Produktentwicklung und Service-dokumentation zur Unterstützung des technischen Kundendienstes

Oliver Thomas¹, Philipp Walter¹, Peter Loos¹, Michael Schlicker², Markus Nüttgens³

¹Institut für Wirtschaftsinformatik (IW_i)
im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI),
Universität des Saarlandes
Stuhlsatzenhausweg 3, Geb. D3 2, D-66123 Saarbrücken
{oliver.thomas|philipp.walter|peter.loos}@iwi.dfki.de

²INTERACTIVE Software Solutions GmbH
Saarterrassen, Hochstraße 63, 66115 Saarbrücken
michael.schlicker@interactive-software.de

³Universität Hamburg
Von-Melle-Park 9, 20146 Hamburg
markus.nuettgens@wiso.uni-hamburg.de

Abstract

Dieser Beitrag behandelt die prozessorientierte Integration von Produktentwicklung und Service-dokumentation zur Unterstützung des technischen Kundendienstes (TKD) im Maschinen- und Anlagenbau, exemplarisch in der Branche Sanitär-, Heizungs- und Klimatechnik (SHK). Durch die mit der integrierten Betrachtung verbundene Gestaltung eines hybriden Produkts kann die Effizienz des Vorgehens in der Serviceerbringung mithilfe mobiler Anwendungssysteme erhöht werden. Die Entwicklung und Bereitstellung des hybriden Produkts bedingt dabei eine interdisziplinäre Sichtweise. Im Beitrag detailliert dargestellt werden die Problemstellung, der Lösungsansatz auf Basis hybrider Wertschöpfung, die Struktur des hybriden Produkts, die informationstechnische Konzeption sowie die Umsetzung der Serviceprozessmodellierung. Als Anwendungsfall dient die Fehlerdiagnose an einem Heizgerät zur Erwärmung von Wasser.

1 Einleitung

1.1 Problemstellung

Der Maschinen- und Anlagenbau ist mit ca. 862.000 Beschäftigten die größte Industriebranche Deutschlands [VDMA06]. Dem gestiegenen Wettbewerbsdruck begegnen die Unternehmen vor allem durch Kundenbindung. Ein zentraler Aspekt ist hierbei die Ausweitung und Verbesserung ihres Serviceangebots speziell im technischen Kundendienst (TKD), der Schnittstelle zwischen Herstellung und Nutzung der Produkte [Kroo66; Meff82; Muse88; Teic94; Harm99; Breu01; Harm03]. Hier agieren sowohl werkseigene Serviceorganisationen des Herstellers als auch ausgelagerte klein- und mittelständische Unternehmen (KMU) und Handwerksbetriebe, welche die im Produktlebenszyklus anfallenden Inspektions-, Wartungs- und Instandhaltungsarbeiten ausführen [Will87].

Um die damit verbundenen Aufgaben adäquat erfüllen zu können, muss ein TKD mit dem richtigen „Informations-Mix“ versorgt werden. Ein zentrales Problem ist hierbei die Beantwortung der Frage nach dem Umfang, Zeitpunkt und Verdichtungsgrad der entscheidungsrelevanten Informationen [Herm99; BuSG00]. Aktuelle Ansätze zur Unterstützung des TKD scheitern oftmals an der gestiegenen Komplexität der Maschinen und dem hiermit verbundenen gestiegenen Bedarf zur Repräsentation der Serviceerbringungsprozesse. Die Folge sind fehlerhafte Inbetriebnahme-, Wartungs- und Reparaturarbeiten und daraus resultierend eine Verlängerung von Maschinenausfallzeiten, die letztlich in erhöhten Kosten für die Kunden und Marktverluste beim Hersteller münden.

1.2 Zielsetzung und Lösungsansatz

Dem zuvor in Abschnitt 1.1 beschriebenen Umstand wird in dem Projekt PIPE¹ durch die integrierte Entwicklung von physischem Produkt und servicerelevanten Informationsbausteinen sowie der Zusammenführung dieser beiden Produktionsfaktoren zu effizienten Serviceprozessen, die dem TKD mobil zur Verfügung gestellt werden, entgegengewirkt. Die zentrale These dieses Konzepts ist, dass durch die mit der integrierten Betrachtung verbundene Gestaltung eines neuen hybriden Produkts die Anforderungen des TKD an eine kundengerechte Inbetriebnahme, Instandhaltung, Wartung und Reparatur von Maschinen und Anlagen gewährleistet sowie die Ef-

¹ Das Akronym PIPE steht für „Prozessorientierte Integration von Produktentwicklung und Servicedokumentation zur Unterstützung des technischen Kundendienstes“. Das Verbundprojekt wird vom BMBF im Rahmen des Konzepts „Innovation mit Dienstleistungen“ gefördert (Förderkennzeichen: 01FD0623).

fizienz des TKD erhöht werden können. Innovativ an diesem Ansatz ist, dass durch die frühzeitige Verzahnung von Produktentwicklung, Dokumentation, TKD, Prozessberatung und moderner Informations- und Kommunikationstechnologie ein hybrides Produkt entsteht, welches die Erstellung integrierter prozessorientierter Produkt- und Serviceinformationen beim Hersteller mit vertretbarem Aufwand ermöglicht. Serviceorganisationen, wie ein werkseigener Kundendienst oder der Kundendienst eines klein- und mittelständischen Handwerksbetriebs, können auf diese Informationen mobil zugreifen.

1.3 Konkretisierung der Anwendungsdomäne

Zur Erzielung richtungsweisender Forschungsergebnisse in Bezug auf Problemstellung, Zielsetzung und Lösungsansatz eignet sich der Wirtschaftszweig Sanitär-, Heizungs- und Klimatechnik (SHK) idealtypisch. Zum einen stellen die Hersteller dieser Branche technisch komplexe Produkte her, zum anderen wird der TKD zum größten Teil von den Handwerksbetrieben und Serviceorganisationen des SHK-Handwerks ausgeführt [Mose87; HoSa96] (vgl. Abb. 1).

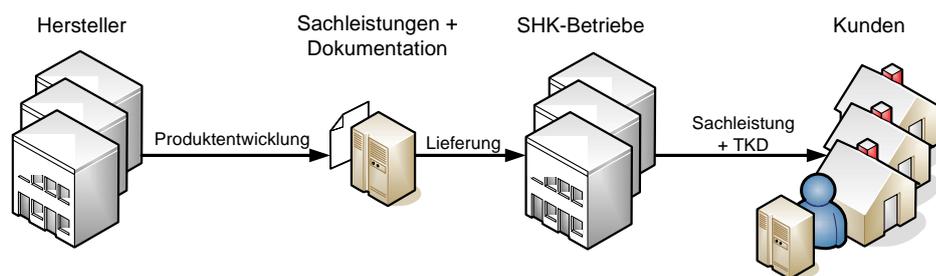


Abb. 1: Status quo der Wertschöpfungskette im SHK-Bereich

Die Instandhaltungsobjekte der SHK-Branche werden in sehr unterschiedlicher Ausprägung gefertigt, sodass Instandhaltungsarbeiten an einfach aufgebauten Produkten, z. B. die Reparatur eines defekten Spülkastens, ebenso anfallen wie die Störungsbehebung innerhalb einer sehr komplexen Wärmeerzeugungs- und Verteilungsanlage [Bill97; ScWa04; West04; BIBB04].

Der Beitrag ist wie folgt organisiert: In Abschnitt 2 werden die zentralen Herausforderungen für den technischen Kundendienst im SHK-Bereich aus den Perspektiven der Hersteller, der Handwerksbetriebe sowie der Kundendiensttechniker erklärt. Anschließend wird in Abschnitt 3 ein Lösungsansatz vorgestellt, der geeignet ist, diesen Herausforderungen zu begegnen. Die Einsatzpotenziale des erarbeiteten Konzepts werden in Abschnitt 4 anhand eines realen Anwendungsfalls der Sanitär-, Heizungs- und Klimatechnik veranschaulicht. Eine Zusammenfassung der Ergebnisse und ein Ausblick in Abschnitt 5 schließen den Beitrag ab.

2 Kundendienstprozesse der Sanitär-, Heizungs- und Klimatechnik

2.1 Herausforderungen aus Sicht der Hersteller

Die Hersteller der SHK-Branche bedienen den Markt mit ihren Produkten überwiegend über die ca. 50.000 SHK-Fachbetriebe und deren ca. 300.000 Mitarbeiter. Der TKD wird dabei sowohl vom Werkskundendienst des Herstellers als auch durch vom Hersteller ausgewählte Servicepartner oder die SHK-Fachbetriebe der Branche ausgeführt [Mose87; Will87; HoSa96]. Die Herausforderung für die Hersteller im Bereich des TKD besteht darin, den Kundendienstorganisationen das Reparatur- und Produktwissen zu vermitteln. Zu diesem Zweck werden TKD-Schulungen angeboten, telefonische Unterstützung in der Reparaturausführung über Call-Center eingerichtet und technische Unterlagen papierbasiert oder elektronisch, z.B. auf CD-ROM, zur Verfügung gestellt. Daraus ergibt sich für die Hersteller ein sehr hoher Aufwand bezüglich der Wissensvermittlung und -bereitstellung. So werden in manchen Betrieben in der technischen Beratung die Beraterplätze ständig aufgestockt, um der steigenden Nachfrage nach Reparaturinformationen begegnen zu können. Trotz dieser hohen Aufwendungen der Hersteller werden aber bei der Arbeitsausführung im TKD immer noch viele Fehler gemacht. Aufgrund fehlerhafter Wartungs- und Reparaturarbeiten entstehen dem Hersteller einerseits Kosten für zusätzliche Leistungen (z.B. Garantie, Gewährleistung, Kulanz), die nicht auf den Kunden umgelegt werden können. Andererseits besteht für den Hersteller die Gefahr, dass bei er dauerhafter Kundenzufriedenheit bedeutsame Marktanteile verliert.

2.2 Herausforderungen aus Sicht der SHK-Betriebe

Auch der SHK-Fachbetrieb muss sich von seinen Wettbewerbern abheben, vorhandene Kunden an sein Unternehmen binden und neue Kunden gewinnen [Breu01; BIBB03]. Dies erfolgt heute stärker als in der Vergangenheit über den TKD. Die Herausforderung für den SHK-Betrieb im Bereich des TKD besteht darin, dass Produkte unterschiedlicher Hersteller zu bearbeiten und aus der Fülle der von den Herstellern angebotenen Informationsquellen die für eine bestimmte Reparatursituation richtigen Informationen herauszufiltern sind [Bunk04] (vgl. Abb. 1). Dabei ist es in der Praxis schwierig, unterschiedliche Gerätekenntnisse der Kundendiensttechniker auszugleichen, die Informationen in eine adäquate Reparaturhandlung umzusetzen und den Wissensverlust im Unternehmen durch das Ausscheiden erfahrener Mitarbeiter auszugleichen. In Analogie zur Argumentation aus der Perspektive der Hersteller ergeben sich auch für die SHK-Betriebe Kosten, die i. d. R. aus zusätzlichen Kundendiensteinsätzen resultieren.

2.3 Herausforderungen aus Sicht der SHK-Kundendiensttechniker

Die Art der Arbeitsausführung hat sich von der funktionsorientierten Arbeitsteilung hin zur prozessorientierten Arbeitsausführung gewandelt. Dabei steht der gesamte Prozess des Kundenauftrags im Mittelpunkt der Betrachtung – dies gilt insbesondere im TKD. Der Kundendiensttechniker erbringt die Leistungen überwiegend im „Alleingang“ vor Ort. Er ist verantwortlich für die korrekte Arbeitsverrichtung, das Identifizieren der benötigten Ersatzteile und die Ersatzteilbeschaffung. Die erfolgreiche Ausführung eines Reparaturauftrags – und damit auch der wirtschaftliche Erfolg des ausführenden SHK-Unternehmens – werden dabei wesentlich von der Effektivität und Effizienz seiner Arbeitsausführung bestimmt. Das Problem für den Kundendiensttechniker liegt vor allem in der hohen Anzahl der zu betreuenden Hersteller und Produkte. Die hieraus resultierende Komplexität der Aufgaben im TKD ist selbst für erfahrene Kundendiensttechniker kaum zu bewältigen. Daher wächst die Bedeutung der Identifizierung und optimalen Gestaltung der Serviceprozesse und der Unterstützung im TKD durch mobile, internetbasierte Informationssysteme, über die ein Kundendiensttechniker zu jeder Zeit und an jedem Ort auf aktuelle Serviceinformationen zugreifen kann [Lehn03; HöSa04; Kirs06].

3 Hybride Wertschöpfung als Innovationsmotor

3.1 Strategischer Lösungsansatz

Leitgedanke des Forschungsprojekts PIPE ist eine Effizienzsteigerung des TKD im Maschinen- und Anlagenbau. Dazu wurde auf Basis der integrierten prozessorientierten Betrachtung von Produktentwicklung und Servicedokumentation eine Methodik zur Entwicklung hybrider Produkte gestaltet und ein solches hybrides Produkt am Beispiel der SHK-Branche prototypisch umgesetzt. Die Forschungsergebnisse sind generell auf die Branche des Maschinen- und Anlagenbaus anwendbar und ermöglichen sowohl die „Hybridisierung“ bestehender als auch zukünftig zu entwickelnder technischer Erzeugnisse.

Bei der materiellen Komponente des hybriden Produkts handelt es sich um ein technisches Erzeugnis des Maschinen- und Anlagenbaus sowie dessen Dokumentation. Diese bestehende oder zukünftig zu entwickelnde technische Anlage wird zu einem hybriden Produkt aufgewertet, indem Dienstleistungen zur Entwicklung, Bereitstellung, Anwendung und Überarbeitung integrierter Serviceprozessbeschreibungen konzipiert werden, die so den kompletten Lebenszyklus

der Serviceprozessdokumentation abdecken. Darüber hinaus soll ein Informationssystem die effiziente Erhebung und Modellierung relevanter Serviceinformationen beim Hersteller ermöglichen. Serviceorganisationen sollen auf die durch das System bereitgestellten Serviceinformationen mobil zugreifen können. Zwei wesentliche Implikationen des skizzierten Lösungsansatzes sind:

1. Durch die mit der integrierten Betrachtung verbundene Gestaltung eines neuen hybriden Produkts können die Anforderungen des TKD an eine kundengerechte Inbetriebnahme, Instandhaltung, Wartung und Reparatur von Maschinen und Anlagen gewährleistet sowie die Effizienz des Vorgehens im TKD erhöht werden.
2. Durch die frühzeitige Verzahnung von Produktentwicklung, Dokumentation, TKD, Prozessberatung und moderner Informationstechnologie entstehen hybride Produkte, welche den Lebenszyklus integrierter prozessorientierter Produkt- und Serviceinformationen beim Hersteller erstmals mit vertretbarem Aufwand abbilden.

Insgesamt ergibt sich damit im Projekt PIPE das in Abb. 2 dargestellte Szenario zur hybriden Wertschöpfung im Maschinen- und Anlagenbau, welches als Erweiterung des Status quo der Wertschöpfungskette im SHK-Bereich (vgl. nochmals Abb. 1) zu verstehen ist.

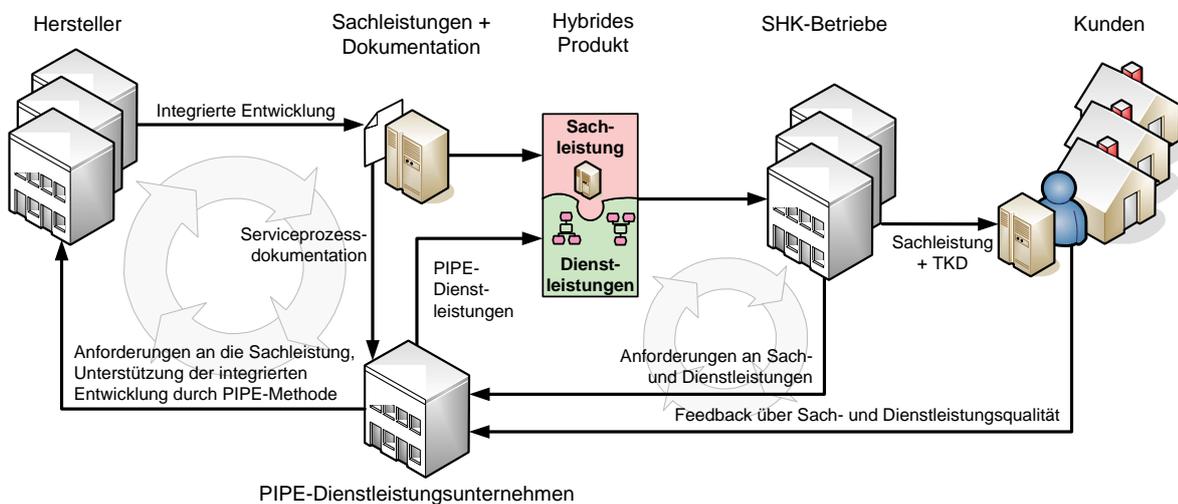


Abb. 2: Hybride Wertschöpfung mit dem PIPE-Konzept

Es bestehen *zwei* Kreisläufe, die zur kontinuierlichen Verbesserung des hybriden Produkts und damit *erstens* zur Verbesserung der Produktentwicklung seitens der Hersteller und *zweitens* zur Verbesserung des Dienstleistungsangebots seitens der SHK-Betriebe beitragen (vgl. Abb. 2). Der erste Kreislauf besteht zwischen dem Gestaltungsprozess des neuen hybriden Produkts

durch den PIPE-Dienstleister² und den beiden Feedbackprozessen, die von den SHK-Betrieben ausgehen (vgl. Abb. 2, rechts). Das Feedback bezieht sich einerseits auf die Anforderungen der SHK-Betriebe an die Komponenten des hybriden Produkts und andererseits auf die Beurteilung der tatsächlich durch die SHK-Betriebe unter Verwendung des hybriden Produkts erbrachten Qualität der Komponenten, die ergänzend durch die Endkunden zu beurteilen ist.

Der zweite Kreislauf besteht zwischen den Herstellern und dem PIPE-Dienstleistungsunternehmen, an das die Hersteller technische Dokumentationen sowie die grundlegenden Informationen über die Serviceprozesse weiterleiten (vgl. Abb. 2, links). In der Gegenrichtung gibt der PIPE-Dienstleister *erstens* das Feedback der SHK-Betriebe an die Hersteller weiter und unterstützt *zweitens* die herstellerseitige Produktentwicklung durch die PIPE-Methodik. Im Ergebnis liegt damit eine *prozessorientierte Integration von Produktentwicklung und Servicedokumentation* vor, die zur Verbesserung des TKD im Maschinen- und Anlagenbau dienen kann.

3.2 Struktur des hybriden Produkts

Im Wesentlichen besteht das zu gestaltende hybride Produkt neben der materiellen Komponente (z.B. Heizungsanlage) aus mehreren Dienstleistungskomponenten. Diese sind, wie in Abb. 3 gezeigt, thematisch um die Servicedokumentation zentriert und in vier Bereiche gegliedert.

Der erste Bereich umfasst alle mit der Serviceprozessmodellierung verbundenen Dienstleistungen: die Entwicklung einer Modellierungsmethode, die Erstellung der eigentlichen Serviceprozessmodelle nach dieser Methode sowie Tests zur Qualitätssicherung der Modelle.

Der zweite Bereich fokussiert die Anwendung der Serviceprozessmodelle, insbesondere die mobile Anwendung und damit die Bereitstellung mobiler Dienste im TKD-Bereich. Dazu zählt zunächst deren Bereitstellung für den Online-Zugriff via Netzwerk sowie den Offline-Zugriff über eine „Stand Alone“-Anwendung auf CD-ROM. Die Übertragung der Serviceprozessmodelle umfasst neben deren Transport auch eine geeignete Konvertierung – die Darstellung auf einem PDA erfordert z.B. eine stärkere Informationsreduktion als auf einem Desktop-PC.

Der dritte Bereich umfasst das Controlling der Serviceprozesse mit den Zielen, Optimierungspotenziale bei den Serviceprozessen aufzudecken und konstruktionsbedingte Schwachstellen in der technischen Anlage selbst zu identifizieren (ein Indiz für eine solche Schwachstelle ist z.B. eine Häufung von Reparaturen an einer bestimmten Baugruppe). Grundlage für das Controlling ist ein Bewertungsschema, das dem TKD-Mitarbeiter ein strukturiertes Feedback ermöglicht,

² Die Darstellung des Geschäftsmodells des Dienstleistungsunternehmens, das die Dienstleistung anbietet, welche eine technische Anlage zum hybriden Produkt aufwertet, ist nicht Gegenstand dieses Beitrags.

das in den Produktlebenszyklus der Anlage und in den Prozesslebenszyklus des Serviceprozesses integriert werden kann.

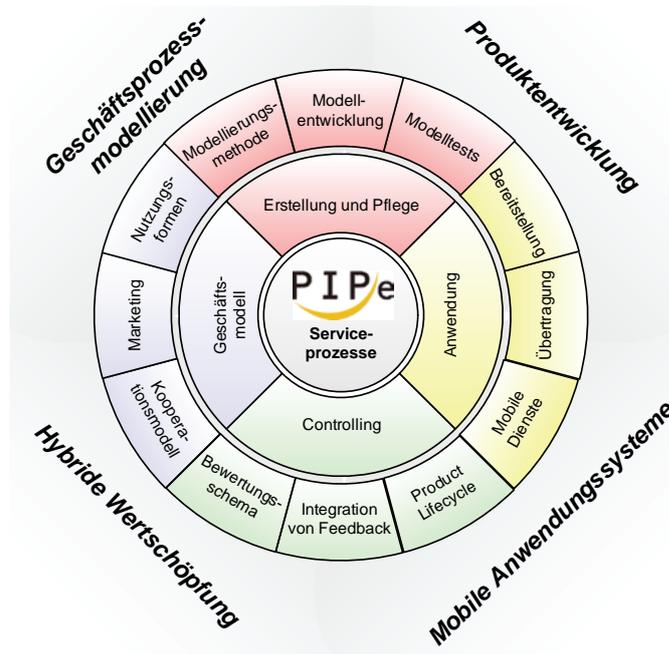


Abb. 3: Struktur des hybriden Produkts nach PIPE

Der vierte Bereich adressiert die wirtschaftliche Nutzung der Serviceprozesse, indem ein Geschäftsmodell mit mehreren Komponenten entwickelt und gepflegt wird. Sein Kern ist ein Kooperationsmodell, das die wirtschaftliche Interaktion der beteiligten Parteien (z.B. Hersteller, Anwender, Modellierer, Portalbetreiber) skizziert und dabei ihren wirtschaftlichen Interessen Rechnung trägt. Weitere Komponenten sind die Identifikation von und Anpassung an neue Nutzungsformen für die Serviceprozessmodelle (z.B. der Einsatz im Rahmen von Schulungen) sowie die Bereitstellung von Werkzeugen für das Marketing.

Dieser hybride Lösungsansatz verdeutlicht, dass das Projekt PIPE stark interdisziplinär angelegt ist und unterschiedliche Themenfelder auf innovative Weise miteinander verbindet, wie in Abb. 3 exemplarisch dargestellt. Die Geschäftsprozessmodellierung fungiert als methodische Grundlage der Serviceprozessmodellierung. Die technische Produktentwicklung stellt die inhaltliche Grundlage der Serviceprozessmodelle dar. Das Forschungsfeld mobiler Anwendungssysteme ist eng mit der mobilen Anwendung und dem Controlling der Serviceprozessmodelle durch den TKD vor Ort verbunden. Das Schema der hybriden Wertschöpfung schlägt sich besonders in der Verzahnung von Produkt- und Serviceprozesslebenszyklus nieder.

3.3 Informationstechnische Konzeption

Basis der Nutzung der Prozessmodellierung für mobile Anwendungssysteme im technischen Kundendienst ist die in Abb. 4 dargestellte PIPE-Systemarchitektur. Sie unterstützt die Erstellung, Bereitstellung und das Controlling der immateriellen Leistungen im hybriden Produkt.

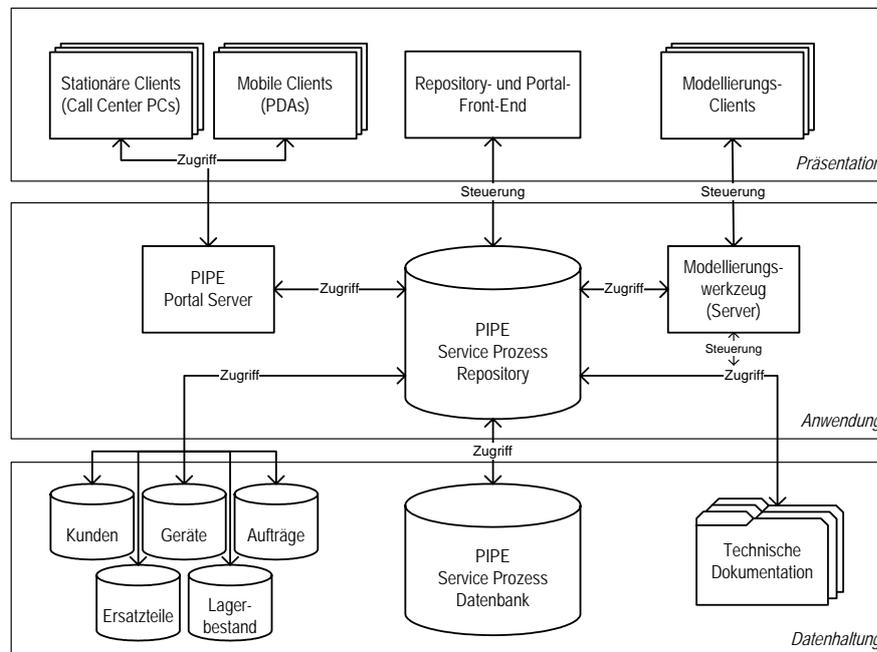


Abb. 4: PIPE-Systemarchitektur

Im Mittelpunkt des hybriden Produkts stehen die mit den Serviceprozessmodellen verbundenen Dienstleistungen der Erstellung und Pflege, der Anwendung und des Controllings. Kern der Architektur bildet daher ein Repository für Serviceprozessbeschreibungen und Verknüpfungen zu damit zusammenhängenden Stammdaten (z.B. Kunden, Geräte, Teile) sowie für technische Dokumentationen, die i. A. in unstrukturierter Form vorliegen (z.B. in Form von PDF-Dateien). Das Repository vereinigt die heterogenen Datenquellen unter einer prozessorientierten Sicht und bildet so die datentechnische Grundlage der in Abschnitt 3.2 umrissenen Leistungen.

Um das Repository herum sind weitere Komponenten angesiedelt, welche die Durchführung der PIPE-Dienstleistungen unterstützen. Erstellung und Pflege der Modelle im Repository sind dezentral über eine Client-Server-Modellierungsanwendung realisiert, bei der auch mehrere Modellierer gleichzeitig über ihren jeweiligen Client auf einen zentralen Modellierungsserver zugreifen können. Der Modellierungsserver unterstützt dabei die Nebenläufigkeit der verschiedenen Modellierungsprozesse, indem etwa simultane Zugriffe auf unteilbare Ressourcen durch Sperren und Freigeben gesteuert werden. Über das Modellierungssystem wird ebenfalls die Einbindung technischer Dokumente in die Modelle gesteuert.

Anwendung und Controlling werden über einen Portalserver unterstützt, auf den die verschiedenen Clients, mobile oder stationäre, zugreifen. Der Portalserver übernimmt dabei die Kommunikation mit den Clients in zwei Richtungen. In der Richtung vom Repository zum Client ist im Portalserver zunächst der Auswahlmechanismus implementiert, über den ein Client ein Serviceprozessmodell suchen und selektieren kann. Im nächsten Schritt wird das gewählte Modell vom Portalserver entsprechend den Anforderungen des Clients konvertiert und zum Client transportiert. Auf dem Client wird der Serviceprozess dann visualisiert und interaktiv unterstützt. In der Rückrichtung vom Client zum Repository nimmt der Portalserver Controllingdaten vom Client entgegen und integriert sie in das Repository. Das Repository und der Portalserver werden über ein integriertes Front-End konfiguriert und gesteuert.

Betreiber des Portals ist das PIPE-Dienstleistungsunternehmen. Somit steht die IT-Architektur in Verbindung mit beiden in Abb. 2 gezeigten Kreisläufen. Der rechte Kreislauf, d.h. die kontinuierlichen Verbesserung der Serviceprozesse, wird hierbei durch die Benutzung der Serviceprozessmodelle und das anschließende Feedback der Benutzer angestoßen, die mit dem PIPE-Portalserver in Kontakt stehen, über den sie Modelle abrufen und Feedback zurücksenden. Das Benutzerfeedback bildet die Schnittstelle zum Modellierer, der die Serviceprozesse mit dem Modellierungswerkzeug auf Basis des Inputs verbessert. Gespeichert werden die Feedback-Informationen zusammen mit den Serviceprozessmodellen im Repository, das aus IT-Perspektive den Schnittpunkt zwischen beiden Kreisläufen darstellt, denn der linke Kreislauf, die Weiterentwicklung des technischen Produkts, wird ebenfalls durch Benutzerfeedback aus dem Repository angestoßen. Bei dem Feedback muss es sich nicht zwangsläufig um dedizierte Bewertungen des Benutzers handeln – eine Analyse der Häufigkeit bestimmter Fehler oder die Verteilung der Fehlerzahl über die Baugruppen eines Gerätes können schon die Notwendigkeit einer konstruktionstechnischen Überarbeitung anzeigen. Aus diesen Informationen kann das PIPE-Dienstleistungsunternehmen Anforderungen an die Sachleistung ableiten und dem Hersteller zur Verfügung stellen. Nach Verbesserung der Sachleistung schließt sich der linke Kreislauf dann mit der Einarbeitung der Änderungen in die Serviceprozessmodelle. Neben aktualisierten Dokumentationen zur Sachleistung, die im Repository abgelegt werden, müssen ggf. auch Serviceprozessmodelle mit Hilfe des Modellierungswerkzeugs angepasst werden.

3.4 Implementierung und Umsetzung

Dieser Abschnitt behandelt die technische Umsetzung der bisher vorgestellten Konzepte. Dabei findet eine Fokussierung auf die Serviceprozesse statt, die den Kern des hybriden Lösungsan-

setzes und damit auch der Systemarchitektur bilden. Im Folgenden sollen die Strukturierung der Serviceprozesse und ihre Einordnung in das Gesamtkonzept im Vordergrund stehen.

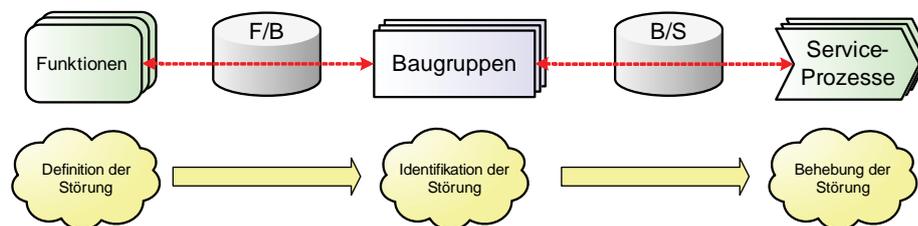


Abb. 5: Modellierung technischer Anlagen mit Funktionen und Serviceprozessen

Kernidee der Serviceprozessmodellierung, die dem PIPE-Ansatz zugrunde liegt, ist die Unterscheidung von Funktion, Aufbau und Serviceprozessen technischer Anlagen. Sowohl die Funktionen als auch die Serviceprozesse einer technischen Anlage sind unmittelbar mit ihrem Aufbau verknüpft. Abb. 5 zeigt, wie dieser Zusammenhang in PIPE abgebildet und genutzt wird, indem auf Basis der Funktionsstruktur die Störung definiert wird, die für die Funktionserfüllung relevanten Baugruppen identifiziert und entsprechende Serviceprozesse vorgeschlagen werden.

3.4.1 Funktionen

Die Modellierung der Funktionen einer technischen Anlage ist die Grundlage für die Definition einer Störung, die als „Ausfall einer Funktion“ definiert wird. Im Gegensatz zu Definitionen möglicher Störungen sind die Funktionen einer technischen Anlage i. A. aufzählbar und – wie im nächsten Abschnitt ausführlich dargestellt – technischen Baugruppen zuordenbar, die letztendlich die Bezugspunkte für die Arbeit des TKD sind.

Die Modellierung der Funktion technischer Anlagen in ablauflogischer Form als Prozess wäre nicht zielführend – die Verortung einer Funktionsstörung im Prozessablauf der Anlage kann zwar ggf. bei der Organisation der Diagnoseschritte helfen („von vorne nach hinten an der Wasserleitung entlang“), für die Störungsdefinition ist diese jedoch nicht relevant.³ Sinnvoller ist eine Gliederung von Funktionen und eine Zuordnung ihrer jeweiligen Teilfunktionen, wie sie in Abb. 6 dargestellt ist. Sie ermöglicht es, Störungen effizient zu lokalisieren. So kann von der allgemeinen Störung der Funktion „Wasser erhitzen“ die Störung auf die Teilfunktion „Brenner betreiben“ und „Heizöl verbrennen“ eingegrenzt werden. Dies ermöglicht einen effizienten Einstieg in die Störungsdiagnose durch die Vorgabe weniger, klar abgegrenzter Alternativen, die zudem einfach zu verifizieren sind („verminderte Heizleistung“, „Rauchentwicklung“).

³ Die Organisation der Diagnoseschritte wird dem Nutzer später vom PIPE-System empfohlen, z.B. auf der Basis von empirisch ermittelten Ausfallhäufigkeiten der einzelnen Funktionen oder dem jeweiligen Diagnoseaufwand.

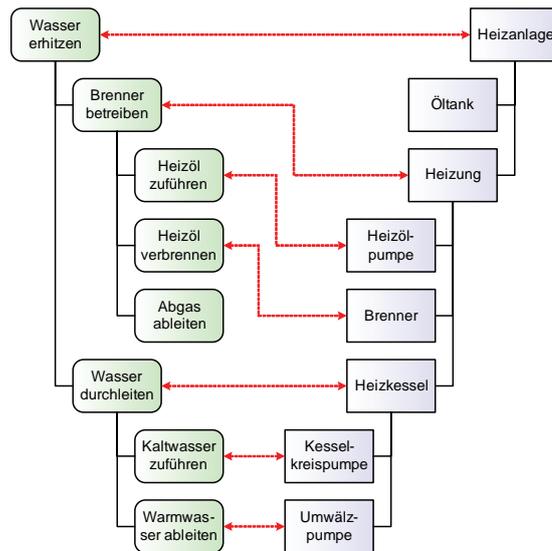


Abb. 6: Relationen zwischen Funktionen und Baugruppen

3.4.2 Baugruppen

Der funktionalen Gliederung gegenüber steht die technische Gliederung der Anlage in verschiedene Baugruppen (vgl. Abb. 6, rechts). Sie ergibt sich aus der technischen Produktentwicklung, bei der zunächst einfache Bauteile hergestellt und dann in mehreren Stufen zu komplexeren Baugruppen zusammengesetzt werden.⁴

Die Verknüpfung von Baugruppen mit Funktionen erfolgt auf Grundlage der technischen Aufgabenteilung – eine Baugruppe wird den Funktionen zugeordnet, für deren Erfüllung sie notwendig ist. Ziel ist dabei die Identifikation von Baugruppen, die für eine Störung als Ursache in Frage kommen. Daraus ergibt sich eine m:n-Beziehung zwischen Baugruppen und Funktionen. In Abb. 6 sind die m:n-Relationen zwischen Funktionen und Baugruppen der Übersichtlichkeit halber vereinfacht als 1:1-Relationen abgebildet. Weiterhin wurde eine Baumstruktur zur Gliederung der Funktionen und Baugruppen verwendet. Hier sind auch andere Ansätze denkbar, z.B. m:n-Relationen zwischen den Funktionen bzw. zwischen den Baugruppen.

3.4.3 Serviceprozesse

Ziel der Störungsdiagnose ist die Störungsbehebung, die durch die Durchführung von Serviceprozessen auf den Baugruppen der technischen Anlage erreicht wird. Dies wird durch eine Verknüpfung von Serviceprozessen mit Baugruppen auf unterschiedlichen Detaillierungsebenen ermöglicht (vgl. Abb. 7). Durch die Zusammensetzung mehrerer Baugruppen ergeben sich

⁴ Bei der industriellen Herstellung kann die technische Gliederung des Produkts als gegeben vorausgesetzt werden, da über die entsprechenden Stücklisten beispielsweise auch die Materialdisposition erfolgt.

Funktionen, die über die Summe der Einzelfunktionen der Baugruppen hinausgehen – eine Heizung kann z.B. Wasser erhitzen, was keine ihrer Baugruppe alleine vermag. Analog ergeben sich bei der Assemblierung von Baugruppen neue Serviceprozesse, welche die Zusammenwirkung der einzelnen Komponenten betreffen. Daher ist auch hier eine hierarchische Anordnung von Serviceprozessen und Teilprozessen vorgesehen, wie Sie z.B. in [ScTh05] erläutert wird. Die Zusammensetzung einer Baugruppe führt so nicht nur zur Übernahme der Prozesse ihrer Komponenten, sondern macht auch die Ergänzung um neu hinzugekommene Prozesse möglich. Sind über die oben beschriebene Funktionsgliederung Baugruppen als mögliche Träger der Störung identifiziert, können über die Verknüpfung mit den Serviceprozessen Maßnahmen zur Behebung der Störung gewählt und durchgeführt werden. Dabei wird auch die Baugruppen-Prozess-Verknüpfung als m:n-Relation abgebildet; das Vorhandensein mehrerer Serviceprozesse für eine Baugruppe (1:n) ist offensichtlich, aber auch umgekehrt kann ein generischer Serviceprozess („Strom abschalten“) auf mehrere Baugruppen (m:1) anwendbar sein.

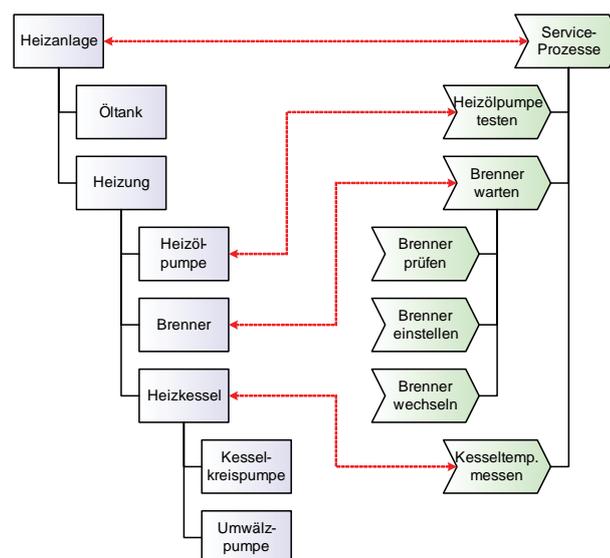


Abb. 7: Relationen zwischen Baugruppen und Serviceprozessen

4 Anwendungsszenario „Warmwasser wird nicht warm“

4.1 Generelle Beschreibung

Der oben dargestellte Lösungsansatz soll durch ein realistisches mobiles Anwendungsszenario exemplarisch veranschaulicht werden. Dieses Szenario beschreibt einen Fehlerfall mit einem Heizgerät zur Erwärmung von Wasser, das der Kundendiensttechniker beim Kunden vor Ort

defekt vorfindet. Ausgangspunkt des Szenarios ist der Fehler „Warmwasser wird nicht warm“, d.h. es ist eine Störung an dem Gerät eingetreten, die seine Funktion beeinträchtigt, sodass das Wasser im Gerät nicht mehr wie vorgesehen erhitzt wird.⁵ Die Bearbeitung dieses Reparaturprozesses stellt hohe Anforderungen an den TKD, da nahezu jedes Bauteil der Heizungsanlage als Ursache für den Fehler in Frage kommt. Folglich ist dieses Szenario ideal geeignet, die grundsätzliche Machbarkeit des in Abschnitt 3 dargestellten Lösungsansatzes zu demonstrieren. Darüber hinaus wird nachfolgend umrissen, wie die PIPE-Systemarchitektur aus Abschnitt 3.3 und die Implementierung aus Abschnitt 3.4 im Anwendungsszenario eingesetzt werden.

4.2 Vorbereitung der Störungsbehebung

Zur Vorbereitung der Störungsbehebung muss der Hersteller zunächst für den Typ des defekten Geräts die notwendigen Serviceprozessinformationen über das PIPE-Repository verfügbar machen (vgl. Abb. 4). Zu den Serviceinformationen gehören die Funktions-, Produkt- und Serviceprozessstruktur (vgl. Abschnitt 3.4) sowie die Serviceprozessmodelle, die mithilfe der Modellierungssoftware erstellt und mit den relevanten technischen Unterlagen verknüpft werden. Im Rahmen dieses Szenarios ergibt sich ein Serviceprozess, wie er in Abb. 8 dargestellt ist.

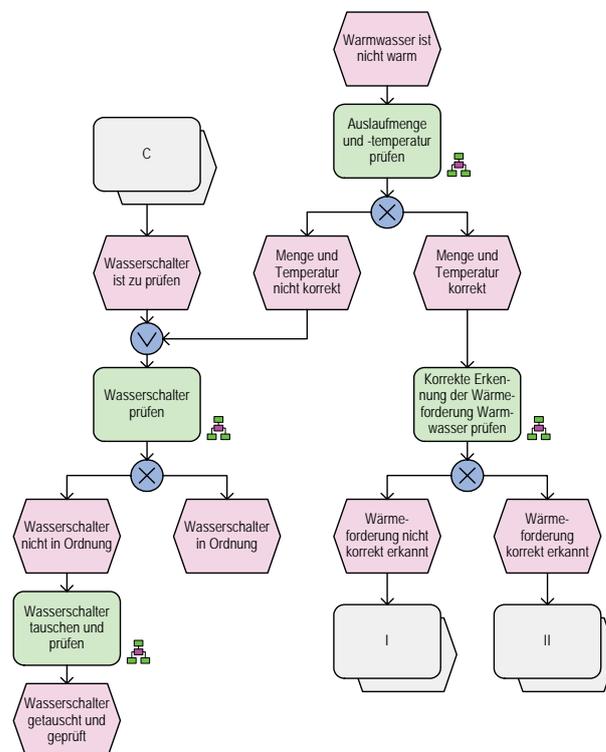


Abb. 8: EPK-Modell „Fehlerbehebung Warmwasser wird nicht warm“ (Ausschnitt)

⁵ Das reale Szenario basiert auf einem Heizgerät des Typs Cerastar ZWR 18–3 KE 23 der Robert Bosch GmbH, Geschäftsbereich Thermotechnik, Produktbereich Junkers.

Als Modellierungssprache wurde die Ereignisgesteuerte Prozesskette (EPK) verwendet [KeNS92]. Das Gesamtmodell enthält auf der obersten Hierarchiestufe 28 Funktionen. Abb. 8 zeigt aus Platzgründen lediglich einen Ausschnitt des Modells. Basis der Modellkonstruktion war die Identifikation derjenigen Bauteile des Heizgeräts, die als Ursache für den Fehler in Frage kommen. Hierbei wurden 8 Bauteile ermittelt und in Abhängigkeit ihrer Bearbeitungsreihenfolge zur Fehlerbehebung geordnet. Jedes der Bauteile wurde auf Prüf- sowie Tausch- und Prüffunktionen abgebildet. Der Reparaturprozess ist um Funktionen zur Erhebung allgemeiner Geräte- und Anlagendaten ergänzt, die sich z.B. auf das Erfassen der Warmwasserauslaufmenge oder das Prüfen des Gasanschlussfließdrucks beziehen.

4.3 Durchführung der Störungsbehebung

Leitmotiv des Projekts PIPE ist es, den TKD zu befähigen, schnell und effizient die geschilderte Störung zu beheben. Dazu erhält der Kundendiensttechniker ein mobiles Endgerät (PDA, Notebook), mit dem er über den PIPE-Portalserver auf das Serviceprozessrepository zugreifen kann (vgl. Abb. 9). Die Störungsbehebung erfolgt in zwei Phasen: Identifikation des Geräts sowie Diagnose und Behebung der Störung.

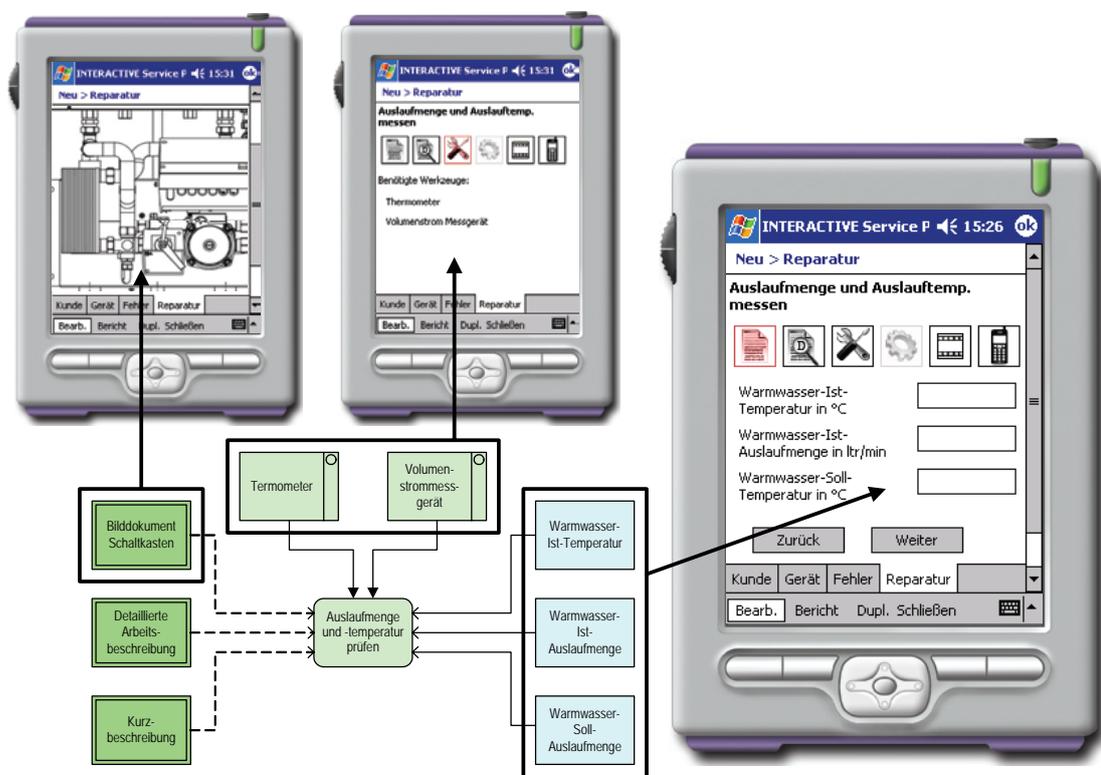


Abb. 9: Mobile Anwendung „Interactive Service Portal“

4.3.1 Identifikation des defekten Geräts

Als ersten Schritt der Fehlerbehebung identifiziert der Kundendiensttechniker das defekte Gerät. Dies kann mobil vor Ort erfolgen, da zur Vorbereitung des TKD-Einsatzes keine gerätespezifischen Unterlagen mitgeführt werden müssen. Durch den Online-Zugriff auf das PIPE-Repository steht dem TKD eine Bibliothek an Serviceinformationen zur Verfügung. Somit entfällt die zeitraubende Identifizierung des Geräts durch den Kunden via Telefon oder durch den SHK-Betrieb auf Basis alter Rechnungen und es müssen keine Serviceunterlagen in Papierform mehr verwaltet, gesucht und transportiert werden. Der Gerätetyp wird nach der Identifikation in den Kontext des Dialogs mit dem PIPE-Server aufgenommen, sodass bei allen nachfolgenden Operationen des TKD nur die für das Gerät relevanten Serviceinformationen sichtbar sind.

4.3.2 Diagnose und Behebung der Störung

Nach der Identifikation des Gerätes steht dessen Produkt-, Funktions- und Serviceprozessstruktur dem Kundendiensttechniker über sein mobiles Endgerät zur Verfügung (vgl. Abb. 9). Nun beginnt der Störungsdiagnoseprozess, wie er in Abschnitt 3.4 beschrieben ist: Auf Basis der ausgefallenen Funktion werden zunächst die Bauteile des Heizgerätes identifiziert, die für den gewählten Fehler in Frage kommen (zum Prozess vgl. Abb. 8). Die Reihenfolge der Diagnoseschritte kann z.B. durch eine empirische Analyse von Ausfallwahrscheinlichkeiten oder durch Aufwandsschätzungen für die Diagnoseschritte dynamisch bestimmt werden. Dabei ist auch die Möglichkeit für den Kundendiensttechniker vorgesehen, den vorgeschlagenen Diagnoseweg zu verlassen. Das kann z.B. notwendig sein, falls eine Störung im System nicht eindeutig abgebildet wird oder fehlt. In diesem Fall stehen zwar keine Serviceprozessmodelle zur Verfügung, technische Dokumente können aber dennoch für die einzelnen Baugruppen abgerufen werden.

4.4 Nachbereitung der Störungsbehebung

Nach der Behebung einer Störung ist eine Bewertung der IT-Unterstützung durch den Kundendiensttechniker vorgesehen. Die gewonnenen Feedback-Informationen werden über den PIPE-Portalserver wieder in das Repository integriert, z.B. durch kontinuierliche Pflege von Metadaten der Serviceprozessmodelle. Der Hersteller kann die aggregierten Feedback-Informationen zu seinen Geräten bspw. zur Verbesserung seiner Serviceprozesse, aber auch der Geräte selbst heranziehen. Ist etwa die Ausfallwahrscheinlichkeit einer bestimmten Baugruppe signifikant hoch, kann eine Änderungskonstruktion in Betracht gezogen werden.

5 Zusammenfassung der Ergebnisse und Ausblick

Der technische Kundendienst im Maschinen- und Anlagenbau steht vor der Herausforderung, seine Serviceleistungen bei stetig steigender Komplexität der technischen Produkte effizient erbringen zu müssen. Im Zentrum des hier vorgestellten Lösungsansatzes stehen Modelle zur effizienten Beschreibung von Serviceprozessen, die mobil kommuniziert werden können. Diese IT-Unterstützung durch mobile Anwendungssysteme ermöglicht es dem technischen Kundendienst, Servicewissen strukturiert vorzuhalten und kontinuierlich zu verbessern. Zur Realisierung des Konzepts ist eine Vielzahl von Technologien und Dienstleistungen notwendig, die z. T. eng miteinander verwoben sind. Daher wurde ein Lösungsansatz durch hybride Wertschöpfung gewählt, in dessen Mittelpunkt ein hybrides Produkt rund um die Serviceinformationen steht. Aus technologischer Sicht wurden eine Systemarchitektur präsentiert, die für die Realisierung des Konzepts geeignet ist, sowie eine Struktur zur Modellierung von Serviceinformationen. Diese einzelnen Punkte wurden anhand eines mobilen Anwendungsszenarios illustriert.

Die Frage, ob es sich bei einer Problemlösung um eine Sachleistung, eine Dienstleistung oder um eine Kombination handelt, wird zukünftig von Seiten eines Kunden immer schwieriger zu beantworten sein und in ihrer Bedeutung in den Hintergrund rücken – die Grenzen zwischen Sach- und Dienstleistung verschwimmen. Für die Wirtschaftsinformatik besteht daher auch in Zukunft eine zentrale Herausforderung darin, die hybride Wertschöpfung durch die Gestaltung adäquater Informationssysteme zu unterstützen.

Literaturverzeichnis

- [BIBB03] Bundesinstitut für Berufsbildung (Hrsg.): *Ein Kundenauftrag im SHK-Handwerk*. Konstanz : Christiani, Techn. Inst. f. Aus- und Weiterbildung, 2003.
- [BIBB04] Bundesinstitut für Berufsbildung: *Anlagenmechaniker/-in für Sanitär-, Heizungs- und Klimatechnik – ein neuer Name oder mehr? Neuordnung*. Bielefeld : Bertelsmann, 2004. – Elektronische Ressource, CD-ROM
- [Bill97] Billesberger, U. B.: *Praxisanleitung zur Stärken-Schwächen-Analyse im Sanitär-, Heizungs- und Klimatechnik-Handwerk*. München : IHW, 1997
- [Breu01] Breunig, L.: *Technischer Kundendienst : Kunden gewinnen und halten mit aktiven Servicestrategien*. Augsburg : WEKA, 2001
- [Bunk04] Bunk, H. D.: *Evaluation als eine Methode der kontrollierten Zielerreichung für Qualifizierungsmaßnahmen im SHK-Handwerk*. Aachen : Shaker, 2004
- [BuSG00] Bullinger, H.-J.; Schuster, E.; Gudszend, T. (Hrsg.): *Optimale Informationsunterstützung für den Technischen Kundendienst : Entwicklungen und Anwendungen im Service des Maschinen- und Anlagebaus ; Forum mit Fachausstellung am 21. März 2000, Stuttgart*. Stuttgart : Fraunhofer IAO, 2000

- [Harm03] Harms, V.: Produktbegleitende Dienstleistungen/Kundendienst. In: Pepels, W. (Hrsg.): *Betriebswirtschaft der Dienstleistungen: Handbuch für Studium und Praxis*. Herne : Verl. Neue Wirtschafts-Briefe, 2003, S. 129–157
- [Harm99] Harms, V.: *Kundendienstmanagement : Dienstleistung, Kundendienst, Servicestrukturen und Serviceprodukte ; Aufgabenbereiche und Organisation des Kundendienstes*. Herne : Verl. Neue Wirtschafts-Briefe, 1999
- [Herm99] Hermes, P.: *Entwicklung eines Customer Self-Service-Systems im technischen Kundendienst des Maschinenbaus*. Heimsheim : Jost-Jetter, 1999
- [HöSa04] Höpfner, H.; Saake, G. (Hrsg.): *Beitragsband zum Workshop "Grundlagen und Anwendungen mobiler Informationstechnologie" des GI-Arbeitskreises Mobile Datenbanken und Informationssysteme, Heidelberg, 23.–24. März 2004*. Magdeburg : Univ., Fak. für Informatik, 2004
- [HoSa96] Hoppe, M.; Sander, M.: SHK – eine Branche im Wandel. *Sanitär + Heizungstechnik* 61 (1996), Nr. 3, S. 38–46
- [KeNS92] Keller, G.; Nüttgens, M.; Scheer, A.-W.: Semantische Prozeßmodellierung auf der Grundlage "Ereignisgesteuerter Prozeßketten (EPK)". In: Scheer, A.-W. (Hrsg.): *Veröffentlichungen des Instituts für Wirtschaftsinformatik*, Nr. 89, Saarbrücken : Universität des Saarlandes, 1992
- [Kirs06] Kirste, T.; Fachgruppe Mobilität und Mobile Informationssysteme (Hrsg.): *Mobile Informationssysteme – Potentiale, Hindernisse, Einsatz : 1. Fachtagung Mobilität und Mobile Informationssysteme (MMS), 20.–22. Februar 2006, Passau, Germany ; [im Rahmen der MKWI]*. Bonn : GI, 2006
- [Kroo66] Krooß, R.: *Der technische Kundendienst als Instrument der Absatzpolitik*. Nürnberg : Spindler, 1966
- [Lehn03] Lehner, F.: *Mobile und drahtlose Informationssysteme : Technologien, Anwendungen, Märkte*. Berlin : Springer, 2003
- [Meff82] Meffert, H. (Hrsg.): *Kundendienst-Management : Entwicklungsstand und Entscheidungsprobleme der Kundendienstpolitik*. Frankfurt a.M. : Lang, 1982
- [Mose87] Mosen, K.: *Marktgerechte Unternehmensführung im Handwerk : dargest. am Beispiel d. Sanitär-, Heizungs- u. Klimatechnik-Branche*. Stuttgart : Gentner, 1987
- [Muse88] Muser, V.: *Der integrative Kundendienst : Grundlagen für ein marketingorientiertes Kundendienstmanagement*. Augsburg : FGM-Verl, 1988
- [ScTh05] Scheer, A.-W.; Thomas, O.: Geschäftsprozessmodellierung mit der ereignisgesteuerten Prozesskette. *Das Wirtschaftsstudium* 34 (2005), Nr. 8–9, S. 1069–1078
- [ScWa04] Schlagnitweit, H.; Wagner, H.: *Sanitär- und Klimatechnik. Heizungs- und Lüftungsinstallation*. Wien : Bohmann Fachbuch im Verlag Jugend & Volk, 2004
- [Teic94] Teichmann, J.: *Kundendienstmanagement im Investitionsgüterbereich : vom notwendigen Übel zum strategischen Erfolgsfaktor*. Frankfurt a.M. : Lang, 1994
- [VDMA06] VDMA (Hrsg.): *Maschinenbau in Zahl und Bild 2006*. Mühlheim am Main : reufurth, 2006. – Stand Februar 2006
- [West04] Westphal, P.: *Anlagenmechaniker, Anlagenmechanikerin für Sanitär-, Heizungs- und Klimatechnik : Umsetzungshilfen zum neu gestalteten Ausbildungsberuf*. Nürnberg : BW Bildung und Wissen, Verl. und Software, 2004
- [Will87] Willerding, T.: *Gestaltungsmöglichkeiten der Kooperation im technischen Kundendienst zwischen Hersteller und Handel*. Bochum : Brockmeyer, 1987

Einführung in den Track

Semantic Web: Web-Infrastruktur für die nächste Generation von E-Business

Prof. Dr. Jürgen Angele

Ontoprise GmbH

Prof. Dr. Rudi Studer

Universität Karlsruhe (TH)

Prof. Dr. Dieter Fensel

Universität Innsbruck

Die Vision des Semantic Web zielt darauf ab, den Inhalt von Informationsquellen sowie die Funktionsbeschreibungen von Services, die im Internet oder in Intranets angeboten werden, maschinell interpretierbar und damit einer inhaltlichen Verarbeitung zugänglich zu machen. Damit können Aufgaben wie Informationsintegration oder die Komposition von Web Services prinzipiell automatisiert werden. Diese Entwicklung ermöglicht auch neue Arten von Applikationen oder Organisationsformen, wie z.B. virtuelle Organisationen.

Der Track "Semantic Web" wird aus Sicht von Forschung und Praxis zum einen aufzeigen, welche neuen Methoden und Werkzeuge gerade entwickelt wurden bzw. in Entwicklung sind und wie damit eine neue Qualität von Anwendungen erreicht werden kann. Zum anderen werden aus Sicht der Praxis Anforderungen an zu entwickelnde Konzepte und Methoden sowie realisierte Anwendungen diskutiert werden.

Programmkomitee:

Dr. Andreas Abecker, FZI Karlsruhe

Dr. Manfred Hauswirth, EPFL Lausanne

Prof. Dr. Matthias Jarke, RWTH Aachen

Dr. Rainer Malaka, EML Heidelberg

Prof. Dr. Heiner Stuckenschmidt, Universität Mannheim

Dr. York Sure, Universität Karlsruhe (TH)

Dr. Ubbo Visser, TZI Bremen

An Ontology Framework for Semantic Business Process Management

Martin Hepp, Dumitru Roman

Semantics in Business Information Systems (SEBIS), DERI Innsbruck
University of Innsbruck, A-6020 Innsbruck, Austria
mhepp@computer.org, dimitru.roman@deri.org

Abstract

A core challenge in Business Process Management is the continuous, bi-directional translation between (1) a business requirements view on the process space of an enterprise and (2) the actual process space of this enterprise, constituted by the multiplicity of IT systems, resources, and human labor. Semantic Business Process Management (SBPM) [HeLD'05]¹ is a novel approach of increasing the level of automation in the translation between these two spheres, and is currently driven by major players from the ERP, BPM, and Semantic Web Services domain, namely SAP². One core paradigm of SPBM is to represent the two spheres and their parts using ontology languages and to employ machine reasoning for the automated or semi-automated translation. In this paper, we (1) outline the representational requirements of SBPM, (2) propose a set of ontologies and formalisms, and (3) define the scope of these ontologies by giving competency questions, which is a common technique in the ontology engineering process.

1 Introduction

Business Process Management (BPM) is the approach of managing the execution of IT-supported business operations *from a managerial process view* rather than from a technical perspective [SmiFi'02]. However, the degree of automation in BPM is still unsatisfying. Also, BPM does not provide a uniform representation of an organization's process space at a semantic level, which would be accessible to intelligent queries or for compliance checks. BPM often includes the modeling of processes in some form. While process modeling is a traditional and well-established topic in Information Systems research, the various possible motivations for modeling a process, the various sources of models, and the resulting variety of requirements on

¹ In order to make this paper self-contained, we include a summary of some materials already presented there in here.

² The most prominent example of this approach is the research project „SUPER“, lead by SAP, with a total project volume of more than 16 Million euro; see <http://www.ip-super.org>

the formalisms used for representing processes are often not considered. This contributes to the dominance of a simplified, workflow-centric view on business processes, i.e. business processes are reduced to the sequencing of activities. Evidence of this workflow-minded notion of processes is that languages and tools for modeling business processes focus on control flow patterns. This is in particular true for BPEL [cf. ACGD'03]. It is only recently that the weaknesses of a merely workflow-centric representation were pointed out by van der Aalst and Pesic [AaPe'06]. In parallel, there has been substantial work on a more comprehensive and richer conceptual model of enterprises and their processes in the “enterprise ontology” research community, see e.g. [FoGr'98], [FBGL'98], [GrAF'00], or recently [Diet'06]. However, process models that could also be executed in production systems were not at the core of this community’s interest. As a consequence, workflow-centric process representations and work on enterprise ontologies are largely unconnected, which contributes to two current shortcomings: Firstly, workflow-centric process representations are not very suitable for accessing the business process space at knowledge level (e.g. for the discovery of processes or process fragments that can serve a particular purpose). Secondly, models created by the “enterprise ontology” community cannot be used with current, workflow-centric BPM tools and infrastructure.

With regard to combining a comprehensive conceptual model of an enterprise with the actual production system and executable workflows, the ARIS methodology [Sche'98] and respective tooling support was a major step. ARIS includes not only the control flow perspective, but also the organizational, data, and functional dimensions of enterprises [Sche'98]. While ARIS is based on such a comprehensive conceptual framework, it has two weaknesses: (1) the expressiveness and degree of formality in the various models is rather limited and (2) the links between the various models are quite weak. This limits the degree of automation in ARIS-based BPM; it also makes access to the enterprise as a whole on a semantic level impossible. With the latter, we mean for instance querying the process space using logical expressions and machine reasoning.

Semantic Business Process Management [HeLD'05] is a novel approach of increasing the level of automation of BPM by representing the various spheres of an enterprise using ontology languages and Semantic Web Services frameworks. The goal is to be able to apply machine reasoning for the translation between these two spheres, in particular for the discovery of processes and process fragments and for process composition. The gain in automation can be because (1) tasks can actually be performed by machines or (2) because SBPM allows for more

intelligent tooling support in BPM, e.g. by detecting conflicts or constraint violations in composite process models assembled by humans. The general vision of Semantic Business Process Management has already been presented in [HeLD'05], and a large group of academic and industrial partners is working on the respective research challenges. However, there does not yet exist a common understanding of the representational requirements of SBPM, i.e. which ontologies and formalisms will be needed, what they shall be used for, and how existing standards, namely BPEL and EPCs, can be integrated into that framework. In particular, there exists no comprehensive stack of respective ontologies available in popular ontology formalisms, e.g. OWL [W3C'04] or WSML [dBLK'05].

1.1 Our Contribution

In this paper, we (1) outline the representational requirements of Semantic Business Process Management, (2) propose a set of ontologies and formalisms, and (3) define the scope of these ontologies by giving competency questions [cf. UsGr'96], which is a common approach in the ontology engineering process. This is an important step in actually building these ontologies.

1.2 Related Work

A substantial part of Information Systems literature deals with aspects of modeling business processes. Thus, a full survey of related literature is obviously outside the scope of this paper. The most important streams of work closely related to ours can be classified as follows:

Enterprise Ontology: Wand and Weber have proposed to use the philosophical discipline of Ontology for building and maintaining information systems. Based on the work by Bunge, they provided a set of “Bunge-Wand-Weber” models for the analysis of IS modeling tasks; see [Webe'97] and [RIRG'06]. As part of the “Toronto Virtual Enterprise (TOVE)” project, Gruninger, Fox, and several others have developed formal ontologies for various aspects of an enterprise, see e.g. [FoGr'98], [FBGL'98], and [GrAF'00]. While much of their work can likely be reused for SBPM, their ontologies are to our knowledge not available in current ontology languages. Recently, Dietz has provided a comprehensive ontological approach to enterprise modeling [Diet'06]. Since his work is not directly linked to common ontology languages and Semantic Web services frameworks, however, the contribution to process composition in Services-oriented Architectures (SOAs) and intelligent queries to the process space is limited.

Business Process Modeling Languages and Standardization Initiatives: The BPM and Web Services communities have yielded a wealth of languages and standardization approaches that

aim at describing business processes, especially from the perspective of Web Services orchestration. The most prominent example is now BPEL4WS [ACGD'03], but other activities covering a varying scope are around, e.g. BPML [Arki'02], BPMN as a complementing graphical notation [BPMI'04a], XLANG [That'01], WSCI [AAFJ'02], and WS-CAF [BCHL'03]. In short, all those focus only on the representation of a part of the process space, mainly the patterns of message exchange (choreography) and the control and data flow in the combination of multiple Web services (orchestration). They were not designed as fully-fledged languages for knowledge representation, but rather for defining process models that can be executed or for abstract processes that can be refined and turned into executable processes later. The shortcomings of a procedural modeling style for the representation of business processes has been recently pointed out by van der Aalst and Pesic; they also proposed a “Declarative Service Flow Language” (DecSerFlow) as a cure [AaPe'06].

Semantic Web Services (SWS): OWL-S [MBDH'04] is a framework and ontology for adding semantics to Web service descriptions. WSMF is a more comprehensive framework [FeBu'02], which has been further developed to the Web Service Modeling Ontology (WSMO). The core specification of WSMO can be found in [RLKB'05], a brief introduction is given in [FeDo'05]. WSML [dBLK'05] is a family of fully-fledged ontology representation languages that supports WSMO. SWS are currently subject to intense research, especially in the DIP project³. The very recent work by Oberle [Ober'06] is very related to our approach, as it is to our knowledge the first approach to combine Semantic Web Services and the management of enterprise IT landscapes.

Workflow management: For an overview of production workflow management including the role of business processes and their modeling, see for example [LeRo'00]. [Leym'96] discusses the impact of using workflow technology on the creation of applications. [LeRo'97] describes an overall environment for modeling, testing, deployment, running, analyzing applications based on business processes (i.e. the lifecycle).

2 Semantic Business Process Management

The basic “semantic bottleneck” in Business Process Management can be described as follows (see also [HeLD'05]): Although a significant part of an enterprise and its process space is already stored in computer systems (e.g. in the form of process models, code fragments as

³ <http://dip.semanticweb.org>

activity implementations, data structures, data, system links, etc.), both *querying* and *manipulating* the process space regularly requires human intervention. When we need a list of processes or enterprise resources that meet a particular definition, we have to ask a business analyst to compose such a document manually, taking into account a lot of implicit knowledge. This implicit knowledge includes in particular generalization and specialization relationships. Modifying the process space, in turn, requires programmers or system engineers to modify program code, change ERP customizing data, or extend database structures in order to materialize a change request in the actual process space. Obviously, there is a functional bottleneck between (1) the *business perspective* on operations and (2) the *actual execution* of operations [cf. BPMI'04b]. In other words, *the fundamental problem is that traversing from one sphere to the other requires manual labor in any of the two directions, i.e. both for querying and manipulating the process space*: If a manager needs to know all billing processes, systems analysts have to try to create an inventory of any such processes; and if a manager needs a new billing process for a new product or service, software engineers have to transform the management requirements into an IT implementation. This leads to the situation that business-process-related activities are, amidst a wealth of IT, surprisingly centric to human labor, and thus slow, costly, and imperfect.

This gap has been already targeted by the emerging field of Business Process Management (BPM) [SmiFi'02]. BPM modeling tools usually put a strong emphasis on the graphical representation of processes, augmented with middleware for workflow and, often, Enterprise Application Integration (EAI) functionality. However, current BPM does not overcome the underlying limitation that *the business process space inside the organization as a whole is not accessible at a semantic level*, especially because business process modeling languages like BPEL4WS [ACGD'03] are an insufficient means of capturing and representing such a domain of discourse.

2.1 Vision

It can be expected that Business Process Management will only come closer to its promise if it allows for a better automation of the bi-directional translation between the business and the systems spheres. The basic idea of Semantic Business Process Management is to combine Semantic Web Services frameworks, ontology infrastructure, and Business Process Management methodologies and tools, namely ARIS, and to develop one consolidated technology. The fundamental approach is to represent both the business perspective and the

systems perspective of enterprises using a set of ontologies, and to use machine reasoning for carrying out or supporting the translation tasks between the two spheres. This principle is shown in Figure 1.

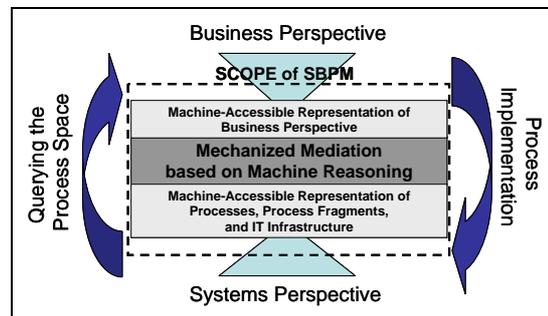


Figure 1. Semantic Business Process Management: Capturing business and IT spheres using ontologies [HeLD'05]

One paradigm of current SBPM research is to provide as much compatibility to existing tools and standards as possible. In particular, processes represented in BPEL or as EPCs should be usable inside an SBPM environment. Also, it should be possible to export processes configured inside an SBPM environment to BPEL so that standard BPEL engines could be used for their execution.

2.2 SUPER Overall Architecture

Currently, the following structure for Semantic Business Process Management emerges as part of the SUPER research project [Leym'06]: First, a **Semantic Service Bus**. This is the foundational layer that deals with brokering service requests from the upper layers. Second, a **Semantic Process Engine**. This is a process engine that instantiates and executes business process models and handles the related service requests from the service bus. Third, the **Semantic Process Modeling Layer**. This layer stores models of business processes plus all additionally relevant spheres of the enterprise, e.g. regarding the organization, resources, etc.

Making SBPM a reality includes a plethora of theoretical and practical challenges. In this paper, we focus on the ontologies and formalisms that constitute the conceptual framework.

2.3 Origins of Process Models

Prior to defining the conceptual framework for SBPM, it is crucial to stress that there are at least three distinct types of sources for business process models that need to be dealt with.

Tools-based, Manual Modeling: The most obvious source of process models are modeling tools used by humans for modeling either existing processes (documenting the “as-is”) or target processes (defining the “should-be”). Typical notations are EPCs [KeNS'91], various

formalisms related to Petri nets, UML activity diagrams, or, more recently, BPMN [BPMI'04a; RIRG'06].

Reference Process Libraries: The gain in spending substantial effort in manually modeling the current status (“as-is”) has been questioned, in particular in ERP-centric corporate environments; instead, Thome and Hufgard have called for dedicating more attention to the implicit process library contained in large Common-of-the-Shelf software packages [ThHu'96]. As a matter of fact, there exists large bodies of process specifications that can be regarded as libraries of best practice; they are an important source of process models for enterprises. Besides reference process models in ERP packages, this includes standards (e.g. RosettaNet PIPs⁴) or the MIT Process Handbook [MaCH'03].

Reconstruction: Process Mining and Reverse Business Engineering: The third source of process models are tools that reconstruct processes from log files and from the customizing data of existing systems. An early description of the fundamental idea of process mining was given by Agrawal, Gunopulos, and Leymann [AgGL'98]. A comprehensive description of challenges in process mining is given in [AaDH'03]. [GrGM'05] describes algorithms for mining workflow management systems for frequent patterns of workflow execution. Very recent work by Jansen-Fullers, van der Aalst, and Rosemann describes work on process mining in the context of ERP solutions [JaAR'06]. Reverse Business Engineering [Ibis'05] is a methodology and family of toolsets that extract and interpret transactional data and program module usage in ERP systems, namely SAP R/3 and mySAP, in order to analyze the process space of an organization. While the main goal of RBE is not the creation of process models, it can be used to select from the ERP reference processes those variants that are in use in the given environment. Process Mining and Reverse Business Engineering can be both the source of process models and at the same time also benefit from richer representations in SBPM environments.

2.4 Process Lifecycle in Semantic Business Process Management

In this section, we describe the process lifecycle in a minimal SBPM environment. It takes into account market structures (e.g. the dominance of EPC- and BPEL-centric tools and BPEL execution engines) and aims at combining the essence of the vision with a realistic scenario. Figure 2 shows the respective lifecycle.

Primary input to an SBPM environment are process models coming from modeling tools, standards, reference process libraries given by ERP vendors, reverse business engineering and

⁴ <http://www.rosettanel.org/>

process mining tools (shown on the left side). These process specifications come in two major types of formalisms: (1) Standardized, control-flow-centric (“procedural”) process notations, e.g. BPEL or EPC, and (2) proprietary representations.

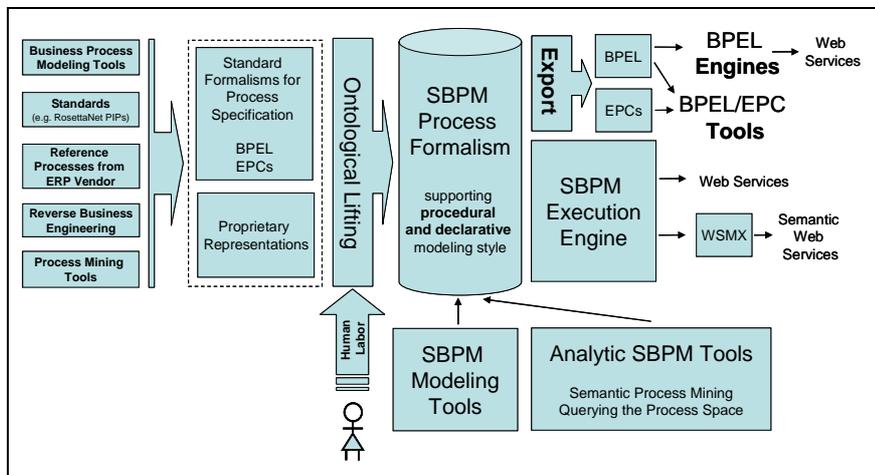


Figure 2. Process Lifecycle in Semantic Business Process Management

All this input must be ontologically lifted, i.e. all data must be augmented by references to ontologies used in the framework. In the case of BPEL processes, this may mean converting a BPEL file to a semantically enriched process specification. Ontological lifting requires almost always human intervention, since the formal semantics of input data must be augmented by annotations or expressed using richer constructs, provided by the target ontologies. A typical example is that the name of the resource is to be replaced by the unique identifier of this resource in the domain ontology.

Once the process specifications are ontologically lifted, they can be stored in the process repository. For this, a rich process formalism is being used that allows both preserving “rigid” control flows of existing workflows, as well as adding declarative specifications, references to resources, and in general any kind of relations to other elements in the SBPM enterprise ontology. This rich SBPM process representation can be accessed by native SBPM modeling tools to create or modify such models directly. It can also be used by native SBPM analytic tools, which may be used to inspect the process space on the basis of this formalism. From the SBPM process formalism, one may export abstract or executable BPEL processes for the use with existing BPEL tooling. These will be procedural reductions of the SBPM process representations that include a defined control flow.

Additionally, the SBPM execution engine can be used to directly execute such SBPM process models. For process models that include the definition of a control flow (e.g. imported BPEL processes), this control flow may be used. For declaratively specified processes, several options

exist: A specific reasoner can be used to create a valid control flow from an existing declarative process. Alternatively, an SBPM tool can be used to guide the user in creating a control flow that meets all constraints. The same two options can also be used for exporting “rigid” control flows in the form of BPEL from SBPM models, in order to have them executed by existing workflow engines. Subsets of SBPM models can also be exported as EPCs for the use in existing EPC-based tooling environments.

3 The SUPER Set of Ontologies for Business Process Management

In this section, we describe a set of ontologies for Semantic Business Process Management that is compatible with the SUPER approach while providing backward compatibility to the ARIS methodology, EPC tools, and BPEL engines and tooling support.

3.1 Methodology

We take the ARIS model of integrated business information systems [Sche'98] as a starting point. ARIS defines the following four views on enterprises: (1) Organization, (2) Data, (3) Control, and (4) Function. For each of these views, we identify suitable ontology sets. Then, we complement the four views by additional spheres, which we assume to be relevant for SBPM tasks. Inside these sets, we try to determine suitable upper ontologies that can be used to ground various more detailed ontologies. It should be stressed that the upper ontologies will have a crucial impact on the degree of automation and interoperability, for they define the common subsets between data from various sources.

3.2 Spheres Relevant for Modeling

In this section, we describe the spheres that contribute to the process space of an enterprise and which need to be represented in a SBPM framework. Figure 3 summarizes these spheres.

Processes: With this we mean the chains of activities that are actually executed, e.g. explicitly designed processes as well as ad-hoc processes.

Process Models: With process models we mean explicit *specifications* of processes, which can be abstract definitions or executable specifications; they may be expressed in a declarative or procedural style. Such models will usually come from past modeling done by humans, software packages in use, or be brought in as requirements from outside the enterprise, e.g. process standards which must be supported in order to participate in a particular value chain.

Organization: Actors in enterprises are bound by the inner structure of the enterprise, e.g. the roles performed by the actors and the line of authority and reporting. Also, all tangible and

intangible resources (e.g. machinery, intellectual property, ...) are part of the organization of an enterprise.

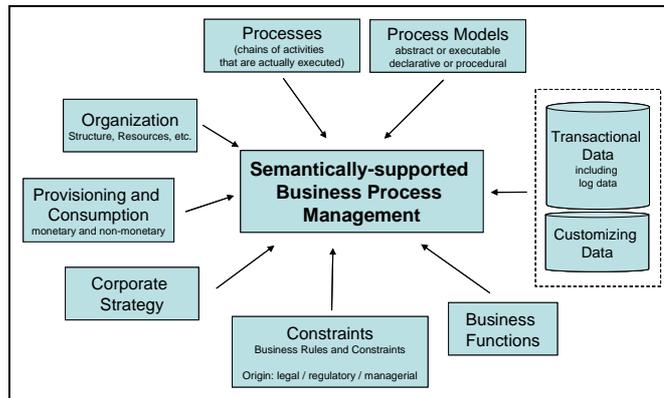


Figure 3. Contributing Spheres

Corporate Strategy: Strategic goals of an enterprise do also influence the process space; in particular, they should be available for ranking alternative process compositions.

Constraints: The process space is further influenced by constraints caused by legal, regulatory, or managerial rules. For instance, if an enterprise has declared to use only organic food in all corporate restaurants or if there if the law of the state prohibits that pregnant employees operate a particular class of machinery, this must be represented for SBPM.

Business Functions: The functional sphere of an enterprise, e.g. typical business functions like Accounting or Human Resources.

Transactional and Customizing Data: The enterprise data stored in corporate databases should also be represented in a SBPM framework. This includes the transactional data and the customizing data, i.e. the settings and process parameters of extensively configurable ERP systems.

In addition to these spheres, there are obviously additional, domain-specific spheres that may need to be included in particular SBPM settings.

3.3 Proposed Ontologies and Formalisms

In this section, we propose a set of ontologies and formalisms for the spheres identified above.

3.3.1 Core Process Modeling: Orchestration

The weaknesses of a workflow-centric representation were recently pointed out by van der Aalst and Pesic [AaPe'06], in particular that capturing a particular sequencing of activities is an over-specification of a business process, since it introduces unnecessary constraints on the execution of this process. If for example, a technical process requires that at least two other tasks are carried out between washing an item and painting it, this is usually expressed in a

workflow-centric modeling language by giving a sequence of activities that meets this constraint (i.e. by putting two other tasks in between), but not by actually capturing this constraint. This has several disadvantages. First, it is much harder to determine whether one process can substitute another, since there is no declaration of the *cause* of this particular ordering. Second, the particular process is bound to a particular set of resources. For instance, if there are additional workstations available, one does not immediately know whether it is valid to perform two tasks in parallel that are modeled as a sequence in the original process. While some of such problems may be solved by avoiding modeling pitfalls, it is quite obvious that a workflow-minded modeling style for processes puts more emphasis on *how* a process is actually executed than necessary. Third, workflow-centric modeling does not clearly distinguish between the constraints of a process (i.e. what must be met) and the actual execution of a process (i.e. how the process is carried out). Fourth, additional knowledge about the process space, e.g. generalization and specialization relationships between activities is not represented. SBPM aims at accessing the process space of an enterprise at a knowledge level, e.g. we want to be able to reason about process composition and substitution and about whether there exists a valid order of execution so that a particular set of Web services with explicit constraints can be employed to constitute a given business process. We may also want to specify processes in a declarative manner. This means that a process ontology that is mainly built around workflow patterns will not be sufficient. On the other hand, we agree that there is sometimes a need to represent rigid workflow patterns.

As a consequence, we propose to define a set of four ontologies for representing processes in SBPM: First, an *Upper Process Ontology* that captures the very basic notions of processes, in particular “activity” and “pre-/post-condition”. This ontology shall also be able to capture existing workflow specifications for a process. Thus, it should also include elements for the most important workflow patterns. Additionally, we propose ontological representations of the three most popular formalisms and notations for processes, i.e. BPEL, BPMN, and EPCs. With “ontological representation”, we mean that we create ontology elements (e.g. concepts and relations) for all (or a subset) of the elements of the respective language and *ground them in the Upper Process Ontology*.

In the following, we give *Competency Questions* [UsGr'96] for the **Upper Process Ontology**:

1. Which is the set of activities constituting a particular process?
2. Which are the pre-conditions, post-conditions, assumptions, and effects of a particular process?

3. For each activity in a particular process, what are the pre-conditions, post-conditions, assumptions, and effects?
4. What is a valid execution order of the activities in a particular process?
5. If the process includes a procedural specification, what is the specified order of execution (i.e.: by which workflow patterns are the activities connected)?
6. If the process includes a procedural specification, is activity 1 {always | sometimes | never} executed {before | directly before | after | directly after | in parallel to} activity 2?

With *pre-conditions*, we mean the information space of an activity or process prior to its execution. With *assumptions*, we mean the necessary state of the world prior to the execution of the activity or process. With *post-conditions*, we mean the information space an activity or process after its execution. With *effects*, we mean the state of the world after the execution of the activity or process. These definitions are taken from [RLKB'05; FeDo'05].

Elements of this ontology will be e.g. *Process*, *Activity*, *Assumption*, *Pre-condition*, *Post-condition*, *Effect*, and typical workflow patterns (e.g. *follows (activity1, activity2)*, *loop (activity1, exit condition)*). For these patterns, we will likely draw on the rich literature on workflow patterns.

It is important to note that these two ways of modeling a process in the Upper Process Ontology are complementing each other; there may even be conflicts such that the specified execution order and the pre-conditions do not match. However, this is something that needs to be dealt with on the operational level. The aim of this foundational ontology is to provide one common point of reference for procedural and declarative process specifications.

The three other ontologies will be ontology variants of popular process modeling formalisms, which will be grounded by references to the Upper Process Ontology. In detail, we will develop **sBPEL**, an ontology version of BPEL [ACGD'03] or a reasonable subset of BPEL; **sBPMN**, i.e. an ontology version of BPMN [BPMI'04a] or a reasonable subset of BPMN; and **sEPC**, an ontology version of EPCs [KeNS'91] or a reasonable subset of EPCs.

3.3.2 Organization and Resources

For the ontological representation of the organizational sphere of enterprises, we can draw upon the rich work done as part of the TOVE project by Fox, Gruninger, and others; see e.g. [FoGr'98], [FBGL'98], or [GrAF'00]. They consider an organization “to be a set of constraints on the activities performed by agents” [FBGL'98]. This ontology provides the vocabulary and constraints for describing the environment in which processes are being carried out. For example, this ontology may be used to describe the actors and the machinery that are being referred to in the process definitions. We propose three major building blocks: First, an *Upper Organizational Ontology* that provides the basic vocabulary and structure for describing

organizations and resources. Second, a *Business Organization Ontology* that defines common types of divisions, roles, and tasks; and third, a *Business Resources Ontology* that defines common types of business resources.

In the following, we give *Competency Questions* for the **Upper Organizational Ontology**:

1. Which employees are members of {the organization | a particular division | a particular subdivision}?
2. Which employees perform a particular role?
3. Which roles does a particular employee perform?
4. Which are the skills of a particular employee?
5. Which are the capabilities of a particular {machine | intangible resource}?
6. Which skills or capabilities are required for a particular role?
7. Who supervises a particular employee?
8. Whom does a particular employee report to?
9. Whom does a particular {employee | system} communicate with?
10. Which resources exist in the enterprise?
11. Which resources are required for a particular task?

These may be extended by aspects of authority and empowerment, e.g. “what resources does an actor have authority to assign?” [FBGL'98].

Elements of this ontology will be e.g. the concepts *Organization*, *Role*, *Task*, *Division*, *Resource*, *Employee*, *MachineOrSystem*, *IntangibleResource*, *SkillCapability*, and the relations *subordinateRoleOf* (*a,b*), *generalizedRoleOf* (*a,b*), *specializedRoleOf* (*a,b*), *reportsTo* (*a,b*), *communicatesWith* (*a,b*), and *supervises* (*a,b*).

The **Business Organization Ontology** will refine the concepts *Role*, *Task*, *Division*, *Employee*, and *SkillCapability* by common types, e.g. *StaffMember*, *Manager*, *CEO*, etc.

The **Business Resources Ontology** will refine the concepts *Resource*, *MachineOrSystem*, and *IntangibleResource* by common types of such things, e.g. *WeldingStation*, *Galvanization-Vessel*, etc. Note that this ontology will just describe the static aspect of resources, not their consumption.

3.3.3 Functions

The processes and process models (and the other elements listed in here) must be regarded in the context of the underlying business functions, i.e. the functional perspective on a enterprise. This includes typical functional groups (e.g. “accounting”) and typical tasks (e.g. “print paycheck”). It is important to note that this is a perspective of its own right, even though there seems to be overlap with the process model and organization sphere. For this, we propose a **Business Functions Upper Ontology** which represents common business functions (e.g. *Marketing*, *Human Resources*, *Operations*, ...) and typical tasks (i.e. activity types). This can

be used as the basis for an enterprise-specific functional ontology. Such grounding will ease several tasks, e.g. benchmarking or systems integration after Mergers & Acquisitions.

3.3.4 Data

The transactional data, usually kept in a relational database management system, is also part of the process space of an enterprise. It is desirable that such data can be included into the unified view on the enterprise. This will likely happen by representing (1) the data structures and (2) the events that create or consume instance data. In ERP-centric IT landscapes, (3) the customizing data of the system must also be regarded, since from this, the set of processes from the process library which are actually enabled can be derived. Also, for processes that can be configured in more detail, the process parameters can be determined.

We propose three ontologies for capturing the data sphere: First, an *Enterprise Data Upper Ontology* that can be used for describing the available data sources and recipients in an enterprise, all metadata, and all events that generate or consume data. (It must be decided on a business level later on whether all such events should be actually annotated; however, a respective vocabulary should be available). Second, a *Transactional Data Ontology* that describes the relational schemas (and their business meaning) of all relevant data sources, and third a *Customizing Data Ontology* that represents the customizing data of ERP systems.

In the following, we give *Competency Questions* for the **Enterprise Data Upper Ontology**:

1. Which database management systems exist in the enterprise?
2. Which databases are in use in the enterprise and on which systems are they located?
3. Which tables (relations) belong to a particular database?
4. Which {business function | task | event | role | machine or system} {writes to | reads from} a particular {data base | table} ?

Elements of this ontology will be e.g. the concepts *DBMS*, *Database*, *DatabaseTable*, and the relations *writesTo* (*a,b*) or *readsFrom* (*a,b*).

3.3.5 Provisioning and Consumption

So far, we have not catered for the fact that the main aspect of business operations is the consumption and production of scarce resources. For the comparison of process alternatives or for activity-based costing, it is necessary to represent the sphere of consumption and provisioning, both of monetary and non-monetary amounts. As a first approximation, we propose a minimal **Provisioning and Consumption Ontology** and define it using the following *Competency Questions*:

1. Which activity {consumes | occupies} [which amount of] which resource at which location at which point in time [for which duration]?

2. Which activity {provides | creates} [which amount of] which resource at which location at which point in time [for which duration]?

3.3.6 *Business Logics: Rules and Constraints*

The process space of an enterprise is also influenced by legal and regulatory constraints and corporate rules based on managerial decisions. Examples of such constraints are that “pregnant employees are not allowed to operate X-ray equipment” or that “machinery used for handling food may not be used for handling non-food materials”. Some of such rules may also be stored in the constraints of the processes themselves. However, in order to avoid redundancy and to ease the maintenance of such constraints, they should be stored separately. For this purpose, we propose an **Enterprise Rules and Constraints Ontology**. In the following, we give preliminary *Competency Questions* for this ontology:

1. Which {states | sequence of activities} {must not occur | should be avoided | should be preferred [over this {state | sequence of activities}]?}
2. What is the reason for any such constraint?
3. Is the reason for the existence of a particular constraint a legal one, a regulatory one, or a managerial decision?

3.3.7 *Strategy*

When composing business processes or monitoring their execution, we also need access to the corporate strategy, for only this will allow us comparing choices. In the following, we give preliminary *Competency Questions* for an **Enterprise Strategy Upper Ontology**:

1. Which are the business goals of the organization?
2. Which are the sub-goals of a particular goal of the organization?
3. Does a particular activity contribute to or conflict with a particular goal or subgoal?

Elements of this ontology will be e.g. the concepts *Goal*, *SubGoal* and the relations *contributesTo* (*Activity*, *Goal*) and *conflictsWith* (*Activity*, *Goal*).

3.3.8 *Domain Ontologies*

It will usually be necessary to complement the ontologies described by specific domain ontologies, which will refine and extend the standard ontologies. For example, the telecommunications domain will require a very detailed ontology of industry-specific concepts, e.g. “router”, “leased line”, “modem” as resources etc. Quite naturally, defining such domain ontologies is outside the scope of this paper.

4 **Evaluation and Discussion**

In the previous section, we have identified the relevant spheres that must be taken into account when building an ontology framework for SBPM. For this, we took the widely accepted ARIS methodology and framework [Sche'98] as a starting point and proposed suitable ontologies.

Since ontologies are agreements between communities and thus include subjective judgments, evaluating them in a meaningful way is difficult and in fact an open research field; for an introduction, see e.g. [FBGL'98] and [GRUB'95]. A practically relevant evaluation of our approach is possible only by coding the respective ontologies and using them in a prototypical SBPM application. Our conceptual framework is based on (1) the ARIS approach as a popular methodology and widely used tooling environment and (2) the process notion of the Web Service Modeling Ontology (WSMO) [RLKB'05] as one leading framework for Semantic Web Services. We take this as two important aspects of a successful representational framework for Semantic Business Process Management, since it will contribute to bridging the respective two research communities.

Two additional aspects require a bit of discussion. First of all, in ERP-centric environments, many facts described by these ontologies will be readily available in the conceptual model of the ERP system. For example, resources and organizational constraints are already stored in the relational databases of the ERP system. This is rather a benefit than a problem, since it will allow us using this large amount of structured data at little cost; it just requires ontology mappings between the various upper level ontologies described in this paper and the database structure of the ERP system.

A second aspect of SBPM that is still open is whether the cost of formalizing the domain of discourse is justified by the business gain in monetary terms. A substantiation of this claim is still lacking from the proponents of SBPM as such. However, in ERP-centric environments, the large amount of structured and semi-structured information in the transactional and customizing data in combination with the vendor-specific process models lets us believe that, at least within the scope of ERP systems, SBPM can be made a reality at reasonable cost. It is also noteworthy that e.g. the ARIS toolset already provides standardized terminology and conceptual models that may be reused.

5 Conclusion

In this paper, we have discussed the representational requirements of Semantic Business Process Management. We proposed a set of ontologies and formalisms and defined the scope of these ontologies by giving competency questions, which is a common approach in the ontology engineering process. In particular, we have shown that a control flow-centric, procedural representation of business processes is often an over-specification of the actual process and should thus not be made the core of an ontological framework for Semantic Business Process

Management. We are currently in the process of a detailed domain capture for the proposed ontologies. Next, we will formalize the ontologies in WSMML and validate their competency by trying to answer the competency questions using our ontologies.

Acknowledgements: We would like to thank Roxana Belecheanu, John Domingue, Agata Filipowska, Frank Leymann, Barry Norton, Carlos Pedrinaci, and all colleagues from the SUPER research project for valuable discussions. The work presented in this paper was partly funded by the European Commission under the projects DIP (FP6-507483), SUPER (FP6-026850), and MUSING (FP6-027097), and by the Trans IT Entwicklungs- und Transfercenter at the University of Innsbruck.

References

- [AgGL1998] Agrawal, R. et al.: Mining process models from workflow logs. Intl. Conf. on Extending Database Technology (EDBT'98). Valencia, Spain, March 3-8, 1998.
- [ACGD2003] Andrews, Tony et al.: Business Process Execution Language for Web Services Version 1.1. <http://www.siebel.com/bpel>, retrieved Nov 30, 2005.
- [Arki2002] Arkin, Assaf: Business Process Modeling Language. <http://www.bpml.org/downloads/BPML1.0.zip>, retrieved Nov 30, 2005.
- [AAFJ2002] Arkin, Assaf et al.: Web Service Choreography Interface (WSCI) 1.0. W3C Note 8 August 2002. <http://www.w3.org/TR/wsci/>, retrieved Nov 30, 2005.
- [BPMI2004a] BPMI.org: Business Process Modeling Notation (BPMN) Version 1.0. <http://www.bpml.org/downloads/BPMN-V1.0.pdf>, retrieved Nov 30, 2005.
- [BCHL2003] Bunting, Doug et al.: Web Services Composite Application Framework (WS-CAF). <http://developers.sun.com/techtopics/webservices/wscaf/>, retrieved Nov 30, 2005.
- [dBLK2005] de Bruijn, Jos et al.: D16.1v0.21 The Web Service Modeling Language WSMML. WSMML Final Draft. <http://www.wsmo.org/TR/d16/d16.1/v0.21/>, retrieved November 7, 2005.
- [Diet2006] Dietz, Jan L. G.: Enterprise Ontology. Springer, Berlin / Heidelberg 2006.
- [FeDo2005] Feier, Cristina; Domingue, John: D3.1v0.12 WSMO Primer, WSMO Working Draft April 1, 2005. <http://www.wsmo.org/TR/d3/d3.1/v0.2/20050401/>, retrieved Nov 30, 2005.
- [FeBu2002] Fensel, Dieter; Bussler, Chris: The Web Service Modeling Framework WSMF. In: Electronic Commerce Research and Applications 1 (2002) 2, pp. 113-137.
- [FBGL1998] Fox, Marc S. et al.: An Organisation Ontology for Enterprise Modeling. In: M. Prietula; K. Carley; L. Gasser (Hrsg.): Simulating Organizations: Computational Models of Institutions and Groups. AAAI/MIT Press, Menlo Park CA 1998, pp. 131-152.
- [FoGr1998] Fox, Marc S.; Gruninger, Michael: Enterprise Modeling. In: AI Magazine (1998) Fall 1998, pp. 109-121.
- [GrGM2005] Greco, Gianluigi et al.: Mining and Reasoning on Workflows. In: IEEE Transactions on Knowledge and Data Engineering 17 (2005) 4, pp. 519-534.
- [GRUB1995] Gruber, Thomas R.: Toward principles for the design of ontologies used for knowledge sharing. In: International Journal of Human-Computer Studies 43 (1995), pp. 907-928.
- [GrAF2000] Gruninger, Michael et al.: Ontologies to Support Process Integration in Enterprise Engineering. In: Computational & Mathematical Organization Theory 6 (2000) 4, pp. 381-394.

- [HeLD2005] Hepp, Martin et al.: Semantic Business Process Management: A Vision Towards Using Semantic Web Services for Business Process Management. IEEE International Conference on e-Business Engineering (ICEBE 2005). Beijing, China, October 18-20, 2005, pp. 535-540.
- [Ibis2005] IBIS Prof. Thome AG: RBE Plus. <http://www.rbe-online.de>, retrieved Nov 30, 2005.
- [JaAR2006] Jansen-Vullers, M.H. et al.: Mining configurable enterprise information systems. In: Data & Knowledge Engineering (2006) (forthcoming).
- [KeNS1991] Keller, G. et al.: Semantische Prozessmodellierung auf der Grundlage "Ereignisgesteuerter Prozessketten (EPK)". Veröffentlichung des Instituts für Wirtschaftsinformatik, Paper 089, 1991.
- [Leym1996] Leymann, Frank: The workflow-based application paradigm. Workshop on Workflow Management - State of the Art. Münster, Germany, April 10, 1996.
- [Leym2006] Leymann, Frank: Super: Overall Architectural Direction (2006).
- [LeRo1997] Leymann, Frank; Roller, Dieter: Workflow based applications. In: IBM Systems Journal 36 (1997) 1, pp. 102-123.
- [LeRo2000] Leymann, Frank; Roller, Dieter: Production Workflow - Concepts and Techniques. PTR Prentice Hall, 2000.
- [MaCH2003] Malone, Thomas W. et al.: Organizing Business Knowledge: The MIT Process Handbook. The MIT Press, Cambridge, MA, USA; London, UK 2003.
- [MBDH2004] Martin, David et al.: OWL-S 1.0 Release. <http://www.daml.org/services/owl-s/1.0/>, retrieved Nov 30, 2005.
- [BPMI2004b] o.V.: Business Process Management Initiative. <http://www.bpmi.org>, retrieved Nov 30, 2005.
- [Ober2006] Oberle, Daniel: Semantic Management of Middleware. Springer, New York 2006.
- [RIRG2006] Recker, Jan et al.: How Good Is BPMN Really? Insights From Theory And Practice. 14th European Conference on Information System (ECIS 2006). Gothenburg (Sweden)2006, pp. 1-12.
- [RLKB2005] Roman, Dumitru et al.: D2v1.2 Web Service Modeling Ontology (WSMO). WSMO Final Draft April 13, 2005. <http://www.wsmo.org/TR/d2/v1.2/>, retrieved Nov 30, 2005.
- [Sche1998] Scheer, A.-W.: ARIS - Vom Geschäftsprozess zum Anwendungssystem. 3rd. Aufl., Springer, Berlin etc. 1998.
- [SmiFi2002] Smith, Howard; Fingar, Peter: Business Process Management. The Third Wave. Meghan-Kiffer Press, Tampa, FL, USA 2002.
- [That2001] Thatte, Satish: XLANG. Web Services for Business Process Design. http://www.gotdotnet.com/team/xml_wsspecs/xlang-c/default.htm, retrieved 2005, Nov 30.
- [ThHu1996] Thome, Rainer; Hufgard, Andreas: Continuous System Engineering. Vogel Verlag, Würzburg 1996.
- [UsGr1996] Uschold, Mike; Grüninger, Michael: Ontologies: Principles, Methods, and Applications. In: Knowledge Engineering Review 11 (1996) 2, pp. 93-155.
- [AaPe2006] van der Aalst, W.M.P.; Pesic, M.: Specifying, Discovering, and Monitoring Service Flows: Making Web Services Process-Aware. BPM Center Technical Report, No. BPM-06-09, 2006.
- [AaDH2003] van der Aalst, W.M.P. et al.: Workflow mining: A survey of issues and approaches. In: Data & Knowledge Engineering 47 (2003), pp. 237-267.
- [W3C2004] W3C: OWL Web Ontology Language Guide. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>, retrieved Nov 30, 2005.
- [Webe1997] Weber, R.: Ontological Foundations of Information Systems. Coopers & Lybrand and the Accounting Association of Australia and New Zealand, Melbourne, Australia 1997.

Automated Selection of Configurable Web Services

Steffen Lamparter, Anupriya Ankolekar

Institute AIFB
Universität Karlsruhe (TH)
Germany
{sla,aan}@aifb.uni-karlsruhe.de

Abstract

To bring service-oriented architectures to their full potential, automatic service discovery and selection mechanisms are required. In this paper, a service selection component is presented that supports offers providing multiple configurations of a service. The selection algorithm ranks the offered services and their configurations according to the requester's preferences and thus facilitates personalized selection strategies. In addition, the approach leverages existing Web standards to provide a maximal degree of interoperability between service providers and their customers leading to significant efficiency gains. The approach is implemented prototypically and the performance is evaluated by means of a simulation.

1 Introduction

Service-oriented architectures (SOA) as a paradigm where applications are built by composing loosely coupled, highly interoperable, reusable services have become increasingly popular in recent years. This flexible style of implementing applications promises cost savings in software development and faster adaption of business processes to changing environments. Realizing such a flexibility requires automatic as well as efficient Web service discovery and selection mechanisms. These mechanisms depend heavily on the way services are described. Approaches that are solely based on WSDL-descriptions are inadequate since XML and XMLSchema do not provide sufficient expressivity to create and relate rich datatypes [MM03]. Therefore, these mechanisms are not amenable to a high degree of automation. In order to tackle this problem several mechanisms have been proposed that depend on much richer service descriptions. Most of them describe services using formal ontologies (e.g. WSMO [KLP⁺04], OWL-S [SPAS03], WSDL-S [AFJ⁺05]). However, none of these approaches address the fact that Web services are highly configurable products that can be offered by multiple parties (providers) with different attributes and under different conditions. This enables a high degree of product differentiation allowing providers to customize their services according to the specific needs of their customers. Clearly, this improves the utility of the transaction for both participants. Consider, for example, a route planning web service, which offers the service of computing a road route between

two locations. Various configurations of the service may take into account the current traffic situation or weather situation when computing the route, or the service may be configured to compute the shortest or quickest route, one that avoids small roads and so on. Naturally, each configuration may have a different price attached. Decision making in markets with such complex services generally requires that both seller pricing functions as well as buyer scoring (preference) functions be taken into account.

In [LAO⁺06] a general policy framework is described that can be used to express pricing and scoring functions. The framework relies on existing or emerging internet standards and thus provides a high degree of interoperability. In this paper, we extend the work in several directions. First, it is shown how the policy framework can be applied to describe configurable Web service offers and requests. Second, the framework is augmented with an abstract selection model, which is independent from specific formalisms and implementations. Third, we present an approach to rank offers in scenarios where both requests and offers are configurable. In this context, we also show how preferences can be expressed directly within a query. The selection algorithm in the previous work deals only either with configurable requests or configurable offers, which are both stored in the knowledge base. Finally, we present a concrete implementation and initial performance evaluations of the system.

The paper is structured as follows. In Section 2 an abstract selection model is presented. In Section 3 this model is formalized with the standardized Web languages OWL, SWRL and SPARQL in order to provide interoperability between Web service requesters and providers. After presenting an implementation and evaluation in Section 4, we discuss related work in Section 5 and conclude the work with a short outlook in Section 6.

2 Abstract Selection Model

First, a *selection* is defined as a decision for the *best* available alternative, i.e. the Web service that is most appropriate to fulfill a certain task. In general, decisions require a choice and criteria by which different choices are judged. Hence, a selection can be regarded as an optimization problem with a certain objective function resulting in an understanding of “better” and “worse”. In this section, we first introduce an abstract notation of the fundamental concepts in the domain and then present the optimization problem that has to be solved in order to derive a preference structure over the service offers.

Definition 1 (Web Service Configuration) *Let $\mathcal{S} = \{s_1, \dots, s_{|\mathcal{S}|}\}$ be the set of all Web services. Then, a Web service $s_i \in \mathcal{S}$ is described by the tuple $s_i = (\mathcal{I}_i, \mathcal{O}_i, C_i)$, where \mathcal{I}_i represent the set of input messages that are required by the service s_i and \mathcal{O}_i the set of output messages that are returned by the service s_i , respectively. Furthermore, a Web service is characterized by a set of feasible configurations C_i with $C_i \subseteq C$, where C is the cartesian product of the discrete*

service attributes $\mathcal{A}_1, \dots, \mathcal{A}_n$, i.e. $\mathcal{C} = \prod_{r \in \{1, \dots, n\}} A_r$. In this context, each $c \in \mathcal{C}$ is a vector $c = (a_{1k_1}, \dots, a_{jk_j}, \dots, a_{nk_n})$, where a_{jk_j} represents the k^{th} value of attribute j .

Recall the route planing example mentioned in section 1. Here the set \mathcal{I}_i represents locations and \mathcal{O}_i the route between these locations. Moreover, we have three attributes: (i) The route type which refers either to the quickest or cheapest route; (ii) one attribute that indicates whether traffic information is taken into account; (iii) and one attribute that allows to include or exclude weather information. These attributes are used to configure the service according to the requesters preferences. For instance, one specific route service configuration is a routing functionality that returns the quickest route including traffic and weather information. Note that according to Definition 1, service functionality is solely defined by inputs as well as outputs and attributes comprise only discrete and static non-functional service properties.

In order to enable Web service transactions between providers and customers, Web service offers and requests have to be specified. Their main purpose is to attach prices as well as preferences to the service configurations. In the following, we formally define how offers and requests can be specified for configurable services. More general approach (e.g. including discount rules) see [BK05].

Definition 2 (Web Service Offer) Assume a set of providers \mathcal{P} as well as a set of Web service offers Θ . Then, $p \in \mathcal{P}$ is an arbitrary provider offering service $o_p \in \Theta$. A Web service offer is characterized by a vector $o_p = (s_p, G_p)$, where $s_p \in \mathcal{S}$ represents the provided service and $G_p : \mathcal{C}_p \rightarrow \mathbb{R}$ the pricing function that assigns a certain price to each configuration $c \in \mathcal{C}_p$ of the service s_p . We assume that the pricing function is described by an additive function, where g_{pj} represents the pricing function of provider p for attribute j . w_j^g can be used to adjust the influence of the different attributes on the price.

$$G_p(c) = G_p(a_{1k_1}, \dots, a_{jk_j}, \dots, a_{nk_n}) = \sum_{j=1}^n w_j^g g_{pj}(a_{jk_j}) \text{ with } \sum_{j=1}^n w_j^g = 1 \quad (1)$$

That means an offer assigns an additive pricing function to a Web service description. The pricing function maps the configurations contained in the service description to a certain price. This approach allows for encoding pricing information in an efficient way. This is required since adding price markup to each configuration would exhibit combinatorial features [BK05]. By means of Function 1, the provider has to define only n pricing functions instead of adding $\prod_{j=1, \dots, n} |A_j|$ price markups. For instance, in the simple route planning example, pricing information can be expressed via six attribute value/price-tuples instead of adding price markups to eight configurations. Of course, for more complex services the efficiency gains are much higher. In the following, a Web service request is defined analogously.

Definition 3 (Web Service Request) Given a set of customers \mathcal{B} , a Web service request is defined as a vector $r_b = (s_b, F_b)$, where $b \in \mathcal{B}$ is the issuer of the request and $s_b \in \mathcal{S}$ the requested service. The preferences of the customer are defined by a standard additive scoring function $F_b : C_b \rightarrow \mathbb{R}$ that assigns a certain score to each requested configuration $c \in C_b$. A configuration which is not requested leads to a score of zero. f_{bj} is the scoring function of requester b for attribute j . Attribute values that are forbidden a score of minus infinity is assigned, i.e. $f_{bj} = -\infty$. w_j^f is the relative importance of attribute j .

$$F_b(c) = \begin{cases} \sum_{j=1}^n w_j^f f_{bj}(a_{jk_j}) & \text{if } c \in C_b, \\ 0 & \text{otherwise.} \end{cases} \quad \text{with } \sum_{j=1}^n w_j^f = 1 \quad (2)$$

For the additive scoring function we have to assume mutual preferential independency between the attributes [KR76]. However, as discussed in [LAO⁺06], from a technical perspective we can express preferences over dependent attributes using a higher dimensional function. Of course, specification of preferences is much harder for the requesters in this case.

In order to find the services that are suitable for a certain task, the inputs and outputs of the requested service s_b are compared with those of the offered services. This is done according to the following matching rule.

Definition 4 (Functional Match) Let r_b be a request for service s_b with the corresponding inputs \mathcal{I}_b and outputs \mathcal{O}_b . Analogously, o_p is a Web service offer containing the service s_p with the inputs \mathcal{I}_p and outputs \mathcal{O}_p . Then, r_b functionally matches o_p iff $\mathcal{I}_p \subseteq \mathcal{I}_b$ and $\mathcal{O}_b \subseteq \mathcal{O}_p$. A functional match is indicated by the notation $r_b =_f o_p$. The set $\Theta^r \subseteq \Theta$ contains the service offers that functionally match a given request r_b , i.e. $\Theta^r = \{o_b \in \Theta | r_b =_f o_b\}$

The intuition behind this approach is that only those Web service offers match that provide at least the requested information while requiring at most the input specified in the request.

Finding the optimal service configuration involves two steps: selecting the best configuration for each provider and choosing the best provider based on the optimal configurations. First, we formulate the optimization problem that allows determining the best configuration $c \in C_p$ offered by a provider $p \in \mathcal{P}$.

The objective function that has to be optimized represents the difference between the score assigned to a certain configuration by the requester and the price of this configuration. The corresponding optimization problem can be formulated as follows:

$$\max_{c_i \in C_p} F_b(c_i) - G_p(c_i) \quad (3)$$

A naive approach to solve this problem is to iterate over all configurations, calculate the individual utilities and rank the configurations accordingly. The overall complexity of such an

algorithm is given by $|\mathcal{O}| \sum_j |\mathcal{A}_j|$, where $|\mathcal{O}|$ is the number of offers and $\sum_j |\mathcal{A}_j|$ the number of possible configurations. Since the number of configurations is exponential in the number of attributes this approach might be inefficient for very complex services. However, in a first step such a naive approach seems to be appropriate, since current Web service descriptions usually consider only relatively few non-functional properties.

Having determined the optimal configuration c_o^* and utility u_o^* for each offer o , we derive a preference structure as follows:

Definition 5 (Preference Structure) *A preference structure on \mathcal{O} is defined by the complete, transitive, and reflexive relation \succeq , which defines a weak preference order over the alternatives as follows: $\forall o_1, o_2 \in \mathcal{O}: o_1 \succeq o_2 \Leftrightarrow o_1$ is preferred to o_2 . Consequently, a preference structure can be constructed from the requester's utility as follows: $o_1 \succeq o_2 \Leftrightarrow u_{o_1}^* \geq u_{o_2}^*$.*

Therefore, a reasonable selection rule would be to select $o^* \in \mathcal{O}$ iff $\forall o \in \mathcal{O}: o^* \succeq o$. Of course, depending on the application also other matching approaches might be required. For example, one might want to select the ten best services to invite them for further negotiations.

Based on the abstract selection model introduced above, the next section deals with implementing this model in an open and heterogenous environment using existing standards and tools.

3 Ontology-based Representation

In this section, we show how the abstract selection model introduced above can be implemented with a standardized logical formalism providing a common understanding between providers and customers. We realize this by means of ontologies, which became an important technology for knowledge sharing in distributed, heterogeneous environments, particularly in the context of the *Semantic Web*.¹

3.1 Ontology formalism

An ontology is a set of logical axioms that formally define a shared vocabulary [Gru93]. By committing to a common ontology, software agents can make assertions or ask queries that are understood by the other agents.

In order to guarantee that these formal definitions are understood by other parties (e.g. in the web), the underlying logic has to be standardized. The Web Ontology Language (OWL) standardized by the World Wide Web Consortium (W3C) is a first effort in this direction [W3C04]. OWL-DL is a decidable fragment of OWL and is based on a family of knowledge representation formalisms called *Description Logics (DL)* [BCM⁺03]. Consequently, our notion of an ontology is a DL knowledge base expressed via RDF/XML syntax to ensure compatibility with

¹<http://www.w3.org/2001/sw/>

DL Syntax	Semantics
\top	Δ^I
\perp	\emptyset
$C \sqcap D$	$C^I \cap D^I$
$C \sqcup D$	$C^I \cup D^I$
$\forall R.C$	$\{a \in \Delta^I \mid \forall b.(a, b) \in R^I \rightarrow b \in C^I\}$
$\exists R.C$	$\{a \in \Delta^I \mid \exists b.(a, b) \in R^I \wedge b \in C^I\}$
$C \sqsubseteq D$	$C^I \subseteq D^I$

Table 1: Selected DL constructs and their model theoretic semantics (for a full list see [BCM⁺03]).

existing World Wide Web languages. The meaning of the modeling constructs provided by OWL-DL like concepts, relations, datatypes, individuals and data values is formally defined via a model theoretic semantics. The mapping for certain common DL axioms is shown in table 1. The meaning of an axiom defines certain constraints on the model. For example, we can define that the concept *Book* is a subconcept of *Product* (i.e. $Book \sqsubseteq Product$). In this case, the interpretation of *Book* has to be a subset of the interpretation of *Product*, i.e. the set of objects that are books is a subset of the set of objects that are products ($Book^I \subseteq Product^I$). By means of the interpretation I the model introduced in section 2 can be expressed using Description Logics in a straightforward way as shown in the next sections.

However, we require additional modeling primitives not provided by OWL-DL. For example, modeling triangle relations between concepts is required. In contrast to OWL, rule languages can be used to express such triangle relation. The Semantic Web Rule Language (SWRL) [HPS04, HPSB⁺04] allows us to combine rule approaches with OWL. Since reasoning with knowledge bases that contain arbitrary SWRL expression usually becomes undecidable [HPS04], we restrict ourself to *DL-safe* rules [MSS05]. DL-safe rules keep the reasoning decidable by placing constraints on the format of the rule, namely each variable occurring in the rule must also occur in a non-DL-atom in the body of the rule. This means the identity of all objects referred to in the rule has to be known explicitly. Since we deal only with known instances in our application, this is no restriction to our approach. To query and reason over a knowledge base containing OWL-DL as well as DL-safe SWRL axioms we use the KAON2 inference engine².

For the reader's convenience we define DL axioms informally via UML class diagrams, where UML classes correspond to OWL concepts, UML associations to object properties, UML inheritance to subconcept-relations and UML attributes to OWL datatype properties [BVEL04]. For representing rules we rely on the standard rule syntax as done in [HPSB⁺04, MSS05].

²available at <http://kaon2.semanticweb.org/>

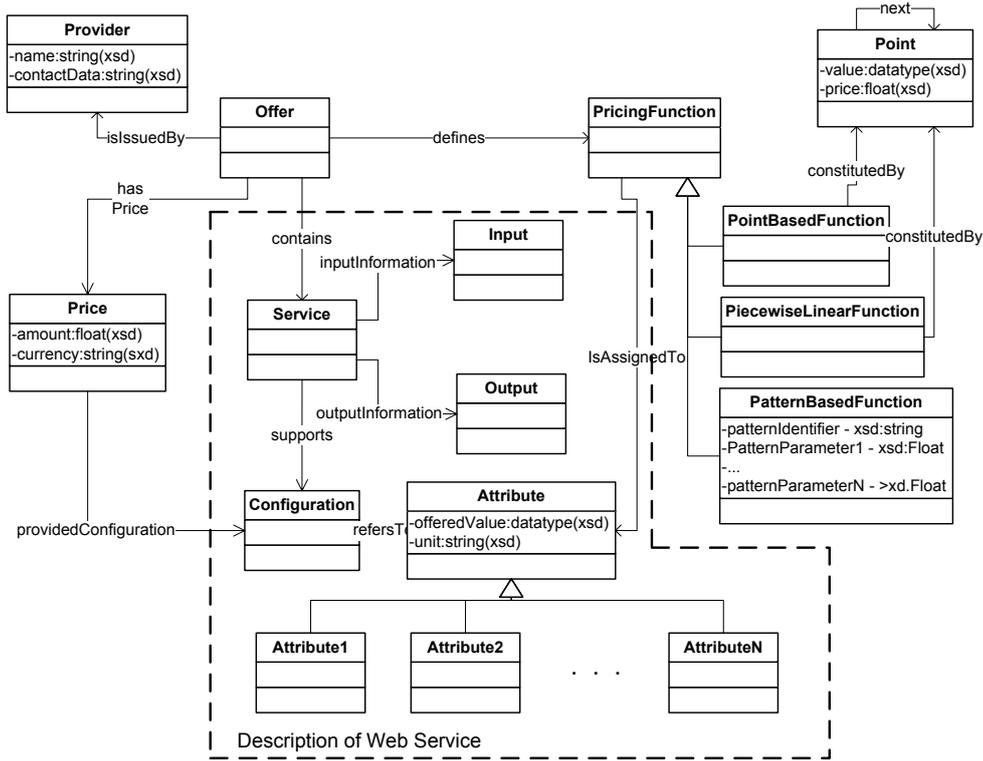


Figure 1: Web Service Offer Ontology

3.2 Modeling Web Services

In this section, it is shown how the abstract Web service model specified in Definition 1 can be formalized using OWL-DL. Note that some additional assumptions are required to derive a formally sound model.

Figure 1 sketches the modeling approach. For modeling Web services the classes within the dashed box are relevant. We introduce the concepts *Service*, *Input*, *Output*, *Configuration* and *Attribute* referring to the sets \mathcal{S} , \mathcal{I} , \mathcal{O} and \mathcal{A}_1 to \mathcal{A}_n in Definition 1, respectively. A *Service* is characterized by their *Inputs*, *Outputs* and *Configurations* involved. This is reflected by the relations *inputInformation*, *outputInformation*, and *supports*, which are formally defined by the following axiom:

$$\begin{aligned}
 \text{Service} \sqsubseteq & \top \sqcap \exists \text{inputInformation. Input} \sqcap \forall \text{inputInformation. Input} \sqcap \\
 & \exists \text{outputInformation. Output} \sqcap \forall \text{outputInformation. Output} \sqcap \\
 & \exists \text{supports. Configuration}
 \end{aligned} \tag{4}$$

This axiom makes sure that each service has at least an *Input* and *Output*. In addition, each service has to support at least one *Configuration*. Recall in our abstract model a configuration $c_i \in \mathcal{C}$ is a vector $c_i = (a_{11}, \dots, a_{ij}, \dots, a_{nm})$ containing one value a_{ij} for each attribute. Hence, we introduce *Configuration* as those individuals that refer to exactly one individual of each

attribute. Assume the concepts $Attribute1 \sqsubseteq Attribute, \dots, AttributeN \sqsubseteq Attribute$ which refer to the sets \mathcal{A}_1 to \mathcal{A}_n . Then a *Configuration* is defined as follows:

$$Configuration \sqsubseteq \top \sqcap_{=1} refersTo.Attribute1 \sqcap \dots \sqcap_{=1} refersTo.AttributeN \quad (5)$$

Each *Attribute* has to have at least two datatype properties: one property representing the *offered value* as well as one property defining the *unit* a certain attribute is measured.

This means in the route planning scenario a service would be described by the three object properties *start*, *destination* and *result*, where *start* and *destination* are subproperties of *inputInformation* and *result* is a subproperty of *outputInformation*.³ Then a route planning service is defined as follows:

$$RoutePlanning \sqsubseteq Service \sqcap_{=1} start.Location \sqcap_{=1} destination.Location \sqcap \exists result.Route \quad (6)$$

Furthermore, the service is described by three *Attributes*: $Weather \sqsubseteq Attribute$, $Traffic \sqsubseteq Attribute$ and $RouteType \sqsubseteq Attribute$. Each of them could be instantiated by two attribute values, which leads to eight possible configurations that can be provided.

After introducing the primitives for modeling Web services in the next sections we show how offers and requests for a certain Web service can be formalized.

3.3 Offer Specification

According to Definition 2 an offer $o \in \Theta$ is characterized by a provider $p \in \mathcal{P}$ offering a service $s \in \mathcal{S}$ under a certain pricing policy $G_p(c_i)$, where c_i represents an arbitrary service configuration $c_i \in C_p$. For modeling the set of offers Θ we introduce the concept *Offer* which *contains* a *Service* and *is issued by* a *Provider*. These simple facts are visualized in Figure 1. Furthermore, prices are attached to the various Web service configurations by means of a function $G_p(c_i)$. To capture this relations the concept *Price* is introduced that relates an *Offer*, a *Configuration* and the corresponding *amount*. This is required since OWL does not support tertiary relations. The datatype property *amount* is determined by function 1. Such functions are modeled using the concept *Pricing Function*. A thorough discussion how such policies can be modeled using OWL-DL and DL-safe SWRL rules is given in [LAO⁺06]. In the following we focus one of the approaches - called *Point Based Function*.

In case of a *Point Based Function* the pricing function is modeled by specifying sets of points in \mathbb{R}^2 that explicitly map attribute values to prices. This is particularly relevant for nominal attributes. As depicted in Figure 1, *Point Based Functions* are *Pricing Functions* that are *constituted by* a set of *Points*. Thus, the datatype property *value* refers to exactly one attribute

³Note that the examples given in this paper are simplified. To be precise one would have to model *Input* and *Output* as roles that are played by information objects representing location or route information. For a more detailed modeling approach based on the foundational ontology DOLCE see [LAO⁺06, OLG⁺06]

value a_{jk_j} and the datatype property *price* to exactly one price $g_j(a_{jk_j})$ that is assigned to this attribute value. OWL datatypes mainly rely on the non-list XML Schema datatypes. Depending on the attribute, *value* either points to a *xsd:string*, *xsd:integer* or *xsd:float*. A *price* is represented by a *xsd:float*. The reference to the attribute j for a function g_j is defined via the *isAssignedTo*-relation. Having this information the overall price $G_p(c_i)$ a provider attaches to a certain configuration c_i can be calculated using formula 1, which can be formalized using the following DL-safe rule:

$$\begin{aligned}
amount(?p, ?pr) \leftarrow & Offer(?o), hasPrice(?o, ?p), providedConfiguration(?p, ?c), \\
& \bigwedge_{j \in \{1, \dots, n\}} (Attribute_j(?a_j), refersTo(?c, ?a_j), offeredValue(?a_j, ?av_j), \\
& assignedTo(?f_j, ?a_j), constitutedBy(f_j, p_j), value(?p_j, ?v_j), \\
& equal(?v_j, ?av_j), price(?p_j, ?pr_j)), sum(?pr_1, \dots, ?pr_n, ?pr) \quad (7)
\end{aligned}$$

Note that rule 7 is not generic with respect to the number and types of the attributes used. However, a more general rule definition is not possible since this would require allquantification in rule bodies, which is not possible using SWRL. Since the rule is generic with respect to the offers, it is possible to generate the rule once for a certain service type. Hence, we believe that this is no restriction in practice. To improve readability, the conversion of measurement units that might be required is omitted in rule 7. *Point Based Function* thus allow to decrease the number of required price specifications from $\prod_{j=1}^n |\mathcal{A}_j|$ to $\sum_{j=1}^n |\mathcal{A}_j|$. Using other function specifications, like *Pattern-based Functions* or *Piecewise Linear Functions*, further improvements are possible.

Having shown how offers can be specified using OWL-DL and DL-safe SWRL rules, in the next section we focus on formalizing request in a way that allows a customer to derive a set of services ordered according to her preferences.

3.4 Request Specification

The aim of this step is to formalize the requester's goals in a way that facilitates the discovery of offers meeting these goals. According to Definition 3, a request defines the properties of an object that is required by a customer in an abstract way, i.e. without referring to a concrete name or identifier. In databases, *queries* are seen as such "intentional" denotations of objects [LL87]. For expressing queries we rely on the emerging standard SPARQL⁴, which provides a protocol and query language for RDF and OWL-DL ontologies. SPARQL is supported by the reasoner KAON2. The discovery of suitable offers has to be sound and complete, i.e. all relevant offers are returned while no irrelevant offers are contained in the result set. In addition, a ranking of the results according to the requesters preferences should be provided. In the following,

⁴SPARQL, W3C Candidate Recommendation (6 April 2006), available at <http://www.w3.org/TR/rdf-sparql-query/>

we gradually show how a SPARQL-query is formulated that enables expressing a Web Service Request (Definition 3).

In a first step, we determine the set of suitable services, i.e. those services that provide the right functionality. In this context, we mainly implement the functional matching introduced in Definition 4. A functional match is realized if $\mathcal{I}_p \subseteq \mathcal{I}_b$ and $\mathcal{O}_b \subseteq \mathcal{O}_p$. Consequently, in DL this amounts to checking the entailment of concept subsumption, i.e. $Input_p \sqsubseteq Input_b$ and $Output_b \sqsubseteq Output_p$. This is a very common approach already used for service matching in [PKPS02, NSDM03, LH03]. However, note that this is very restrictive and for some applications better solutions might exist (cf. [GMP04]). The following SPARQL-query formulates the matching rule for our route planning service example.

```
PREFIX wsm: <http://ontoware.org/emo/1.1/>
SELECT ?offer
WHERE {
    ?offer contains ?service . ?start rdf:type wsm:Location .
    ?dest rdf:type wsm:Location . ?result rdf:type wsm:Route .
    ?service wsm:start ?start ; wsm:destination ?dest ; wsm:result ?result . }
```

In the second step, we add constraints to this query in order to reduce the number of returned matches. For each attribute \mathcal{A}_j the subset $\bar{\mathcal{A}}_j = \{a_{jk} \in \mathcal{A}_j | f_{bj}(a_{jk}) = -\infty\}$ is determined for which the utility function of a customer is zero. Based on the set $\bar{\mathcal{A}}_j$ SPARQL-filter conditions are automatically added to the query. For example, assume a requester lives in a very busy quarter of London and he thus mandatorily requires a route planner with traffic information. In this case we would add the following filter condition:

```
SELECT ?offer, ?configuration
WHERE { ...
    ?configuration wsm:refersTo ?traffic .
    ?traffic rdf:type wsm:Traffic ; wsm:offeredValue ?trafficValue .
    FILTER ( ?trafficValue = "yes" ) .
}...
```

A clause in the filter condition is added for all $a_{jk} \in \bar{\mathcal{A}}_j$ and for all attributes j . Note that from a functional point of view the filter conditions are not required since configurations, which contain at least one attribute that is valued by $-\infty$, are ranked very low and thus are neglected in the selection process. However, by introducing the filter conditions the number of possible configurations that have to be ranked is reduced. This might increase the ranking performance. Finally, to facilitate the selection of a service with the corresponding configuration, we order the set of matches according to the preferences of the user. This means we have to determine which of the configurations $c \in C_p$ offered by a provider $p \in \mathcal{P}$ is most advantageous according to the requester's preferences $F_b(c)$. Technically there are two possibilities to realize this: First, the query above is issued and the results are sent to the client, where a preference-based selection is executed locally. Although this approach avoids revealing the customer's preferences, it may

lead to a high communication overhead, where in the worst case the entire repository has to be transferred to the requester. We thus opt for the second alternative. Here preferences are included as part of the request and the ranking is done by the server. This allows deriving only the best matches from a high number of suitable services in the repository. Methods for adding rich preferences to a query language are well-known in literature [Kie02, LL87, AW00]. Usually such queries are called *preference queries* and are implemented by database built-ins as well as SQL syntax extensions. The KAON2 system enables preference queries by allowing built-in predicates in the SPARQL query. Therefore, the SPARQL syntax is extended by the “EVALUATE” keyword. Again we exemplify the approach using a *Point Based Function* definition. But this time we encode the function as a *String* in the query rather than adding it to the knowledge base (as done for offers in Section 3.3). In order to realize this approach the predicate *pbF* is introduced. *pbF* takes a *String* representation of the tuples representing the *Point Based Function* and an attribute value. The predicate evaluates the *Point Based Function* and returns the score of the attribute value. The query below illustrates this approach using the three attributes of our route planning example. Note that for simplicity reasons scaling as well as weighting issues are omitted.

```
SELECT ?offer, ?configuration, ?u
WHERE { ...
  ?offer wsm:hasPrice ?price .
  ?price wsm:providedConfiguration ?configuration .
  ?configuration wsm:refersTo ?traffic ; ?weather ; ?routeType .
  ?traffic rdf:type wsm:Traffic ; wsm:offeredValue ?trafficValue .
  EVALUATE ?val1 := pbF("(yes,1),(no,0.3)",?trafficValue).
  ?weather rdf:type wsm:Weather ; wsm:offeredValue ?weatherValue .
  EVALUATE ?val1 := pbF("(yes,0.8),(no,0.6)",?wheaterValue).
  ?routeType rdf:type wsm:RouteType ; wsm:offeredValue ?routeTypeValue .
  EVALUATE ?val1 := pbF("(quickest,0.5),(cheapest,0.5)",?routeTypeValue).
  EVALUATE ?val := sum(?val1,?val2,?val3) .
  EVALUATE ?u := sub(?val,?price) }
ORDER BY DESC(?u)
```

The result of this query is a set of matches ordered according to the preference structure introduced in Definition 5. Thus, the corresponding selection rules can be applied.

In the next section, we present a concrete implementation of our selection algorithm. In addition, we discuss the performance of the algorithm.

4 Implementation

The algorithm presented in this paper is implemented within a larger framework consisting of two components: A server component provides a repository for Web service offers. The repository is a DL knowledge base that can be queried using the KAON2 reasoner. KAON2 is

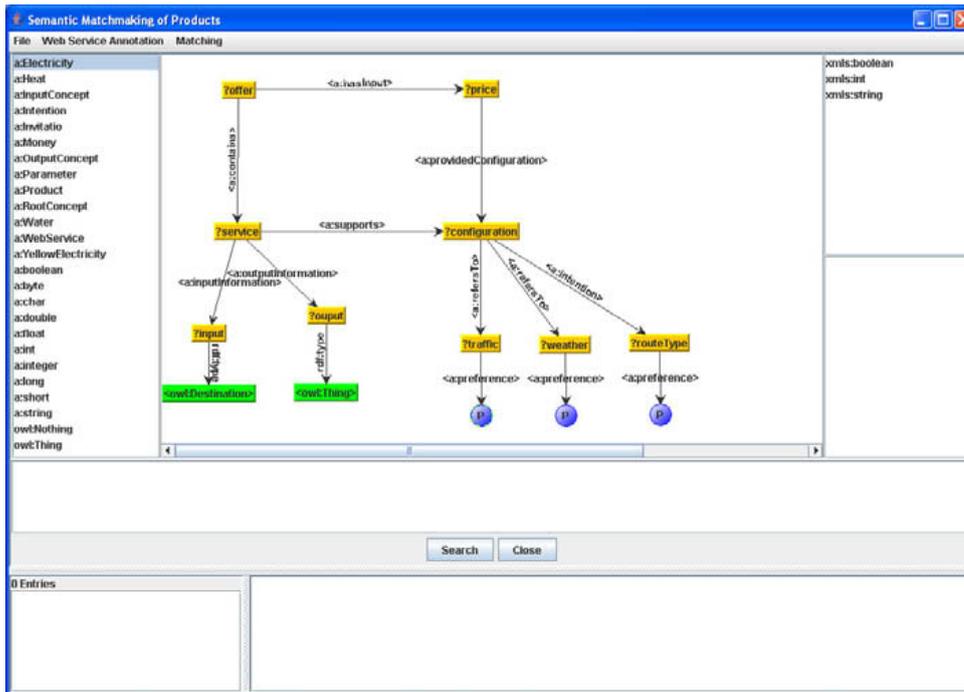


Figure 2: Generation of SPARQL-query.

chosen because it supports the logical fragment required for our offer and request descriptions, while being optimized for query answering [MS06], which is the main focus of the repository. In addition, there are components that transform WSDL and HTML forms to ontology-based descriptions and a Web crawler, which searches the Web for available service descriptions. The second component is a client that facilitates the specification of Web service offers and requests. Since the terminology used by participants might be different, mapping between ontologies can be specified using the formalism presented in [HM05]. Generally, the framework supports more expressive service descriptions than we use in this paper. For example, the service description could include behavioral aspects as presented by Agarwal and Studer [AS06].

Figure 2 shows the request generator which is part of the client tool. It can be used to graphically compose a query. Preferences can be added to all attributes in the query (visualized by a blue circle with a 'P'). By pressing this symbol a window pops up that allows for expressing *Point Based Functions*, *Piecewise Linear Functions* and *Pattern Based Functions*. Once the search button is pressed, the query is formalized using SPARQL and sent to the server. Ranked results are shown and can be browsed using the windows below the search button.

In order to evaluate the performance of the algorithm we conducted a simulation. In this context, offers and requests are randomly generated using a uniform distribution. The set of offers is stored in the knowledge base and the requests are used to generate the queries. The time between sending the query and receiving the result is measured. In order to avoid possible network delays the simulation is done on a single machine. Several simulation runs are conducted with varying

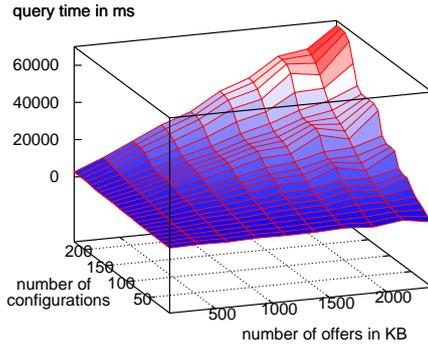


Figure 3: Query time.

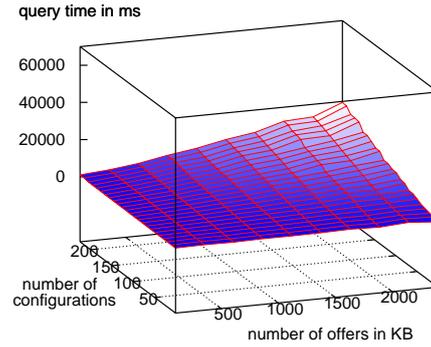


Figure 4: Query time using caching.

number of offers in the knowledge base and varying service complexity. For each setting the average query time is determined based on ten simulation runs. We analyzed the worst case scenario, where all offers functionally match the request, no hard constraints are defined, and all offers provide all configurations. We evaluated two versions of the algorithm:

Algorithm 1: In the first algorithm we simply query the knowledge base and calculate the price of an offer using Rule 7. This has to be done for every relevant configuration, i.e. $|\Theta||C|$ configurations in the worst case.

Algorithm 2: To avoid unnecessary repetition of the same evaluations we introduce *caching* of prices in algorithm 2. After calculating the price for a certain configuration according to the pricing policies in the offer, we store this price as an instance in the knowledge base which avoids additional evaluations for further requests. However, this approach increases the size of the knowledge base by additional $5|\Theta||C| - 4|\Theta|$ axioms compared to Algorithm 1. In a first step, we do not limit the cache size. However, in case of limited storage/memory capacity an adequate caching replacement strategy has to be introduced.

Figure 3 and 4 show the interdependency between the number of offers, the number of configurations in an offer, and the query time for algorithm 1 and 2, respectively. For scenarios with a low number of orders (< 500) or a low number of configurations (< 100) query answering is realized under five second. However, query answering slows down to 70 seconds with 2500 orders and 250 configurations in the knowledge base, which is mainly due to the high number of price calculations. Therefore, the caching of prices used in Algorithm 2 speeds up service selection considerably. However, one should be aware that Algorithm 2 is significantly more resource demanding.

Generally, the evaluation shows that our selection algorithm might be applicable in scenarios where either the number of offers or the number of configurations per offer is moderate or where the selection performance is not crucial. To improve the performance introducing an instance for each configuration should be avoided and the optimization should be done directly based on the pricing functions, possibly with a built-in predicate implementing a simplex algorithm.

Moreover, we could exploit the additive structure of pricing and scoring functions, perform the maximization per attribute and aggregate the maximal values for all attributes. Thereby, the complexity of the problem can be considerably reduced.

5 Related Work

First approaches addressing the configurability of services are policy languages like WS-Policy, EPAL and WSPL. They allow to define which configuration are supported by a provider or desired by a requestor, but no prices or preferences can be attached. WS-Agreement [Gri05] extends WS-Policy in this direction and provides the means for attaching prices and preferences to configurations. However, all these specifications are based purely on XML and thus lack formal semantics. KAOs [UBJ04] and REI [Kag04] are ontology-based approach for expressing policies. However, they also evaluate either to true or false and thus provide no ranking of the alternatives.

In literature we can identify three major branches of work that strive for ranking of suitable services. First, there are logic-based approaches (e.g., [PKPS02], [GMP04], [NSDM03]) that allow for different degrees of matches based on partial or incomplete matches. However, such rankings are typically rather coarse and one can argue that pure logical matchmaking without value reasoning is not sufficient [SRT05, KFKS05]. Second, there are matchmaking approaches purely based on information retrieval techniques like [BK06]. Here rankings are calculated by defining similarity measures for service properties. Klusch and colleagues [KFKS05] extend logic-based matchmaking with syntactic measures. This is similar to our approach, since we use a logic-based approach for matchmaking of service functionality. In addition, similarity measures can be seen as special preference functions. However, our selection approach is not limited to similarity-based preferences. It is also possible to specify the valuation of certain alternatives explicitly. This is also possible in the system presented by Balke and Wagner [BW03]. They use SQL-based preference queries introduced in [Kie02]. However, their work is not based on semantic service annotations. A third branch of work relies on logical rule languages to specify preferences (as in our case SWRL). Prominent examples are SweetDeal [GP03] and the work by Oldham and colleagues [OVSH06]. Although these approaches show how preferences can be formally represented, they currently lack a formal selection model as we presented in section 2.

All approaches above consider only the case of selecting a single service. Approaches beyond this simple case are presented in [SNVW06] and [ZBN⁺04]. [SNVW06] presents an multi-attribute combinatorial auction for grid service markets with multiple sellers and buyers bidding on complex service bundles. Zeng and colleagues [ZBN⁺04] present two methods for selecting a service based on several QoS-criteria. In this context, they consider not only one single service, but optimize the utility for an entire composition of services using integer programming.

However, the problem of service selection for multiple service requests is beyond the scope of this paper.

6 Conclusion

The work in this paper addresses a problem that arises with the increasing differentiation of services in electronic markets. This requires mechanisms to describe configurable services and algorithms to discover and select suitable services and configurations. It is shown, how the abstract selection model can be implemented by leveraging standardized web languages, such as OWL-DL, SWRL and SPARQL. This facilitates interoperability in heterogenous and open environments. In a first step, a rather simple optimization algorithm is applied to rank the service offers according to the preferences of a requester. The algorithm is evaluated by means of a simulation. The evaluations indicate that, while being suitable for rather simple service descriptions or small repositories, the system is not yet capable of selecting from a large repository of complex services in run-time. We are currently working on the scalability of the system and plan to add a built-in predicate which efficiently implements the optimization problem. In doing this, iterating over all configurations can be avoided and the number of instances in the repository can be reduced. Moreover, we plan to extend our approach in order to allow specifying preferences also for functional, dynamic and continuous service properties.

Acknowledgments

This work was funded by the German Research Foundation (DFG) in scope of Graduate School Information Management and Market Engineering.

References

- [AFJ⁺05] R. Akkiraju, J. Farrell, J. Miller, M. Nagarajan, M. Schmidt, A. Sheth, and K. Verma. Web Service Semantics - WSDL-S, " A joint UGA-IBM Technical Note, version 1.0,. Technical report, April 2005.
- [AS06] S. Agarwal and R. Studer. Automatic matchmaking of web services. In *Int. Conf. on Web Services (ICWS'06)*, Chicago, USA, 2006. IEEE Computer Society.
- [AW00] R. Agrawal and E. L. Wimmers. A framework for expressing and combining preferences. In *SIGMOD '00: Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 297–306, New York, NY, USA, 2000. ACM Press.

- [BCM⁺03] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory Implementation and Applications*. Cambridge University Press, 2003.
- [BK05] M. Bichler and J. Kalagnanam. Configurable offers and winner determination in multi-attribute auctions. *European Journal of Operational Research*, 160(2):380–394, January 2005.
- [BK06] A. Bernstein and C. Kiefer. Imprecise RDQL: Towards Generic Retrieval in Ontologies Using Similarity Joins. In *21th Annual ACM Symposium on Applied Computing (SAC)*, New York, NY, USA, 2006. ACM Press.
- [BVEL04] S. Brockmans, R. Volz, A. Eberhart, and P. Löffler. Visual modeling of OWL DL ontologies using UML. In S. M. et al., editor, *Proc. of the 3rd International Semantic Web Conference*, pages 198–213, Hiroshima, Japan, November 2004. Springer LNCS.
- [BW03] W.-T. Balke and M. Wagner. Towards personalized selection of web services. In *Proc. of the 12th Int. World Wide Web Conference*, 2003.
- [GMP04] S. Grimm, B. Motik, and C. Preist. Variance in e-business service discovery. In *Semantic Web Services: Preparing to Meet the World of Business Applications, workshop at ISWC 2004*, 2004.
- [GP03] B. Grosz and T. Poon. SweetDeal: Representing agent contracts with exceptions using XML rules, ontologies, and process descriptions. In *Proc. of the 12th Int. Conf. on the World Wide Web (WWW 2003)*, Budapest, Hungary, May 2003.
- [Gri05] Grid Resource Allocation and Agreement Protocol Working Group. Web services agreement specification. <https://forge.gridforum.org/projects/graap-wg/document/WS-AgreementSpecification/en/7>, June 2005.
- [Gru93] T. R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.
- [HM05] P. Haase and B. Motik. A mapping system for the integration of owl-dl ontologies. In A. Hahn, S. Abels, and L. Haak, editors, *IHIS 05: Proceedings of the first international workshop on Interoperability of heterogeneous information systems*, pages 9–16, 2005.
- [HPS04] I. Horrocks and P. F. Patel-Schneider. A proposal for an OWL rules language. In *Proceedings of the 13th International Conference on the World Wide Web (WWW 2004)*, pages 723–731, New York, USA, 2004. ACM Press.

- [HPSB⁺04] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosz, and M. Dean. SWRL: A semantic web rule language combining OWL and RuleML. W3C Submission, available at <http://www.w3.org/Submission/SWRL>, May 2004.
- [Kag04] L. Kagal. *A Policy-Based Approach to Governing Autonomous Behavior in Distributed Environments*. PhD thesis, University of Maryland Baltimore County, Baltimore MD 21250, November 2004.
- [KFKS05] M. Klusch, B. Fries, M. Khalid, and K. Sycara. Owls-mx: Hybrid semantic web service retrieval. In *Proceedings of 1st Intl. AAAI Fall Symposium on Agents and the Semantic Web*, Arlington VA, USA, 2005. AAAI Press.
- [Kie02] W. Kießling. Foundations of preferences in database systems. In *VLDB'02: Proc. of the 13th Int. Conf. on Very Large Databases*, pages 311–322, 2002.
- [KLP⁺04] U. Keller, R. Lara, A. Pollers, I. Toma, M. Kifer, and D. Fensel. WSMO Web Service Discovery, November 2004. <http://www.wsmo.org/TR/d5/d5.1/v0.1>.
- [KR76] R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. J. Wiley, New York, 1976.
- [LAO⁺06] S. Lamarter, A. Ankolekar, D. Oberle, R. Studer, and C. Weinhardt. A policy framework for trading configurable goods and services in open electronic markets. In *Proceedings of the 8th Int. Conference on Electronic Commerce (ICEC'06)*, New Brunswick, Fredericton, Canada, August 2006.
- [LH03] L. Li and I. Horrocks. A software framework for matchmaking based on semantic web technology. In *WWW '03: Proceedings of the twelfth international conference on World Wide Web*, pages 331–339. ACM Press, 2003.
- [LL87] M. Lacroix and P. Lavency. Preferences; putting more knowledge into queries. In *VLDB '87: Proceedings of the 13th International Conference on Very Large Data Bases*, pages 217–225, San Francisco, CA, USA, 1987.
- [MM03] D. J. Mandell and S. McIlraith. Adapting BPEL4WS for the Semantic Web: The Bottom-Up Approach to Web Service Interoperation. In D. Fensel, K. P. Sycara, and J. Mylopoulos, editors, *2nd Int. Semantic Web Conference (ISWC)*, volume 2870 of *LNCS*, pages 227–247, Sanibel Island, FL, USA, 2003. Springer.
- [MS06] B. Motik and U. Sattler. A comparison of reasoning techniques for querying large description logic aboxes. In *Proc. of the 13th International Conference on Logic for Programming Artificial Intelligence and Reasoning (LPAR 2006)*, Phnom Penh, Cambodia, November 2006.

- [MSS05] B. Motik, U. Sattler, and R. Studer. Query answering for OWL-DL with rules. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 3(1):41–60, JUL 2005.
- [NSDM03] T. D. Noia, E. D. Sciascio, F. M. Donini, and M. Mongiello. A system for principled matchmaking in an electronic marketplace. In *WWW '03: Proceedings of the twelfth international conference on World Wide Web*, pages 321–330. ACM Press, 2003.
- [OLG⁺06] D. Oberle, S. Lamarter, S. Grimm, D. Vrandecic, S. Staab, and A. Gangemi. Towards ontologies for formalizing modularization and communication in large software systems. *Journal of Applied Ontology*, 1(2), 2006.
- [OVSH06] N. Oldham, K. Verma, A. Sheth, and F. Hakimpour. Semantic WS-Agreement partner selection. In *Proc. of the 15th Int. WWW Conference*, Edinburgh, UK, 2006.
- [PKPS02] M. Paolucci, T. Kawamura, T. R. Payne, and K. P. Sycara. Semantic matching of web services capabilities. In *1st International Semantic Web Conference (ISWC)*, LNCS 2342, pages 333–347, 2002.
- [SNVW06] B. Schnizler, D. Neumann, D. Veit, and C. Weinhardt. Trading grid services - a multi-attribute combinatorial approach. *Forthcoming in European Journal of Operational Research*, 2006.
- [SPAS03] K. Sycara, M. Paolucci, A. Ankolekar, and N. Srinivasan. Automated discovery, interaction and composition of semantic web services. *Journal of Web Semantics*, 1(1), 2003.
- [SRT05] A. P. Sheth, C. Ramakrishnan, and C. Thomas. Semantics for the semantic web: The implicit, the formal and the powerful. *Int. J. Semantic Web Inf. Syst.*, 1(1):1–18, 2005.
- [UBJ04] A. Uszok, J. M. Bradshaw, and R. Jeffers. KAoS: A Policy and Domain Services Framework for Grid Computing and Semantic Web Services. In *Trust Management: 2.d Int. Conference, iTrust 2004, Oxford, UK*, volume 2995 of LNCS, pages 16–26. Springer, 2004.
- [W3C04] W3C. Web ontology language (OWL). <http://www.w3.org/2004/OWL/>, 2004. W3C Recommendation.
- [ZBN⁺04] L. Zeng, B. Benatallah, A. H. Ngu, M. Dumas, J. Kalagnanam, and H. Chang. QoS - aware middleware for web services composition. *IEEE Transactions on Software Engineering*, 30(5):311–327, 2004.

Einführung in den Track

Service Oriented Computing

Prof. Dr. Torsten Eymann

Universität Bayreuth

Prof. Dr. Manfred Grauer

Universität Siegen

Dr. Alfred Geiger

T-Systems

Service Oriented Computing auf der Basis von Peer-to-Peer (P2P)- und Grid-Technologie ist Gegenstand intensiver weltweiter Forschungstätigkeit. Dabei erfolgt die Realisierung heute in den meisten Fällen mittels Standard-Webservices. Zu den erhofften Vorteilen gegenüber traditionellen Lösungen zählen neben besserer Skalierbarkeit, dezentralem Management, reduzierter Produktentwicklungszeit, höherer Fehlertoleranz und niedrigeren Betriebskosten auch neue Formen und Dimensionen inner- und überbetrieblicher Integration, Koordination und Bereitstellung von Ressourcen.

Service Oriented Computing beschreibt ein neues Paradigma für Distributed Computing, welches das Potential hat, die Erstellung, Architektur, Verteilung und Ausführung von Softwareanwendungen grundlegend zu verändern. Services stellen autonome, plattformunabhängige Softwaremodule dar, die zu kollaborativen SW-Applikationen vernetzt werden können. Standardisierte Protokolle wie die Web-Service Technologie ermöglichen eine übergreifende Beschreibung, Dienstfindung und Orchestrierung dieser Systeme.

Programmkomitee:

Prof. Dr. Martin Bichler, Technische Universität München

Prof. Dr. Wolfgang Gentzsch, D-Grid-Koordinator und GGF

Dr. Sven Graupner, HP Palo Alto

Prof. Dr. Wilfried Juling, Universität Karlsruhe (TH)

Prof. Dr. Stefan Kirn, Universität Hohenheim

Dr. Jörg Noack, BMW München

Prof. Dr. Detlef Schoder, Universität zu Köln

Prof. Dr. Ulrich Trottenberg, Fraunhofer Institut SCAI und Universität zu Köln

Negotiations in Service-Oriented Architectures

Falk Kretzschmar

Dept. of Management Information Systems
Martin-Luther-University Halle-Wittenberg
Universitätsring 3
D-06108 Halle, Germany
kretzschmar@wiwi.uni-halle.de

Abstract

The software infrastructure of today's enterprises is transforming from monolithic application software into more flexible component-based and modular architectures. In this context Service-Oriented Architectures (SOAs) are discussed as a paradigm where application components can be accessed and exchanged through a service broker. Existing standards for implementing the service broker in SOA only allow searching for services based on functional parameters. Especially in a B2B setting where the involved enterprises are economically independent instances the price for the exchanged services is of particular interest. This causes the increasingly important question of how the economic perspective can be added to SOA. Negotiations between service provider and service requestor are one possible extension to the basic architecture. Subject of these negotiations can be non-functional parameters such as price and Quality of Service (QoS) for the exchanged services.

1 Introduction

Negotiations in Service-Oriented Architectures (SOAs) are a current and interdisciplinary field of research. The idea that applications can be automatically composed out of reusable components seems to be very promising for researchers as well as for practitioners. One foundation to achieve this flexible application architecture is to choose the right components based on their functional and non-functional parameters. A further step to a more flexible composition of reusable components is to negotiate the terms of use for these components.

In this article we summarize key concepts for negotiations in general and we discuss their applicability for extending SOA. The article is organized as follows: After a short introduction and the definition of fundamental terms the second part of the article introduces main negotiation concepts for including an economic perspective in SOA. The third part suggests technologies for the implementation of the described negotiations with special emphasis on semantic description for the exchanged services. The article ends with a discussion of the proposed concepts and a short outlook on future work.

The main idea of SOA is to provide the functionality of applications as a service and to allow a simple way to access this service via a web infrastructure. Important goals for using a SOA are reuse of existing components, interoperability, loose coupling, and the possibility to flexibly adjust the applications to the company's business processes [Alon03], [Zimm03]. According to the World Wide Web Consortium (W3C) specification SOA consists of three main parts: The service provider publishes and provides services, the service broker contributes to publishing and finding of the services and the service requestor finds the adequate services. Figure 1 shows the mentioned SOA components and their interaction [W3C104].

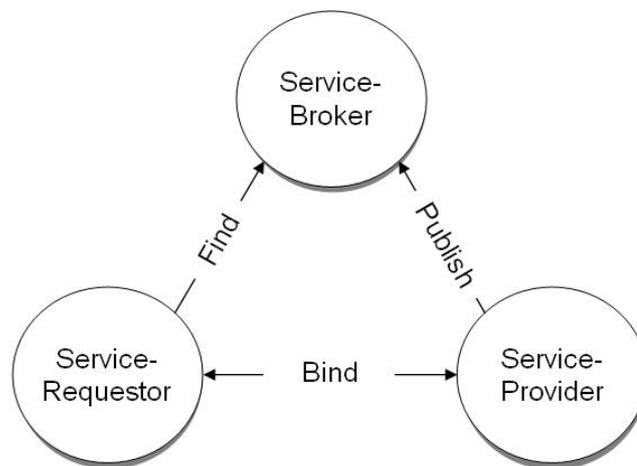


Figure 1: Components of a Service-Oriented Architecture

Web services are currently the most popular technology for implementing a SOA. The basic web services standards are Simple Object Access Protocol (SOAP) for messaging, Web Service Description Language (WSDL) for interface description, and Universal Description, Discovery, and Integration (UDDI) as an optional technology for implementing the service broker. Besides this set of XML-based technologies there exist many extensions to SOA in terms of composition and coordination of services as well as to ensure transactions and security.

In addition to these extensions there exists the approach to use the ideas of the semantic web in SOA. Standards like Web Ontology Language (OWL) and Resource Description Framework (RDF) were developed to describe the semantics of web resources. Web service discovery based on OWL descriptions is a very promising idea, but there is still a lot to be done with regard to the semantic description of services to finally provide an automated semantic search for web services. This automated semantic search is one of the main foundations for negotiations between service provider and service requestor. Before actually implementing a negotiation mechanism in SOA, there needs to be a description of the traded services which is understood by both negotiation parties. This means a common ontology is needed to specify the context of the negotiation and the terms of use for the exchanged services. Currently this is not possible with the basic SOA standards: UDDI allows the search for services based on their functionality. Other search criteria such as price and Quality of Service are not considered, which means the negotiation about economic criteria is still taking place outside the basic architecture in a separate process. This contradicts the above mentioned goals for using SOA, in particular the flexible on-demand access to services and the loose coupling between service provider and service requestor.

2 The economic perspective on SOA

2.1 Definition and classification

From a technological perspective SOA is an architecture for a distributed system that provides the functionality of application components as services. These services are accessed by using the common internet protocols. Whereas in the domain of computer science and information systems the technological aspects of SOA are well researched the economic perspective on SOA is often neglected.

From an economic perspective the service provider is offering a digital service and the service requestor is consuming it on-demand. So compared to the technological perspective on SOA technical interactions become economic transactions.

Negotiations in SOA describe the process of elaborating an agreement between service provider and service requestor based on economic criteria. Besides functional search criteria non-functional criteria such as price and Quality of Service are considered as important negotiation

criteria for exchanging services in SOA [Bole04], [Dan04], [Ludw03], [Sing05]. On the one hand the service provider describes a service offer based on these criteria and on the other hand the service requestor is searching for adequate services based on his search criteria. Negotiating an agreement between service provider and service requestor leads to an economic coordination between the involved negotiation parties. To analyze the economic coordination in more detail it is a common approach to look at the negotiation process using a phase model. Using this model the economic coordination in SOA is a process consisting of four transaction phases. During the information phase the service requestor is getting information about offered services and possible transaction partners. In the negotiation phase the transaction partners are selected and the contract for using the service is elaborated. This contract is called Service Level Agreement (SLA), which is established and agreed upon in the agreement phase. The fulfillment phase finally comprises service delivery, payment and further support. Figure 2 shows the four transaction phases that are similar to the transaction phases in electronic markets [Lind98].

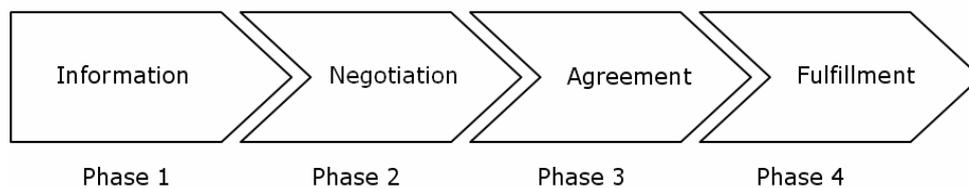


Figure 2: Transaction phases for the economic coordination in SOA

For implementing the negotiation phase there exist many different concepts. Among them are the very prominent auction mechanisms. These negotiation concepts are often referred to as general concepts without an explicit relation to negotiations in SOA. In [Sege99], [Bich01] it is mentioned that simple auctions, extended auctions, negotiation support systems, and agent-based negotiations are main concepts for implementing the negotiation phase. Table 1 shows a classification of negotiation concepts, which considers the architecture, negotiation rules, number of negotiation attributes, and the degree of automation.

Category:	Negotiation concepts:
Architecture	1:1 negotiation, m:n marketplace, brokered marketplace
Negotiation rules, number of negotiation attributes	Simple auction mechanisms: English auction, Dutch auction, Vickrey auction
	Extended auction mechanisms: double, multi-unit, multi-attribute auction
Degree of automation	Negotiation Support Systems, Agent-Based Negotiations

Table 1: Classification of negotiation concepts

These different concepts for implementing the negotiation phase derive from the parameters that characterize a particular negotiation scenario. Many authors state that besides the pure negotiation mechanism the actual setting of a negotiation is of main importance [Kris02], [Milg04]. To describe the setting of negotiation scenarios in SOA we use the parameters that are listed in table 2.

Category:	Negotiation parameters:
Negotiation partners	Number of negotiation partners, Strategic behavior of negotiation partners
Negotiation mechanism	Rules for matchmaking of negotiation partners, Rules for pricing the traded service
Negotiation subject	Properties of the traded service
Technological setting	Network and server performance, Standards and communication protocols, Semantic description of services and negotiation ontology

Table 2: Parameters for the negotiation setting

The above mentioned parameters can be used for a general description of negotiation scenarios. Depending on these parameters the outcome of the negotiation phase varies considerably. For instance the number of negotiation partners influences the negotiation power of one particular

negotiation partner and leads to specific strategic behavior. Changing the number of negotiation partners means that their original strategy is not appropriate anymore. In addition to the strategy derived from the number of participants it is possible to define a basic strategy for each actor or for types of actors. This strategy is based on the rules that are defined in the negotiation mechanism.

We explained earlier that in the negotiation phase transaction partners are matched and the price of the traded service is determined. The second category of negotiation parameters describes the actual matchmaking and the pricing of services. In the case of negotiations between service provider and service requestor the properties of the traded service can be important negotiation parameters as well as technical parameters like network and server performance or the availability of semantic descriptions for services.

2.2 Negotiations in SOA and possible architectural extensions

For the scope of this article the negotiation phase is of special interest. We examine the question of how the transaction partners should be matched and what pricing mechanism should be used for trading services in SOA. In 2.1 we named general concepts for implementing the negotiation phase. In real-world negotiations the different negotiation concepts can occur as a combination of the above mentioned categories. For instance, on a brokered marketplace a multi-attribute auction can be used for negotiating between software agents. Our goal is to adapt the general negotiation mechanisms so that they match with the special setting for negotiations in SOA.

SOA is an architectural concept for application software which is often referred to as software as a service [Elfa04], [Sing05]. So in broad terms the negotiation subject is a service, the negotiation partners are applications or organizations that run these applications and the technological setting is the Internet. Because of the SOA goal to flexibly adjust to a company's business processes by selecting the best suited service for the current task, we need automated selection of services based on a semantic search. For developing a concept for negotiations in SOA this means we also need an approach for automated negotiations. Negotiation Support Systems include a human decision and are therefore not matching with the above mentioned SOA goals and the technological setting. By contrast there are also negotiation concepts that allow automated negotiations between negotiation partners. Agent-based negotiations fall into that category. In agent-based negotiations a strategy based on a fixed set of rules can be

implemented. This means the software agent is able to fulfill the task of an automated search and selection of services.

For the remainder of this article we focus on an extended service broker for automated negotiations in SOA. By definition SOA implies the use of a service broker, but from the technological perspective SOA allows 1:1 negotiation, m:n marketplace as well as a brokered marketplace for negotiations between service provider and service requestor. Until now many companies do not use a service broker when implementing SOA, because it is not necessary for them. Integrating existent applications within a company is usually done without a service broker, although from the technology perspective the service broker can be used for intra-company transactions as well as for inter-company transactions. Despite the fact that the idea of a global marketplace for service has not been implemented yet, we consider the service broker to be a good starting point for extending SOA, because the broker can be used for storing metadata about services and service providers. Extending the service broker leads to a central instance that allows automated negotiations of service parameters. In table 3 main concepts for an extended service broker in SOA are listed. Applying the different pricing models means that the standard service broker is transformed into a simple catalogue, an extended catalogue, a simple auction mechanism, or an extended auction mechanism. Agent-based negotiations can also be used for automated negotiations in SOA, but they do not determine a special pricing model.

Pricing model:	Service broker:
Fixed price, “take it or leave it” offer	Simple catalogue
Fixed price, differentiated price	Extended catalogue
Dynamic pricing	Simple auction mechanisms
Negotiation of price, Quality of Service and other contract parameters	Extended auction mechanisms, Agent-Based Negotiations

Table 3: Concepts for an extended service broker in SOA

The above mentioned concepts for negotiations in SOA describe different pricing models for the traded service. Fixed prices and differentiated prices are a simple way to incorporate the economic perspective in SOA. The service provider is getting paid for the functionality of a software component and for the promise to deliver it on-demand. Ignoring the cost of developing and providing services would mean that the service provider will use other channels to distribute his software components. The problem with fixed prices and differentiated prices is that the service provider needs to know the demand function for the traded services. In real-world scenarios this is often not the case. Incomplete information about the market can be a problem for trading traditional products as well as for trading software services. With the different auction mechanisms the pricing of the traded services is done dynamically and according to the auction type specific rules. The service provider is getting information about the demand function for the traded services and usually earns higher profits.

3 Technologies for implementing negotiations in SOA

3.1 Existing standards

UDDI is often called a global marketplace for offering web services. In fact it is a catalogue that can be used by service providers to publish their services and by service requestors to find them. Because of the deficit of semantic service description in UDDI, it does not support automated negotiations between the involved transaction partners.

Nevertheless UDDI is a good starting point for extending SOA, since it already implements a structure for storing metadata: “white pages” include provider information, “yellow pages” provider and service information, and “green pages” information for calling the service. How these categories are matched with the UDDI data structure is described in figure 3. In addition to that it is shown that UDDI finally provides the reference to a WSDL document [Sing05].

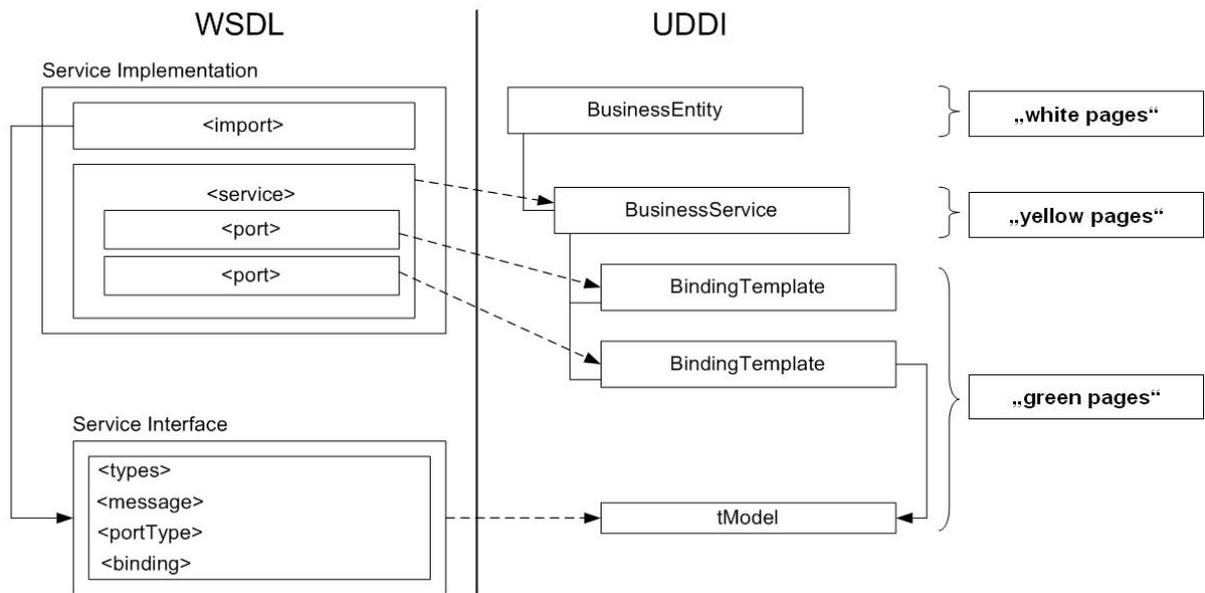


Figure 3: Link between UDDI metadata categories, UDDI data structure, and WSDL

With this reference the binding of the actual service is done through direct interaction between service provider and service requestor. Service binding is done either statically or dynamically. With the static approach services are connected to each other at design time. This can be problematic, in the case of changes of interfaces, messages, transport protocols, and network addresses. Furthermore static binding is an inflexible and in the case of frequent changes costly way to operate a distributed system. By contrast the dynamic approach allows the binding of services at runtime. With this approach UDDI delivers a set of services that implement exactly the same interface.

For automated negotiations in SOA this is not enough, because the same interface does not guarantee the same semantics of two services. In addition to that the metadata about services and service providers in UDDI is not sufficient. For that reason there already exist a number of extensions to the basic architecture. In majority these extensions were developed in research projects and make use of semantic web technologies to extend SOA. We differ between the following two categories:

1. Service broker with semantic extensions
2. Interface description with semantic extensions.

Extending the service broker is a sound method for developing a catalogue that includes more information on services than UDDI. When incorporating negotiation mechanisms like auctions the service broker is the central instance for clearing offers and counteroffers. Research projects and prototypes that fall into the category of a service broker with semantic extensions are UDDIe, OWL-S Matchmaker, and WS-QOS [Shai03], [Chen03], [Srin04], [Paol02], [Gram03]. Interface descriptions with semantic extensions match with SOA implementations that do not use a service broker, because the functional and non-functional parameters of services are described by the interface. In this case lightweight approaches of service discovery like Web Service Inspection Language (WSIL) can be used, but implementing UDDI is not necessary. Research projects and prototypes that fall into the category of interface descriptions with semantic extensions are WSOL, SLang, WSML, WSDL-S, and WSLA [Pate03], [Lama03], [Verh04], [Tian04].

3.2 Extended service broker architecture for automated negotiations in SOA

For implementing automated negotiations in SOA we propose the extended service broker architecture. It consists of a standard UDDI registry extended by Web Ontology Language for Services (OWL-S) service profiles, the OWL-S matchmaker and the negotiation mechanism. A conceptual overview of this extended service broker is depicted in figure 4.

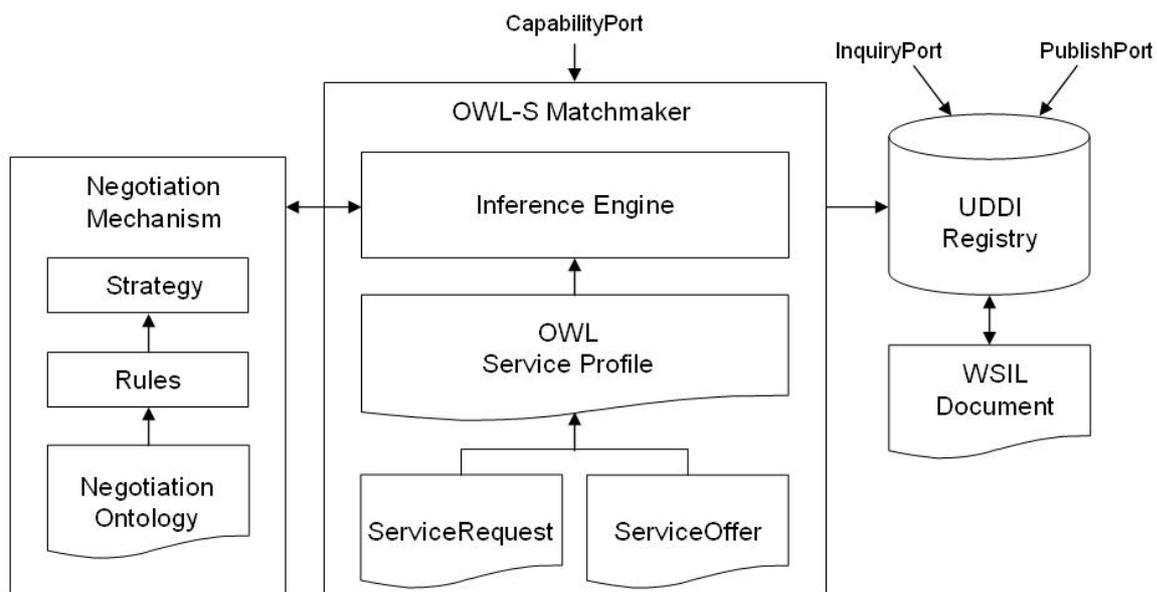


Figure 4: Extended service broker architecture for automated negotiations in SOA

Compared to a standard service broker with this architecture it is possible to describe service profiles. With the help of these service profiles the service requestor can define a service request, which contains parameters the service provider should guarantee. For the service provider the service offer contains service parameters he is able to guarantee. Based on that semantic description of services the inference engine of the OWL-S matchmaker matches the two profiles and by using the UDDI registry or the WSIL document it can finally discover and select the requested service. How the actual matching is done is implemented in the negotiation mechanism that defines the negotiation rules and the negotiation strategy.

An important part of the extended service broker architecture is the negotiation ontology. OWL-S can be used to define an ontology that allows a common understanding of the negotiation subject [Buss02], [Paol02]. For automated negotiations in SOA we propose a negotiation ontology that consists of several non-functional service parameters. Table 4 shows the main non-functional service parameters service price and Quality of Service as well as additional parameters like reputation of negotiation partners or duration of service lease.

Non-functional service parameters:	
Service price	
Quality of Service	Server performance: response time, availability, throughput, reliability Network performance: bandwidth, latency, jitter
Security	
Negotiation mechanism	
Number of negotiation partners	
Reputation of negotiation partners	
Duration of service lease	
Sanctions or penalties in case of non-compliance to the service level agreement	

Table 4: Non-functional service parameters for negotiations in SOA

Negotiating a Service Level Agreement between service provider and service requestor is a complex task. Before the actual negotiation process starts the service profiles for searching

services and for providing services have to be defined. In addition to that for the service provider it is necessary to establish an architecture for measuring and monitoring the service parameters. During the negotiation phase all parameters must be understood by both negotiation parties. The negotiation ontology describes the meaning of the above mentioned non-functional service parameters in a machine-readable way. Using this description the rules of the negotiation mechanism consider the service parameters and their meaning for the negotiation parties.

We argue that for the main non-functional parameters price and Quality of Service it is likely that service provider and service requestor have a common understanding. Furthermore service provider and service requestor know how to measure these service parameters. With the additional non-functional parameters such as reputation of negotiation partners and security it is difficult to come to an agreement. Measuring and monitoring the additional non-functional service parameters is more complex, because it requires a reputation mechanism and policies for secure service delivery.

At this point the proposed extended service broker architecture for automated negotiations in SOA is a conceptual architecture. Semantic service discovery, which is part of the architecture, is a main foundation for implementing automated negotiations in SOA.

4 Discussion and future work

This article gave a short overview of existing technologies and further concepts for implementing negotiations in SOA. We argued that in the case of economically independent enterprises the economic perspective for exchanging services is of particular interest. We introduced possible extensions to SOA and suggested current technologies for their implementation. The most important extension is an architecture that allows automated negotiations of price and other parameters such as Quality of Service for the exchanged services.

The use of semantic web technologies like OWL-S is a very promising approach to allow an economic coordination in SOA. The proposed extended service broker architecture makes use of a negotiation ontology that allows a common understanding of the negotiation subject. Service profiles are used to describe service requests and service offers that are matched by the inference engine of the OWL-S matchmaker.

For further research it is necessary to work on a SOA specific and more flexible negotiation mechanism. The goal to adapt the mentioned standard negotiation mechanisms so that they match with the technological setting of SOA can be achieved by conducting further research on the outcome of negotiations in SOA. Our future work will include an analysis of design decisions and their impact on profit of service providers, achieved utility of service requestors, and delivery rate of services.

References

- [Alon03] Alonso, G.: Web Services: Concepts, Architectures and Applications. Springer, 2003.
- [Bich01] Bichler, M.: The Future of E-Markets: Multi-Dimensional Market Mechanisms. Cambridge University Press, Cambridge, 2001.
- [Bole04] Boles, D.: Erlösformen für Web Content und Services. Informatik Springer, 2004.
- [Buss02] Bussler, C., Fensel, D., Maedche, A.: A conceptual architecture for semantic web enabled web services. ACM SIGMOD Record, v.31 n.4, December 2002.
- [Chen03] Chen, Z. et al.: UX -An Architecture Providing QoS-Aware and Federated Support for UDDI, ICWS '03, School of Computer Engineering, Nanyang Technological University, 2003.
- [Dan04] Dan, A. et al.: Web services on demand: WSLA-driven automated management. IBM Systems Journal, v.43 n.1, p.136-158, 2004.
- [Gram03] Gramm, A.: WS-QOS - Ein Rahmenwerk für Dienstgüteunterstützung in Web Services, Institut für Informatik, Freie Universität Berlin, 2003.
- [Elfa04] Elfatraty, A., Layzell, P.: Negotiating in Service-oriented Environments. Comm. ACM, 47 (8), 103-08., 2004.
- [Kris02] Krishna, V.: Auction Theory. Academic Press, 2002.

- [Lama03] Lamanna, D. et al.: SLAng: A language for Service Level Agreements, Departement of Computer Science, University College London, 2003.
- [Lind98] Lindemann, M.A., Runge, A.: Electronic Contracting within the Reference Model for Electronic Markets, 6th European Conference on Information Systems (ECIS-98). Aix-en-Provence, France, 1998.
- [Ludw03] Ludwig, H.: Web Services QoS: External SLAs and Internal Policies Or: How do we deliver what we promise? In Proc. of the First Web Services Quality Workshop, IEEE Computer Society, Rome, 2003.
- [Milg04] Milgrom, P.R.: Putting Auction Theory to Work (Churchill Lectures in Economics). Cambridge University Press, 2004.
- [Paol02] Paolucci, M. et al.: Importing the Semantic Web in UDDI, in Proceedings of Web Services, E-business and Semantic Web Workshop 2002.
- [Pate03] Patel, K., Tasic, V.: Web Service Offerings Language (WSOL): Characteristics, Applications, and Tools. 2003.
- [Shai03] Shaikh A. et al.: UDDIe: An Extended Registry for Web Services, Department of Computer Science, Cardiff University, UK, 2003.
- [Sege99] Segev, A., et al.: On Negotiations and Deal Making in Electronic Markets, Information Systems Frontiers, v.1 n.3, p.241-258, October 1999.
- [Sing05] Singh, M., Huhns, M.: Service-Oriented Computing: Semantics, Processes, Agents. Wiley, 2005.
- [Srin04] Srinivasan, N. et al.: Adding OWL-S to UDDI, implementation and throughput, First International Workshop on Semantic Web Services and Web Process Composition, San Diego, California, USA, 2004.
- [Tian04] Tian, M. et al.: A Survey of current Approaches towards Specification and Management of Quality of Service for Web Services. Freie Universität Berlin, Institut für Informatik, 2004.

- [Verh04] Verheecke, B. et al.: AOP for Dynamic Configuration and Management of Web Services, *International Journal of Web Services Research*, 1(3), 25-41, 2004.
- [W3C104] Booth et al.: *Web Services Architecture*. W3C Working Group Note, 2004.
<http://www.w3.org/TR/ws-arch/>
- [Zimm03] Zimmermann, O., Tomlinson, M., Peuser, S.: *Perspectives on web services: applying SOAP, WSDL, and UDDI to real-world projects*. Springer, Berlin, 2003.

Evaluation of Economic Resource Allocation in Application Layer Networks – A Metrics Framework

Werner Streitberger, Torsten Eymann
Chair of Information Systems
University of Bayreuth, Germany
95440 Bayreuth
{streitberger,eymann}@uni-bayreuth.de

Daniel Veit
Chair of Information Systems, E-Business and E-Government
University of Mannheim, Germany
veit@uni-mannheim.de

Michele Catalano
Department for Economics
University of Ancona, Italy
catalano@dea.unian.it

Gianfranco Giulioni
Department of Quantitative Methods and Economic Theory
University of Chieti-Pescara, Italy
giulioni@dea.unian.it

Liviu Joita, Omer F. Rana
School of Computer Science and the Welsh eScience Centre
University of Cardiff, UK
{l.joita,o.f.rana}@cs.cardiff.ac.uk

Abstract

Support for economic resource allocation in Application Layer Networks (such as Grids) is critical to allow applications and users to effectively exploit computational and data infrastructure as a utility. The evaluation of resource allocation strategies plays a major part in the selection of a resource allocation method. This paper presents an evaluation framework for resource allocation methods in Application Layer Networks that attempts to support both a technical and an economic evaluation. The model uses a layered metrics pyramid with different aggregation levels and statistical methods.

On top of the pyramid, only one number, the social utility, is able to characterize an economic resource allocation method. This single number may serve to compare different resource allocation strategies. However, one can also determine values obtained for metrics at intermediate levels. We demonstrate using a prototype application how such a metrics pyramid may be deployed in reality, focusing on the implementation of the lower levels of the pyramid.

1 Introduction

An Application Layer Network (ALN) describes an abstraction that integrates different overlay network approaches, like Grid and Peer-2-Peer systems on top of the physical connectivity provided by the Internet. A common characteristic of ALNs is the redundant, distributed provisioning and access to data, computation or application services, while hiding the heterogeneity of the service and resource network from the user. One of the main issues in the development of future ALNs is the efficient exploitation of services and resources available in the network.

A key requirement of ALNs is to support scalable, dynamic, and adaptable allocation mechanisms. This issue is being addressed by a number of Grid and P2P projects such as Globus [FoKe97], Legion [CKKG99], Nimrod/G [BuAG00], CATNETS [EACC05], and Gnutella [AdHu00]. Although considerable progress has been made developing software architectures and allocation methods, which allow clients to obtain services “on demand”, the evaluation of these methods with respect to each other is rarely undertaken. What is missing is a standard set of metrics that could be used as the basis for such an evaluation. It is important that such metrics allow comparison of both the application “behaviour”, along with the infrastructure behaviour.

Currently, no metric framework exists that takes into account the characteristics of applications that may be deployed over ALNs – especially ones that measure the performance of the resource allocation strategy being used. Current Grid and Peer-2-Peer applications often use a service-oriented architecture, which is characterized by dynamic and heterogeneous resources. In such environments one of the key issues is the assignment of resources to services. This paper assumes that resource allocation is being undertaken in a “market” based environment – allowing service users and providers to sell and buy services and resources [ReSE06]. It is also assumed that two concurrent markets co-exist – a “resource market” and a “service market”. A service instance can be provisioned by different possible, even heterogeneous resources; this heterogeneity, however, is shielded from the user, who acquires services on a virtualized, homogeneous service market. The presence of a market allows to use economic concepts with reference to the allocation strategy being adopted [ENRS06], e.g. different auction types, bilateral bargaining, market-oriented programming or computable general equilibrium approaches.

The characteristics of e.g. Grid and P2P applications in ALNs define particular resource allocation requirements which have the following characteristics [MoBa03]:

- **Dynamicity:** There are changing environments and there is a need for adaptation to changes.
- **Diversity:** Requests may have different priorities and responses should be assigned according to them. It is therefore possible for a number of different types of requests to co-exist, and determine how these should be handled becomes significant.
- **Large scale:** There are often a large number of nodes, and it is often necessary to group these elements to create hierarchical organisations.
- **Partial knowledge:** It is not possible to determine at any point in time the full state of the system or network.
- **Complexity:** Learning mechanisms are necessary to adapt to changes in infrastructure, and optimal solutions are not easily computable.
- **Evolutionary:** The system is open to changes which cannot be taken into account in the initial setup.

The goal of the evaluation framework presented here is to define a general set of metrics for performance analysis, which can be used in projects evaluating and comparing the performance of resource allocation mechanisms running in network exhibiting the above characteristics. Such a framework should include both technical parameters (such as quality of service factors – like response time) and economic parameters (such as overall cost of computation or data access), to compare different resource allocation methods with each other.

This paper is structured as follows: Section 2 presents the metrics used for the evaluation framework. Section 3 discusses how these metrics may be used for a technical and economical evaluation. A demonstration of the approach using a Grid prototype application is presented in Section 4, and a summary of the approach is provided in Section 5.

2 The Metrics Pyramid

It is often useful to be able to compare two allocation methods using a single index or number. Such an index provides an aggregated behaviour of an allocation method with reference to a number of features. The application of statistics is necessary to achieve the index, but is useful only in conjunction with a detailed measurement framework. The results of the measurement are intended to allow providing feedback on system behaviour, improve chances of successful adaptation, improve implementation, and increase accountability.

Figure 1 shows the logical structure of data and indices. It sums up the underlying methodology establishing the shape of a pyramid. Data are the basic units of information, collected through technical monitoring of the application layer network. Parameters which are likely to be of significance within the application and the resource allocation mechanism have to be selected for measurement. These parameters define the raw disaggregated data. To ease the analysis of the raw data, they are collected from different experiments (simulation runs) into a database. Disaggregated indicators provide the first stage of evaluation, and may comprise of a number of independently measured values. They help to improve the implementation of the resource allocation mechanisms.

For further evaluation, it is obligatory to take into account a set of characteristics that are not directly comparable because these characteristics correspond to variables of different dimensions and unit of measurement. Thus, they have to be made comparable, e.g. by

normalization, and then grouped into indicators. An indicator is defined as a ratio (a value on a scale of measurement) derived from a series of observed facts, which can reveal relative changes as a function of time. They provide feedback on system behaviour and also allow the analysis of performance and predictions of future performance.

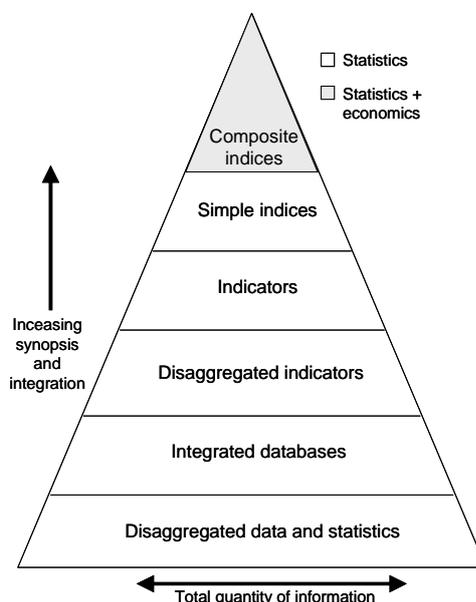


Figure 1: General method to obtain a composite index [PiZC00]

Finally, the simple and composite indices are computed, which represent performance benchmarks. They express information in ways that are directly relevant to the decision-making process. In general, indicators help the assessment, the evaluation, and most important, they help to improve accountability. Examples of applications using these techniques can be found in social sciences and in economics [Horn93][CoHR95][UNDP99].

Although highly aggregated indices are attractive because of their simplicity, they also carry risks. Most of all, aggregates tend to mask real-life complexity and detail for policy-making. Highly aggregated indices are important in giving a view of overall progress, but they should be readily disaggregated into its components that may help to specify reasons for the index going up or down and also answer questions of interest to decision-makers working on lower scales. The pyramid with its different layers supports this analysis progress, it enables both analysis at lower scales and decision-making on highly aggregated indices.

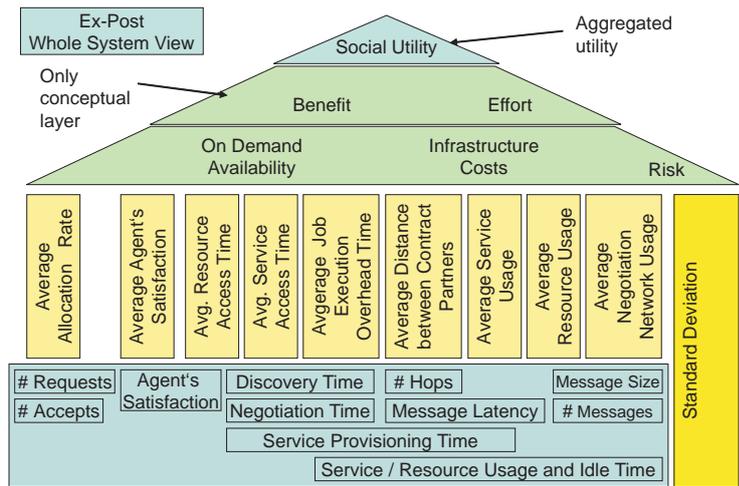


Figure 2: Metrics Pyramid

To support both technical and economic parameters, the evaluation process is divided into two layers. The first uses basic statistical concepts, while the second uses economic principles. The technical parameters, at the lower end of this pyramid, provide the basis for economic parameters that lie higher up in the pyramid, as illustrated in figure 2. A prerequisite is, that the key parameters at the bottom layer can be technically measured in the system to be analysed.

Aggregate indices may be generated out of several of these low level parameters. The parameters at the higher layers of the pyramid have economic semantics and therefore a higher significance. At the top of the pyramid there is a composite index defined as (global) “social utility” – providing a single metric to specify the overall performance of the allocation strategy being used in a particular application.

The selection of the technical parameters, and their aggregation to simple and composite indicators, can vary, depending on the resource allocation method and/or the measurable technical metrics of the application domain. The framework gives the possibility to add and remove metrics to optimize the framework for a special application domain without changing the general concepts used to build up this pyramid.

3 Technical and economical metrics

This section defines and illustrates various layers and aggregation steps of the metrics pyramid, (cf. figure 2), as used in the project CATNETS [EACC05] to analyse centralized and decentralized economic resource allocation methods.

The approach to technical metrics focuses on providing generic, easy to measure parameters, which can subsequently be aggregated. Technical layer metrics can be classified into: (I) efficiency measures (e.g. number of requests, number of acceptances); (II) utility measures (e.g. agent satisfaction); (III) time metrics (e.g. discovery time, negotiation time, service provisioning time) which are measures of the rate of change of market processes; and message-based metrics (IV) to measure the activity of users to communicate to find resources and services. The technical metrics include:

- **Number of Requests:** This metric measures demand, counted as the number of requests for services and resources.
- **Number of Acceptance:** The number of acceptance measures successfully requested (acknowledged) services and resources.
- **Agent Satisfaction:** The individual agent satisfaction measures the utility gain of a single transaction, which is the distance between his lowest price willing to pay in this transaction and the final price of the agreement. It is defined as the ratio between the subjective reservation value and transaction price.
- **Discovery Time:** This metric is used to measure the time to find a set of possible negotiation partners.
- **Negotiation Time:** The negotiation time measures the time needed to finish the negotiation between one buyer and one or several sellers. The measurement of the negotiation time starts after service discovery has completed, and ends before service provisioning.
- **Service Provisioning Time (Effective Job Execution Time):** The evaluation framework defines the service provisioning time as the service usage time of one transaction. It optionally includes also setup times, etc.
- **Hops:** The number of hops describes the distance between a service consumer and a service provider, counting the number of hops a message needs to traverse between them (this may be averaged over all the messages exchanged between a consumer and a provider).
- **Message Latency:** The message latency measures the time a message needs to arrive at the communication partner. Message latency is a parameter that indicates the performance of the physical network link and its message routing capabilities. It is expected that a large distance (cf. also number of hops) between a service consumer and

service provider should lead to higher message latency. Whether hops or latency is used to measure distance should be decided casually.

- Message Size: This metric is used to measure the message size (e.g. in kilobytes).
- Number of Messages: This value counts the number of messages needed for service allocation. The traffic on a physical network link could be computed by multiplying the message size with the number of messages sent.

These metrics are then used in two aggregation steps, as described below.

3.1 Simple indicators first aggregation

This layer defines a set of independent metrics which are normalized between 0.0 and 1.0. This makes it easier to find valid functions for the layers above, such as on demand availability and infrastructure cost. The technical metrics may be combined to obtain a framework that enables evaluation of different service oriented architectures. The aggregated metrics at the simple indicator layer are:

- Allocation Rate: This metric is a measure of the efficiency of the allocation process, which is computed using the number of requests and number of accepts. A buyer can demand services, but there is no guarantee that the allocation mechanism (centralized or decentralized exchange) performs a match between demand and supply.

$$allocation_rate = accepts/requests$$

- Agent Satisfaction: This metric implicitly shows the average surplus of the service provider or user (agent) in the system (normalized to the interval 0.0 and 1.0). A low value means that an agent has not been able to complete its goals successfully during the negotiation process. A high value means that the agent can constitute good results satisfying its requirements.
- Access Time: This indicator evaluates the time needed from the starting point of discovery until the final delivery of the service.

$$access_time = discovery_time + negotiation_time + provisioning_time$$

- Job Execution Overhead: This is the additional time needed for negotiation. It refers to the overhead introduced during the service negotiation process. The overhead is the sum of the service access time and the resource access time.

$$job_execution_overhead = access_time_service + access_time_resource$$

- Distance between Contract Partners: Message latency is the messaging time incurred by agents, and it is proportional to the distance between the sending and receiving nodes. The latter would be the ratio between the actual distance and the maximum distance between agents.
- Service/Resource Usage: The network usage will be evaluated by the ratio between provisioning time and the total simulation time. This evaluation would be conducted for both services and resources.

$$service_usage = provisioning_time / simulation_time$$

$$resource_usage = provisioning_time / simulation_time$$

- Network Usage: This metric is used to measure the total number of messages exchanged between two agents.

3.2 Composite indicators - second aggregation

Composite indicators are an aggregation of simple indicators, and are closer in scope to the application than to the infrastructure. Simple indicators are normalized between a [0; 1] interval in order to be comparable with each other. From the economics perspective, composite indicators measure the benefit from exchange (on-demand availability), and costs incurred from activities in the market (infrastructure costs). An optimal allocation mechanism would have an on-demand availability value = 1, and infrastructure costs = 0.

On DeMand availability (ODM) is a composite indicator obtained as mean of simple indicators, and may be derived as:

$$ODM = 1/3 (allocation_rate + agent_satisfaction + job_execution_overhead)$$

Infrastructure Cost (IC) is calculated in the same way. It is also a mean of multiple simple indicators, and may be derived as:

$$IC = 1/4 (distance + service_usage + resource_usage + network_usage)$$

It may be possible to model some of these metrics as stochastic variables, giving a mean and standard deviation over which the given metric varies. In economic applications, variance would be a measure for the overall “risk” to achieve stability of a given metric development.

3.3 The social utility - a composite index

To be able to use economic concepts into the evaluation process, a virtual policy maker is introduced, who aims to optimize the allocation. The behaviour of the policy maker is described with a goal function which is to be minimized under some constraints. The constraints are

determined by the structure of the economy, i.e. the underlying laws of the economy. The policy maker does not have complete knowledge of the economy, but uses a model of the economy based on statistical cause-and-effect relations. The policy maker collects all the metrics from the previous layer, has some preferences about ODM and IC and is aware about the distribution of profits and costs. Analogies exist in macroeconomic theory, and are used here with regard to two related models: Barro Gordon [BaGo83] and Poole [Pool70].

Mapping these models to the evaluation framework leads to the goal of improving the utility of society and simultaneously lowering the fluctuation in the economy. If the policy maker minimizes the inverse ODM and IC by reducing the distance to zero of those variables (optimal values), the social utility is improved. Moreover, he knows that inverse ODM and IC are distributed across the population (i.e. he knows the statistical moments of those variables) in a sense that he looks at the distribution of benefits and costs over the agents. A result of the final optimization rule is that the policy maker is interested in minimizing inequality (measured with respect to the metrics defined previously) between agents.

In order to follow this approach for ALNs, a loss function is built using the composite indicators ODM and IC. ODM and IC are taken as stochastic variables and their distribution across a population (the set of service users and providers) by the first and second moments are considered. Renaming the variables for the presentation X (inverse on demand availability index) and Y (costs index), the ALN loss function is

$$L = E[\alpha(X - X^*)^2 + \beta(Y - Y^*)^2],$$

where α and β are weights and X^* and Y^* are target values. It is important to give a value to the targets. As X is the inverse of the On Demand Availability it can be concluded that a lower value of X is better for the policy maker. Thus, the target value is set equal to zero ($X^* = 0$). Y is a direct measure of costs so it is natural to choose zero also for this variable ($Y^* = 0$). With this target, the loss function is

$$L = E[\alpha X^2 + \beta Y^2]$$

The random variables can be expressed as

$$X = \mu_x + u \text{ and } Y = \mu_y + z,$$

where u and z are random variables with zero means and positive variances ($\mu_u = 0$, $\mu_z = 0$, $\sigma_u^2 > 0$, $\sigma_z^2 > 0$). Note that variances σ_u^2 and σ_z^2 are the variance across a population.

Subsequently, the final social utility index may be specified as:

$$L = \alpha\mu_x^2 + \alpha\sigma_u^2 + \beta\mu_y^2 + \beta\sigma_z^2.$$

So one possibility for evaluating ALNs is to make use of:

$$\min \alpha\mu_x^2 + \alpha\sigma_u^2 + \beta\mu_y^2 + \beta\sigma_z^2.$$

This means that for the final evaluation of the scenarios, the first and second moments of ODM and IC have to be evaluated, whereas the first moment is the mean of ODM and IC, and second moment the variance of ODM and IC. In fact a policy maker has preference to minimize the inverse of ODM, thereby leading to the minimization of costs and the minimization of variability, i.e. the fair distribution of social utility between agents.

Applying the depicted metrics pyramid, a layered evaluation of different resource allocation methods is possible. The Grid application prototype, which is illustrated in the next section, uses the pyramid for a first evaluation of different scenarios. The focus of the evaluation is to show the relevance of this pyramid when being applied to a real Grid application prototype.

4 The Application Prototype

4.1 Application prototype description

The proof-of-concept prototype has been developed and named Catallaxy Data Mining (Cat-DataMining) [JPBG04]. Figure 3 shows the prototype components and related software agents as buyers and sellers in the Grid service market and Grid resource market. The prototype is composed of four main components: the Master Grid Service (MGS) – a Complex Service requestor; the Catallactic Access Point (CAP) – the entry point to the Complex Services for the Master Grid Service; the Data Mining Services – as types of Complex and Basic Services (BS); and job execution resources – as computational resources. The Complex Service (CS) agent acting on request of the MGS is the buyer entity in the service market, and the Basic Service is the seller entity in the service market.

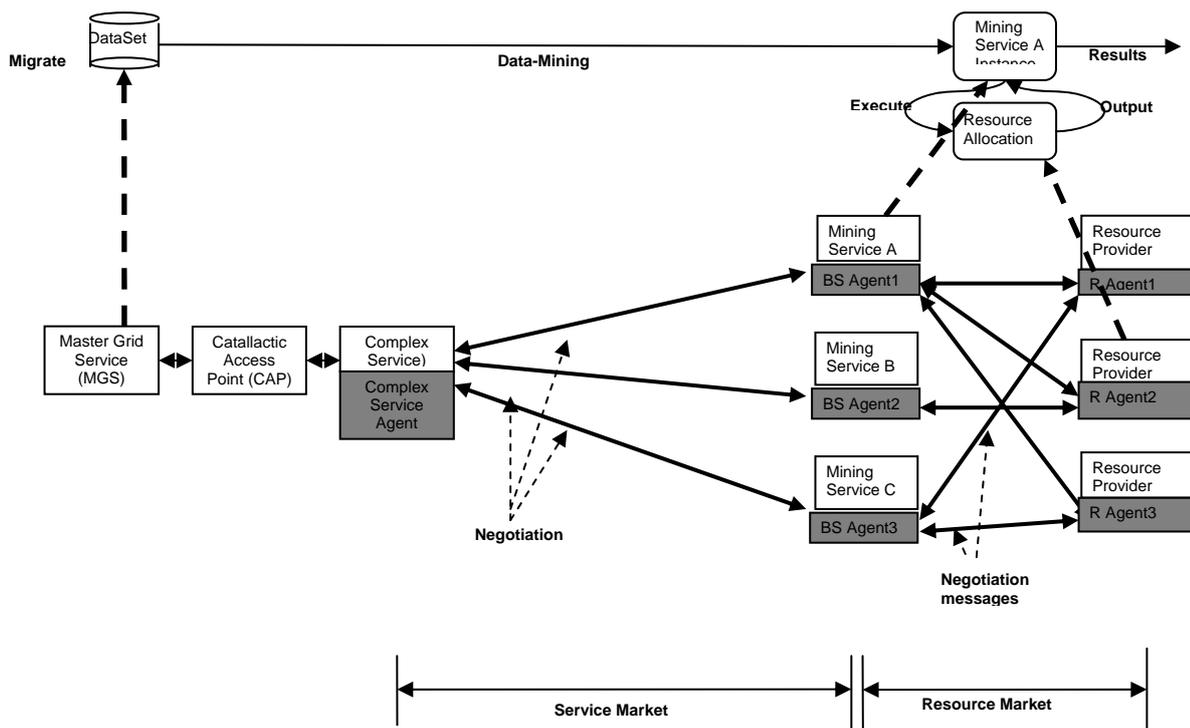


Figure 3: Cat-DataMining prototype – the Catalactic agents and the markets – one service invocation

The BS consists of a data mining job execution environment, which offers the deployment of multiple “slaves” able to execute the data mining task. Within the Cat-DataMining prototype, the data mining BS needs to be able to provide a response time as an important characteristic. With this goal the data mining BS agent buys resources in the resource market. Resource seller entities are able to provide a set of resources via the Local Resource Manager (LRM). The Resource Agents act on behalf of these Local Resource Managers, and provide an interface to the physical resources they manage.

The data mining BS agent is the buyer entity in the resource market, and the LRM agents are the seller entity on the resource market. The main functionalities of BS agents at the resource market are: (I) co-allocation of resources (resource bundles) by parallel negotiation with different resource providers (local resource manager entities); (II) informing the Complex Service about the outcome of resource negotiation.

4.2 Test-beds and preliminary results

We have performed several experiments in a cluster with the current version of the middleware, allowing for a preliminary performance assessment of the Grid Market Middleware (GMM) developed in the CATNETS project [JRCC06]. The goal of the experiments is to evaluate the autonomic behaviour of the GMM in terms of self-organisation, given by decentralized resource

discovery, and in terms of self-optimisation, given by adaptation to load and capacity of the resources.

For these experiments we have used for the economic agents an implementation of the ZIP (Zero Intelligence Plus) agents [PrTo98] which use a gradient algorithm to set the price for resources. Clients initiate negotiations with a price lower than the available budget. If they are not able to buy at that price, they will increase their bids until either they win or reach the budget limit.

Services will start selling the resources at a price which is solely influenced by the node's utilization, following the pricing model presented in [YeBu04]. As services get involved in negotiations, the price will also be influenced by demand. If a service agent is selling its resources, it will increase the price, when resources are sold to identify the price the market is willing to pay. When it no longer sells, it will lower the price until it becomes competitive again or it reaches a minimum price defined by the current utilization of the resource.

In order to test the performance of the market based resource allocation mechanism, we setup controlled experiments deploying several instances of the middleware in a Linux server farm. Each node has 2 CPU Intel PIII 1 GHz and 512 MB of memory. The nodes in the farm are connected by an internal Ethernet network at 100Mbps.

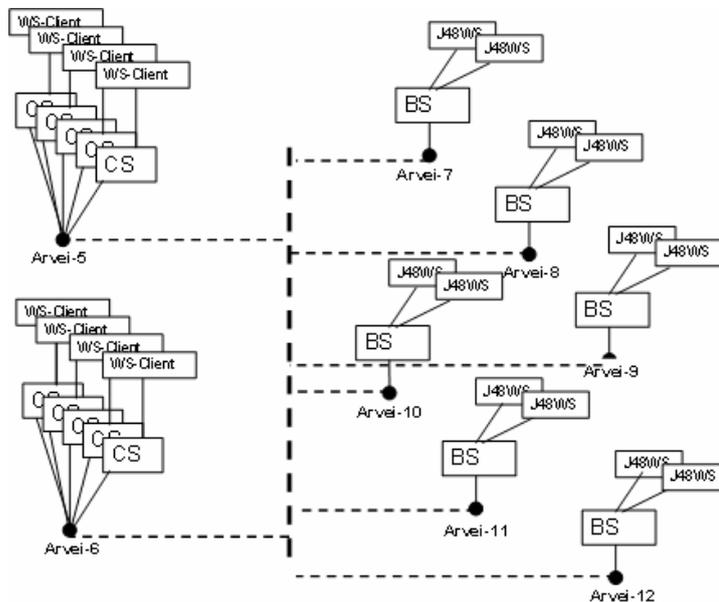


Figure 4: Experiment Setup

We deploy GMM and the Web Services on six nodes (named arvei- $\{7 \dots 12\}$). On each node we also deployed a Web Service which performs a CPU intensive calculation on the machines, increasing load. These Web Services are exposed in a Tomcat server. Access to execute these

Web Services on the Resources is what will be negotiated by the Service with the Clients. Figure 4 shows a schematic view of this deployment.

The experiments consisted in launching 2 Clients concurrently from 2 other nodes which were not running the Web Services. We generate a baseline load on all three nodes of 25% of CPU usage to simulate some background activity. Each Client performs 50 requests, in intervals of 10 seconds. Whenever a Client wins a bid with a Service, it invokes the Web Service in the selected node. The complete experiment runs for about 10 minutes. To better test autonomic load balancing, we artificially stressed one of the nodes (arvei-10) up to 95-100% of CPU usage for a short time during the experiment.

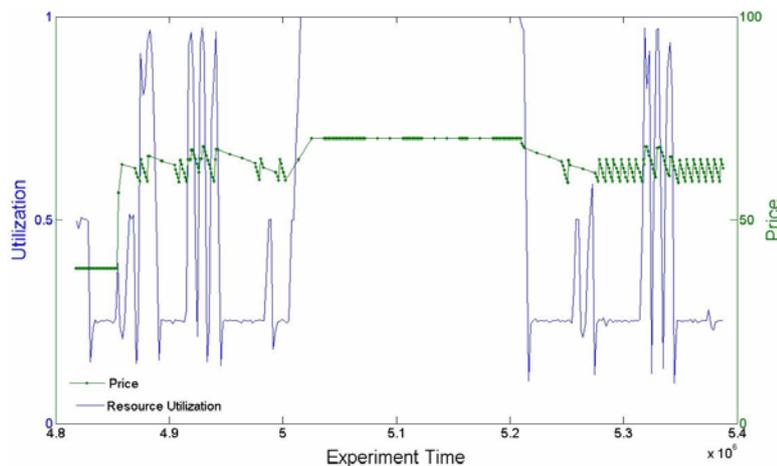


Figure 5: Experiment Results – Price and Utilisation Basic Service at node arvei12

Figure 5 and Figure 6 show the measured results of selected nodes used in the experiment. The measured metric price corresponds to the agent satisfaction of the presented metrics pyramid, the execution time to the provisioning time, and the utilisation metric to the resource usage. The metric negotiation time equals the negotiation time in the metrics pyramid. Both figures show the same experiment time interval. High utilisation of the CPU leads to higher prices of the resource CPU and to higher execution times, whereas the negotiation time doesn't significantly increase. Prices of the CPU are lowered time-delayed; a short increase of the utilisation increases the prices again. As expected, a lower utilisation of the CPU decreases the execution time.

Essentially, each data for each technical metric is referring to a single transaction and it has to be collected by the corresponding agent. Price and utilization data are collected periodically and independently by each transaction, giving a number of observations whereas execution time, negotiation time, and allocation are a dataset of 100 observations, which are collected by the complex services.

The computation works as follows: the first price data that has been collected after the execution data time stamp has been considered. Repeating the above procedure for each execution time, an observation data set of 100 items both for price and utilization has been extracted.

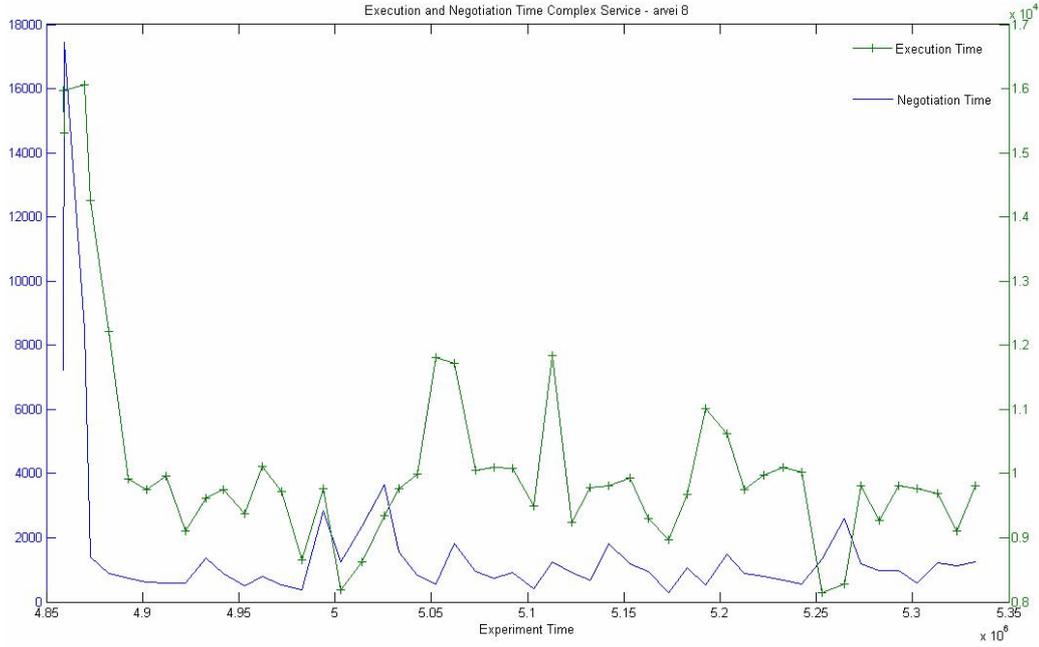


Figure 6: Experiment Results - Execution and Negotiation Time at node arvei8

The first aggregation level contemplates the construction of the normalized metrics set agent satisfaction, access time and resource usage. Agent satisfaction and resource usage are normalized to the interval between 0.0 and 1.0. The access time is normalized as follows:

$$access.time = e^{-\beta(execution.time+negotiation.time)}$$

where the weighting factor $\beta = 0.00005$ is chosen arbitrarily. The on demand availability is defined the sum of the benefit metrics access time and satisfaction and the infrastructure cost indicator equals the resource usage. Applying the mean and the standard deviation of both on demand availability and infrastructure costs (see Table 1), we obtain the final index like follows:

$$L = \alpha \left[1 - 1/2 (\mu_{access.time} + \mu_{agent.sat}) \right]^2 + \alpha (1/2)^2 (\sigma_{agent.sat}^2 + \sigma_{access.time}^2) + \beta (\mu_{resource.usage})^2 + \beta_{resource.usage}$$

<i>Metric</i>	<i>Value</i>	<i>Metric</i>	<i>Value</i>
L	0,43508	$\alpha \beta$	0,5
$\sigma_{resource.usage}$	0,32025	$\mu_{resource.usage}$	0,48686
$\sigma_{agent.sat}$	0,015615	$\mu_{agent.sat}$	0,46295
$\sigma_{access.time}$	0,013218	$\mu_{access.time}$	0,080522

Table 1: Experiment Results - Final Index

5 Conclusion

The design of a general evaluation framework for resource allocation in ALNs has been presented, which consists of a set of technical and economic parameters. The parameters in the lower levels of the framework provide technical data, while higher level economic parameters could be integrated into business models for decision makers.

However, the implementation and possible refinements of the presented concept is ongoing work which is not yet finalized. An exact definition about how to measure the defined metrics for the framework and integrate them is also ongoing work. A first implementation of the presented metrics framework has been discussed – work undertaken as part of the EU CATNETS project. This project develops a resource allocation architecture based on decentralized economic models, allowing resource management polices to be specified in a decentralized, autonomous and infrastructure independent way. The economic concepts of the CATNETS project uses bargaining mechanisms improving the resource allocation in Application Layer Networks (ALNs). An application example is used to demonstrate how metrics associated with the evaluation framework can be specified.

Acknowledgements

This work has partially been funded by the EU in the IST programme “Future and Emerging Technologies” under grant FP6-003769 “CATNETS”.

References

- [AdHu00] Adar, Eytan; Huberman, Bernardo A.: Free riding on gnutella. Technical report, Xerox PARC, 10 Aug. 2000.
- [BaGo83] Barro, Robert; Gordon David: Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics*, Vol. 12, No.1, pp. 101–122, 1983.
- [BuAG00] Buyya, R.; Abramson, J.; Giddy D.: Nimrod/G: An architecture for a resource management and scheduling system in a global computational grid. In *Proceedings of the 4th International Conference on High Performance Computing in Asia-Pacific Region (HPC Asia 2000)*, 2000.
- [CKKG99] Chapin, S. J.; Katramatos, D.; Karpovich, J.; Grimshaw, A.S.: The legion resource management system. In Dror G. Feitelson and Larry Rudolph, editor, *Job Scheduling Strategies for Parallel Processing*, pages 162–178. Springer Verlag, 1999.
- [CoHR95] Cobb, C.; Halstead, T.; Rowe J.: *The genuine progress indicator: Summary of data and methodology*. San Francisco: Redefining Progress, 1995.
- [EACC05] Eymann, T.; Ardaiz, O.; Catalano, M.; Chacin, P.; Chao, I.; Freitag, F.; Gallegati, M.; Giulioni, G.; Joita, L.; Navarro, L.; Neumann, D.; Rana, O. F.; Reinicke, M.; Schiaffino, R.; Schnizler, B.; Streitberger, W.; Veit, D.; Zini F.: *Catallaxy-based grid markets*. In *Proceedings of the First International Workshop on Smart Grid Technologies (SGT05)*, 2005.
- [ENRS06] Eymann, T.; Neumann, D.; Reinicke, M.; Schnizler, B.; Streitberger, W.; Veit D.: *On the Design of a Two-Tiered Grid Market Structure*. *Business Applications of P2P and Grid Computing*, MKWI, 2006.
- [EyPS00] Eymann, T.; Padovan, B.; Schoder, D.: *The catallaxy as a new paradigm for the design of information systems*. In *Proceedings of the 16th IFIP World Computer Congress, Conference on Intelligent Information Processing*, 2000.
- [FoKe97] I. Foster and C. Kesselman. *Globus: A metacomputing infrastructure toolkit*. *The International Journal of Supercomputer Applications and High Performance Computing*, 11(2):115–128, 1997.
- [Horn93] Horn, R.: *Statistical Indicators For the Economic & Social Sciences*. Cambridge University Press, 1993.
- [JPBG04] Joita, L.; Pahwa, J.S.; Burnap, A.; Gray, P.; Rana, O.F.; Miles, J.: *Supporting collaborative virtual organisations in the construction industry via the grid*. In *Proceedings of the UK e-Science All Hands Meeting 2004*, Nottingham, UK, 2004.
- [JRCC06] Joita, L.; Rana, O.; Chacin, P.; Chao, I.; Freitag, F.; Navarro, L. Ardaiz O.: *Application Deployment using Catallactic Grid Middleware*. In: *Proceedings of the 3rd International Workshop on*

Middleware for Grid Computing (MGC 2005), co-located with Middleware 2005, 2005, Grenoble, France.

- [MoBa03] Montresor, A.; Baboglu, O.: Biology-inspired approaches to peer-to-peer computing in bison. In Proceedings of the 3rd International Conference on Intelligent System Design and Applications (ISDA'03), Advances in Soft Computing, pages 515–522. Springer-Verlag, Aug. 2003.
- [PiZC00] Pinter, L.; Zahedi, K.; Cressman D.: Capacity Building for Integrated Environmental Assessment and Reporting. International Institute for Sustainable Development and United Nations Environment Programme, 2000.
- [Pool70] Poole, W.: Optimal choice of monetary policy instruments in a simple stochastic macromodel. Quarterly Journal of Economics, 84:197–216, 1970.
- [PrTo98] C. Preist, M. van Tol, Adaptive agents in a persistent shout double auction. In Proceedings of the First international Conference on information and Computation Economies (Charleston, South Carolina, United States, October 25 - 28, 1998). ICE '98. ACM Press, New York, NY, 11-18.
- [ReSE06] Reinicke, M.; Streitberger, W.; Eymann, T.: Evaluation of Service Selection Procedures in Service Oriented Computing Networks. In: Multi Agents and Grid Systems, Vol. 2, pp. 271-285, 2006.
- [UNDP99] UNDP. Human development report. New York: Oxford University Press, 1999.
- [YeBu04] C.S. Yeo, R. Buyya, Pricing for utility-driven resource management and allocation in clusters. Proceedings of the 12th International Conference on Advanced Computing and Communication (ADCOM 2004), Ahmedabad, India, December 2004. Allied Publisher: New Delhi, India; 32–41.

Service-Oriented Computing for Risk/Return Management

Wolfgang Hackenbroch, Matthias Henneberger

Competence Center IT & Financial Services

University of Augsburg

86135 Augsburg

{wolfgang.hackenbroch,matthias.henneberger}@wiwi.uni-augsburg.de

Abstract

Financial applications in the field of risk/return management demand for powerful and at the same time flexible information technology resources. These requirements seem to be highly suitable for service-oriented computing concepts. In this paper we identify promising risk/return management services and analyze the specific value proposition of service-oriented computing in this context. Taking the calculation of covariance matrices as an example we propose a model to quantify the economic value of such a service and thereby make a contribution towards understanding the business value of service-oriented computing. Moreover, our quantification model can be applied for the pricing of services and for the allocation of resources to services in a market-based environment.

1 Introduction

Risk/return management (RRM) is an important business function, especially for enterprises in the financial services industry because in an increasingly volatile and competitive market environment financial positions are affected by a high number of interacting economic factors that need to be permanently evaluated. To gain competitive advantage banks and other financial services firms apply sophisticated and complex models for the quantification of risk and the optimization of risk and return of portfolios containing all kinds of investment objects. These tasks are time-critical and therefore demand for powerful computational resources. Service-oriented computing based on grid technologies promises to overcome current restrictions by offering high-end computational capabilities at moderate cost and supporting new ways of intra- and inter-enterprise collaboration.

Although the financial services sector is often mentioned as a key industry for service-oriented computing based on grid technologies, see e.g. [Frie03; Midd04], there is a lack of publications concerning applications in this area and important questions are still to be answered. In this paper we are striving to develop an understanding for the applicability of the general concept of service-oriented computing in the specific area of RRM by answering the following research questions: i) Which business functions and concrete methodologies are promising for the application of service-oriented computing in the area of RRM? What is the specific value proposition of service-oriented computing in this domain? ii) How can benefits and costs of service-oriented computing be quantified? For our analysis we will identify and discuss grid services in the domain of RRM from a management perspective. To answer the second research question we will concentrate on one concrete service – the estimation of covariance matrices – and propose an economic model for the quantification of benefits and costs. Based on this example we will discuss how economic models can be applied in the context of service and resource markets.

The remainder of this text is structured as follows: In section 2 we describe our notion of RRM and provide fundamental principles. Section 3 is dedicated to the concept of service orientation and derives the value proposition of service-oriented computing in the domain of RRM. In section 4 we provide a quantification of benefits and costs for the covariance estimation service.

2 Risk/Return Management

In this section we will first provide basic principles of RRM. We will then describe specific aspects that prove to be complex and resource-intensive in practice.

2.1 Definition and Objectives

Almost all value generating business transactions are associated with some kind of uncertainty and thus contribute to an enterprise's overall risk exposure. Enterprises currently face rising pressure from competition on a global scale, delivered on integrated markets (especially integrated financial markets) where more and more complex products are traded on. Hence, in corporate practice an integrated view on risk and return is mandatory. Considering this integrated view, we propose the following definition: RRM is concerned with identification, quantification and control of risk together with the corresponding return associated with all kinds of corporate investment decisions. It is necessary to point out that accordingly, the scope of RRM is by far

larger than risk management alone. It covers all business functions that are concerned with investment decisions and therefore need to evaluate the risk and return of the available alternatives. Therefore, for example, also portfolio management can be seen as a part of an integrated RRM.

It is an important objective of RRM to satisfy the manifold information demands inside an enterprise. Ex ante decision support for the planning of investment or disinvestment decisions has to be provided, answering the question whether (and to what degree) an intended investment or disinvestment will lead to value added for the enterprise. To this end, accurate and up-to-date risk and return measures have to be quantified on the transactional level as well as on various aggregation levels along the organizational hierarchies like departments or business units. Especially for financial services institutions with their typically high exposure to market and credit risks this can be considered as crucial.

Prominent side conditions of RRM are the growing number of laws, rules and regulations aimed at the prevention of illiquidity and bankruptcy by restricting potential losses resulting from risky investment objects. In the financial services industry, regulations are especially tight (e.g. as in the Basel II or Solvency II accord or in the Sarbanes-Oxley Act) and financial institutions are required to abide by certain well-defined risk limits, see e.g. [Schi97, 505]. They keep significant capital reserves in order to retain their solvency even in extreme market situations. These capital reserves then restrict the overall earnings in terms of return from risky investments. Therefore it is an essential task for an integrated RRM to determine a suitable capital allocation against the background of risks being taken.

2.2 Complexity of Risk/Return Management Methods

RRM is a time-critical and resource-intensive¹ undertaking. It is time-critical, because regulatory constraints as well as internal information needs have to be fulfilled in certain, well-defined time frames, ensuring timeliness and relevance of the information delivered. It is resource-intensive, because there is a general trade-off between timeliness and precision of the underlying methods: Usually it is only possible – ceteris paribus – to improve in one dimension at the expense of the other. Excellence in both dimensions can only be achieved employing powerful resources. In the following we will consider fundamental and resource-intensive procedures in the area of RRM.

¹ Here in terms of information technology (IT) resources (see subsection 3).

It is well known that the *estimation of distribution parameters* for risky investment objects is a fundamental problem in RRM, which is in fact often cited as “one of the important problems in finance” [ElGr72, 409]. Many models of portfolio theory or capital market theory require for their practical application the estimation of input parameters, based on empirical values drawing on a multitude of historical data (the “history”). For example, Markowitz’ Portfolio Selection model relies on expectations and covariances of the considered investment objects’ returns in order to select an investor’s optimal portfolio. Since the “true” return distribution parameters are unknown, this constitutes in fact an estimation problem. In this context, higher estimation precision can be achieved only with higher effort, be it by considering longer estimation time intervals or by implementing more complex computation procedures. We refer to e.g. [HuWh98; Kupi95] and the literature referenced therein. [Alex96, 233] provides an overview of techniques used in covariance estimation.

A more specific task is the *estimation of covariance matrices*, containing all pairwise covariances between the investment objects’ returns. In fact, covariances are used for risk quantification and aggregation in almost all areas of RRM, thus constituting a fundamental building block for many financial applications. Depending on the (possibly high) number of investment objects, these matrices typically contain a large count of covariances which all have to be individually estimated. Namely, we have for n investment objects $n(n + 1)/2$ covariances, i.e. the number of matrix elements is increasing with approximately $n^2/2$. Assuming a history with length $N=250$ working days per year for each investment object, we have for $n=1.000$ investment objects already approximately $2Nn^2/2=Nn^2=250.000.000$ historical values to take into account. It becomes apparent that, depending on the estimation method used, especially the estimation of covariance matrices is a very resource and time intensive problem.

A wide-spread method of RRM to deal with uncertain future values is *historical or stochastic simulation*. The latter is often also called “Monte-Carlo simulation”. In this context both forms have in common, that for a possibly large number of investment objects first the empirical return matrix has to be determined. For m so called “market factors” with N empirical (historical) prices each this is a $m \times N$ matrix. In the case of historical simulation one assumes that future values are independent and identically distributed conforming to the distribution of a representative historical sample. Stochastic simulation closely resembles this approach with the difference that future values are determined stochastically on the basis of an assumed distribution. In order to reduce calculation effort, most often the normal distribution is applied [Völk01, 76]. When a

simulation approach is chosen, one usually relies on the “law of large numbers”, i.e. a very large number of simulation runs has to be performed in order to obtain reliable and statistically significant results. For instance, for $m=100$ market factors with a history of one year each, the empirical return matrix with daily returns contains $mN=25.000$ elements (assuming $N=250$ working days per year). Performing e.g. 100.000 simulation runs this already leads to 2.500.000.000 values whose processing requires not only a lot of storage space but also significant computation capacity. We will point out in the following section how these requirements are addressed by service-oriented computing.

3 Service-Oriented Computing in Risk/Return Management

In academic research as well as in corporate practice a new paradigm for the utilization of information technology resources is en vogue. We use the term “service-oriented computing” in this text to address several (in fact converging) aspects: Elements from different domains like service-oriented architectures (abstract concept of services), grid technologies (grid middleware) or distributed computing (load balancing, scheduling) contribute to enable and facilitate the service-oriented computing principle. Services as well-defined software components are the pivot element of service-oriented architectures, whereas virtualization of infrastructure and resource sharing are primarily in the focus of approaches related to grid technologies or distributed computing. In our view the combination of both aspects sets the frame for service-oriented computing.

3.1 From Service-Oriented Architectures to Grid Services

According to [DJMZ05, 7] we define service-oriented architectures (SOA) as the “abstract concept of a software architecture with focus on providing, searching and consuming so-called services over a network”. Before we confine our notion of service-oriented computing, it is necessary to characterize two central terms in this context: resource and service. Following [NeHO06, 206] we regard a resource as the representation of a logical or physical entity (like computing or data capacity, software licenses, hardware or network infrastructure). A service on the other hand provides a specific functionality and therefore aggregates the use of different (and in the context of service-oriented computing most often distributed) resources. Accordingly, a service in this sense is a software component designed to enable or support (part of) an

enterprise's business process. It makes sense to distinguish between two types of services: One speaks of "complex" or "composite" services on a high aggregation level in comparison to "basic" or "atomic" services on a low aggregation level as e.g. in [ENRS06, 44; Doda06, 12]. It is characteristic for SOA that services are "loosely coupled" with each other, implying that they can be reused in a manifold way and that business processes can be dynamically configured with suitable, "best-of-breed" services. We will follow this perspective of "coarse granularity" and will provide examples for both service types offered in a SOA that can be used to "orchestrate" business processes, here in the context of RRM.

Regarding the related concept of grid computing, interestingly enough, the available definitions are mostly of descriptive nature and provide little more than certain essential characteristics. They are in most cases based on the seminal papers of [FoKe98; FoKT01]. Various proponents have thereupon agreed that "a computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive and inexpensive access to high-end computational capabilities". It is an important characteristic, that grid computing reaches beyond administrative domains and thereby defines the so-called "virtual organization" as proposed by [FoKT01]. For grid technologies, an evolution towards service-oriented architectures can be observed [Long04]. We therefore speak of "grid services" in this text, denoting software components that realize resource intensive services based on grid technology. This term also conveys the close relationship between grid services and web services. In fact, grid services are based on specific web service standards, e.g. the specifications (OGSA) and (WSRF). They extend web services insofar as they imply the dynamic, yet for the user transparent, allocation of (physical) resources to services by some kind of a grid middleware.

3.2 Services for Risk/Return Management

In this section we will provide some examples for grid services for RRM and classify them into complex and basic services. From a business perspective, at least two types of services can be distinguished: services for risk quantification and services (using risk/return information) for the management of risk and return, e.g. in the form of portfolio management.

A risk quantification service delivers one or several risk measures on different aggregation levels ranging from single investment objects up to the whole enterprise. Which measure should be applied depends on the type of risk (e.g. calculating credit risk differs significantly from calculating market risk) and on the objectives pursued. In corporate practice the probably most wide-

spread risk measure is Value-at-Risk (VaR)², mainly because of its ease-of-use and simplicity. Furthermore, regulatory constraints require a frequent calculation of VaR on the enterprise level. Besides VaR, there exist a variety of other risk measures like variance or down-side risk measures, e.g. semi-variance. In a service-oriented environment a collection of services is conceivable, each calculating a specific risk measure. For trading securities decision support systems not only have to deliver precise up-to-date estimations of market parameters but also need to provide sophisticated optimization algorithms. In a service-oriented environment this could be achieved by a *portfolio optimization service*, after the necessary input parameters, e.g. variances and covariances are determined. This essentially implies the determination of the efficient frontier, which is a NP-hard problem. Nevertheless, a numerical solution can be found for smaller problems using quadratic programming or applying heuristic approaches. The required precision of these heuristics can only be achieved with significant computational effort. Especially on the trading floor it is furthermore mandatory, that information is available in real-time so that traders can quickly analyze different scenarios and react accordingly.

As already stated in section 2, most business functions in RRM rely on fundamental methods like parameter estimation or simulation that could as well be realized as separated services and seized by complex services. For example, VaR can either be deduced analytically from the variances and covariances or can be determined using historical or Monte-Carlo simulation. Concrete examples for *basic services for parameter estimation and simulation* may be a historical data service, delivering e.g. an empirical return matrix consisting of the historical returns of the relevant investment objects, a regression analysis service, a stochastic or historical simulation service, a covariance matrix estimation service or a time series generation service (e.g. on the basis of ARCH or GARCH models). In section 4 we will describe one selected basic service – the covariance matrix estimation service – in more detail, also evaluating its application for the overall risk quantification of the enterprise.

3.3 Value Proposition of Service-Oriented Computing

The relevant, yet basic, value proposition of service-oriented computing concepts is almost canonical and to date already numerously stated, see e.g. [Abba04; BeFH03; FoKe98]. It delivers on demand computing power at transparent and relatively low cost (in comparison to dedicated, server-based computing) in combination with increased flexibility, scalability and a robust be-

² According to [DuPa97, 3] VaR is the “loss in market value over a time horizon H that is exceeded with probability $1-\alpha$ ”.

behaviour against failure. This is achieved by using existing and/or standardized resources which are geographically and/or logically distributed in more or less autonomous units provoking high percentage utilization. In this context, [NeHO06, 206] distinguish between the possibility to perform computational intensive jobs in a reduced time frame, even enabling the solution of problem classes which so far could not be examined, and the possibility to share (not fully utilized) resources across different organizations. Yet, service-oriented computing is not primarily concerned with the utilization of free CPU cycles (which has been scrutinized for about 10 years by the disciplines of cluster computing and distributed computing). Instead, it provides new capabilities for intra- and inter-enterprise collaboration, enabling for example [Grau06, 71] world-wide communication and collaboration, access to high-performance computers for simulation or utilization of remote data sources. In the following we will elaborate on the basis of subsection 2 in more detail the specific value proposition of service-oriented computing in the context of RRM:

1. Because of computational constraints, RRM (especially the quantification of the overall risk position and adjustment of the capital allocation) to date is performed in relatively large and fixed time intervals. For instance, some enterprises are accumulating risk reports containing all relevant risk (and return) information only once every quarter or once a month, as reported by practitioners like e.g. [Midd04]. In other cases regulatory requirements demand that new information is available the next morning (after overnight batch runs) regarding the enterprise's risk position of the day before, as e.g. in [Base04]. Whereas this might be sufficient in special cases, it is considered a disadvantage in general because investment decisions then are based upon outdated information. We figure that with service-oriented computing the necessary underlying calculations can possibly be dramatically accelerated without the need to invest into additional cost-intensive computing infrastructure. In this context we have the following scenario in mind: Because of integrated and possibly highly volatile markets, an integrated and up-to-date RRM is mandatory. It implies "sensitive behaviour" regarding expected and unexpected market reactions as well as real-time determination of the current overall risk position. RRM takes place (not in fixed time intervals but) dynamically and "event-driven", in much shorter time intervals. In fact we envision the "real-time enterprise" which is able to react appropriately almost immediately to all relevant and/or unforeseen market movements because it has timely and accurate risk/return information at its disposal. As soon as one calculation run is completed, immediately

the next calculation starts. Depending on the current capacity of IT involved and the complexity of computation procedures the RRM time intervals are varying over time.

2. The input data needed for parameter estimation and forecasting in RRM is typically geographically and/or logically distributed. For instance, information about the portfolio of a globally investing enterprise might be scattered over several trading units in different locations. A common infrastructure based on grid services standards not only allows sharing the data across organizational boundaries. The distribution of data also matches the fundamental structure of service-oriented computing and thus can be exploited for distributed processing where the data is available. In this case no centralization of data (causing communication and management complexity) is necessary.

3. There is a trend towards higher frequency of input data first observed and published by [Engl00]. Today, for example stock market data is widely available in a granularity down to the transactional level. The analysis of this “Ultra-High-Frequency” market data is a promising new area with implications for risk management not yet fully discovered. Service-oriented computing can contribute its share to storing and processing this huge amount of data providing up-to-date information and comprehensive analyses.

4. Because of the permanent movement and development of (financial) markets the demand for RRM calculations is itself far from constant. Using a dedicated infrastructure the enterprise is therefore committed to provide at any time a computing capacity aligned to the maximum demand during peak times. Additionally there is always the trade-off between RRM and other operations which have to be performed by the resources at hand. This challenge is met when unused resources can be seized at any time for additional speed and/or accuracy and in turn are available for daily operations in “quiet times”.

5. There exists a variety of RRM applications that do not by all means require high-performance computing power at any given time. In contrast to the typical batch-processing mode they can possibly be executed cost-effectively on a service-oriented infrastructure in the background. With the available (and varying) computing capacity they can be performed continuously, including “slow business hours” like e.g. overnight time, and are still incessantly adding value for the enterprise. The key point here is the automatic allocation of resources whenever they are available, increasing flexibility and manageability of existing corporate infrastructure.

6. Grid services offer new possibilities for intra- and inter-organizational collaboration regarding the integration, coordination and usage of resources and services. Concerning the inter-

organizational provision of services and resources we need to clearly distinguish between high-level services that realize specific business functions (e.g. the quantification of risk) hiding the underlying grid-based calculations from the end-user and low-level services providing only the necessary computational capabilities. As [GRCB06, 4] point out, especially for small and medium enterprises low level services are generally not satisfactory as these institutions most often lack the expertise necessary to implement complete RRM algorithms. Nevertheless we argue that in the financial services industry there is room for both approaches: For large financial services corporations that face fluctuating demand for computing capacity as discussed above, it may be interesting to contract additional computing capacity (on demand) from an external provider. Standardized solutions are not appropriate in this case as these institutions precisely try to get a competitive advantage on financial markets by applying proprietary methods and algorithms. Providers like SUN already offer the possibility to use their grid network to run resource intensive calculations on a pay-per-use basis. On the other hand small and medium enterprises demand for standardized high-level services in the area of portfolio management and risk quantification. Financial software or data suppliers like Reuters, Bloomberg and RiskMetrics already offer internet-based services e.g. for the calculation of portfolio risk, see for instance [Risk06]. Summarizing, we can state that service-oriented computing provides a highly suitable basis for the implementation of corporate RRM services. Exploiting the characteristic properties of grid services (like virtualized hardware at geographically distinct locations) an enterprise can benefit from more efficient risk/return calculations.

4 The Covariance Estimation Service

As pointed out above, covariances are an essential prerequisite for all kinds of financial risk calculations. In the following section we will consider a covariance estimation service that provides its user transparently with up-to-date covariance data for the relevant investment universe and propose an economic model for the quantification of benefits and costs that can be attributed to this service.

4.1 Economic Value of a Covariance Estimation Service

We consider an enterprise that frequently recalculates its risk position and therefore employs a service responsible for the estimation of covariance matrices. To this end we assume that the

only purpose of covariance estimation is risk quantification. Thus the service's economic value can be directly derived from the benefits of risk quantification alone. We perceive the enterprise as the weighted "sum" of its investment objects, i.e. its overall risk position is expressed as the portfolio risk, measured by the variance σ^2 of portfolio returns. We can calculate portfolio risk, resulting from n investment objects (numbered from 1 to n), using the covariance matrix, as $\sigma^2 = \sum_{i=1}^n \sum_{j=1}^n Cov_{ij}$ with Cov_{ij} denoting denoting the weighted covariance between investment objects i and j .

Since the enterprise is acting in an uncertain and dynamic environment its risk position is changing willingly (by making investment decisions) or unwillingly (by "movement" of the underlying markets). Because the estimation of covariance matrices is time-consuming, the covariances at hand are always significantly outdated and therefore the variance calculated from the covariance matrix does not reflect the current risk position. We are in the following recurring to the fact that enterprises are adjusting their risk position to a value somewhere below a certain threshold thus constituting a "safety margin" (in the regulatory context often also called "haircut", see e.g. [Base04, 29]). They are doing so by using the capital allocation between risky and risk-free investment objects for balancing their overall risk position.³ Our basic modelling approach is in the following: whenever covariances are available the safety margin can be adjusted immediately in a way that the resulting (and over time changing) overall risk position of the enterprise with high probability does not exceed the given risk limit at any time.

The estimation of one covariance matrix is assumed to take exactly T periods. The estimation of a new covariance matrix begins immediately after finishing the previous matrix, thus the service can provide a complete covariance matrix every kT periods ($k \in \{1,2,3,\dots\}$). According to [BFHH05, 7], the benefits B of risk quantification subject to T can then be calculated as⁴

$$B(T) = \left(i + \frac{\bar{\sigma}(\mu - i)}{q_\alpha \sigma_\sigma \sqrt{2T}} \right) \cdot K \quad (1)$$

The parameters of the model are defined as follows. $K > 0$ denotes the enterprise's total capital that is always completely allocated to a portfolio containing risky investment objects and/or a risk-free alternative. The portion of the enterprise's capital allocated to the risky portfolio yields

³ At this point it is important to understand that the model presented here is not addressing the evaluation of the efficient set of investment objects or portfolio optimization (both problems also require the calculation of covariances), but the aggregation and management of the risk position of an enterprise.

⁴ This text contains a shortened version of our model. The complete exposition including the assumptions of the model can be found in the working paper [BFHH05, 7] of the same authors.

the expected return μ , whereas the risk-free investment pays the time-invariant risk-free interest rate i . As we assume that always $\mu > i > 0$, the enterprise would fully allocate its capital to risky investment objects. Yet the enterprise is required to abide by the given risk limit $\bar{\sigma} > 0$ and therefore needs to invest a certain share of the available capital into the risk-free alternative, depending on the intended safety margin. The safety margin itself is not modelled explicitly as a separate parameter, but results from the volatility of portfolio risk (denoted as σ_σ) and additionally from the frequency of the estimation, as higher frequency implies a more accurate determination of the current risk position. Consequently, the faster covariance estimations are accomplished (reflected by a smaller T), the smaller the safety margin can be, the less capital needs to be invested into the risk-free alternative and, ultimately, the higher are the overall benefits of the enterprise.

4.2 Value Proposition for the Covariance Estimation Service

We can now analyze our model with respect to the value proposition developed in section 3 and confine our findings for the special case of covariance matrix estimations. The first value proposition we identified was the possibility to generate additional value by accelerating RRM calculations because high-end computing capabilities are available for relatively moderate cost. It is difficult to quantify costs of our grid service. They may comprise (varying) costs of physical resources as well as costs for the implementation, management and maintenance of the service itself, including e.g. user support by business experts. In the following we are only considering computational resources as a cost driver, thus leaving the (end-user) perspective of a coarse granular service. More precisely, we assume costs to be proportional subject to the computing capacity needed for the estimation, reflected by a factor price p . Additionally, we restrict ourselves to the consideration of the time needed for computation, neglecting e.g. latency or transmission times. It is reasonable to assume that the so defined costs are moderate compared to server-based computing: In the case of covariance estimations the corresponding computations can be distributed on several resources, as all pair wise covariances can be calculated independently from each other. Efficiency losses are low and cost advantages actually take effect, as a number of low cost standardized components can provide the same capacity for covariance calculations than an expensive “traditional” server. Moreover, higher utilization levels can be expected due to on-demand allocation of resources. To analyze the effect of a decrease in costs on the overall benefits we first consider the functional relationship between T and the comput-

ing capacity $z > 0$ necessary to complete the covariance matrix within the time frame T . With $n(n+1)/2$ covariances and w denoting the workload per covariance we have

$$T(z) = \left(\frac{n(n+1)w}{2z} \right). \quad (2)$$

A larger z results in a smaller calculation time frame T which in turn leads to increasing benefits. Together with the cost side, expressed by pz , we can formulate the objective function Z (with decision variable z) as the difference of benefits and costs on the basis of equations (1) and (2) as⁵

$$Z(z) := Ki + \frac{K\bar{\sigma}(\mu - i)\sqrt{z}}{q_\alpha \sigma_\sigma \sqrt{n(n+1)w}} - pz \rightarrow \max! \quad (3)$$

Applying the standard optimization procedure (i.e. solving $Z'(z) = 0$ for z) delivers as a distinct solution

$$z^* = \frac{K^2 (\mu - i)^2 \bar{\sigma}^2}{4n(n+1)p^2 w q_\alpha^2 \sigma_\sigma^2} \quad (4)$$

Since $Z''(z) < 0 \forall z$, the so defined z^* is a global maximum of the objective function. It also determines – using equation (2) – the corresponding time frame T^* our service should comply with. Lower costs in equation (4) are reflected by a smaller price p leading to an increasing optimal capacity z^* and a shorter calculation time T^* . Thus it is economically reasonable to allocate more capacity, resulting in increasing overall benefits. More precisely, as the computing capacity depends quadratically on the price p , even a small decrease of p has high impact on capacity and calculation time. We therefore expect covariance calculations to be performed more frequently (even real-time), allowing enterprises to better exploit risk limits.

We also observed in section 3 that service-oriented computing is advantageous concerning the processing of high data volumes. This constitutes an important aspect in the case of covariance matrices as well: First, the number of input values for the calculation is substantial as pointed out in section 2. Second, these values are usually distributed within the enterprise, e.g. geographically distributed according to different financial markets where, third, necessary market data is often provided by different information providers, e.g. Reuters or Bloomberg. A service-oriented infrastructure based on common standards facilitates the integration of data across or-

⁵ The objective function is the result of an analytically necessary and numerically justifiable approximation (see the working paper [BFHH05, 7] for a complete deduction).

ganizational boundaries. Covariances could be calculated independently, e.g. on portfolio level, and then be aggregated as needed.

With our model we can also analyze the effect of changing (market) conditions on the demand for computing capacity (see point 4 and 5 in section 3). As can be observed by equation (4), the demand for computing capacity is influenced by market- or enterprise-specific parameters. For example, the more capital the enterprise has to its disposal the more (in absolute terms) it will invest into risky investment objects. Higher risk exposure in turn increases the importance of RRM which is correctly reflected by a larger value for z^* . The same argumentation holds when the enterprise faces a higher risk limit $\bar{\sigma}$. In this case it should allocate more resources to RRM applications, which is consistently leading to an increasing z^* in our model. Eventually, when the risk premium ($\mu-i$) rises, investing into risky objects becomes more attractive and profitable, resulting in a larger share of risky capital. In order to manage the consequently more voluminous portfolio our model proposes that additional computing capacity is necessary. In contrast to server-based concepts where resource allocation is fixed (at least for a certain time interval), with service-oriented computing resource allocation can be adjusted dynamically, realizing an optimal allocation at any point in time.

Regarding the possibility of service-oriented infrastructures to provide services across organizational boundaries we argued in section 3, that especially standardized services in the area of risk quantification are well suited. Accordingly we envision a covariance estimation service that can be accessed via a service-oriented environment. As observed by [GRCB06, 8] for the manufacturing industry, such a service-oriented environment needs to fulfill several other requirements, e.g. regarding security, accounting and billing or end-user access.

4.3 Model Application

One important characteristic of service-oriented computing is the possibility to share services and resources within and across organizational boundaries. Especially for resource sharing, grid computing concepts propose the application of market mechanisms to ensure an efficient allocation of available resources, see e.g. [WPBB01; BAGS02]. Yet a direct mapping of complex services to resources is not reasonable, as service consumers in general do not have concrete knowledge of the resources necessary to solve the problem at hand. To this end, [ENRS06, 44] propose a two-tiered market. The set of services, together with means for provisioning and pricing, constitute the “service market”. At the same time services demand for different kinds of

resources, which are provided on the corresponding “resource market”. Basic services are responsible for purchasing resources on the resource market.

This is an important application of economic models like the one we described here because for the pricing of services as well as for resource allocation it is necessary to quantify the economic value associated with service and resource consumption. In fact our model perfectly fits in this scenario: Covariance estimation can be regarded as a basic service whose benefits are derived when it is deployed for risk quantification, which may in turn be accomplished by a complex service. On the other hand the covariance estimation service needs to seize physical resources on the resource market to perform its calculations. Here, the model can be employed for resource allocation issues on the resource market. Regarding the pricing of a covariance estimation service it delivers the economic value of covariance estimations depending on the calculation frequency. This can be considered as a quality attribute of the service being the subject of service level contracts. For instance, when the service is provided externally, a concrete calculation time may be contracted. Considering the benefits as an upper bound (an enterprise would not pay more than its benefits), the price of the service is set depending on this calculation time.

5 Conclusion

In this paper we addressed two research questions concerning service-oriented computing and its applications in the field of RRM. First, we argued why complex and resource-intensive RRM calculations are highly suitable to be performed on service-oriented infrastructures: Their specific properties and structure from our perspective almost ideally match with service-oriented infrastructure. We identified a set of potential services off the beaten track, like e.g. the estimation of covariances, to illustrate this close relationship. Second, we provided a model addressing the economic aspects of service-oriented computing.

We moreover argued that our model can be applied in a market-oriented scenario where benefits and costs have to be evaluated. In fact, this in our opinion constitutes a prerequisite for the further development of adequate market-oriented approaches. For instance, auction mechanisms appear to be well-suited to ensure an economically efficient allocation of services and resources on the respective markets [ENRS06, 44]. We are currently working on the transformation of the model results for representing the valuation attributed to covariance estimation by market participants using a bidding language. Thereby we enable the regulation of access to services and

resources depending on the individual priority, measured by the reservation price, ascribed by the service consumer, see [NeHO06, 2007]. Apart from such market-oriented approaches there are open issues that need to be considered. With the exemplary list of services established in this text we intend to foster the development of service-oriented computing in the domain of RRM. Future work is planned to provide the proof-of-concept for basic services (parameter estimation or simulation) as well as for more complex RRM services. Here, problems concerning the technical implementation or security issues still need to be addressed. Additionally, economic effects and new business models resulting from the adoption of service-oriented computing in the area of RRM need to be analyzed. The availability of business applications, e.g. for RRM, is a critical success factor for the wide adoption and further development of service-oriented computing. From our perspective, there is plenty of room for further research in this new area, intertwining some of the most interesting ideas at the intersection of information systems and business domains.

References

- [Abba04] Abbas, Ahmar: Grid Computing: A Practical Guide to Technology and Applications. Charles River Media, Hingham 2004.
- [Alex96] Carol Alexander: Volatility and Correlation Forecasting. In: Carol Alexander (Hrsg.): The Handbook of Risk Management and Analysis. John Wiley & Sons Ltd., West Sussex 1996, S. 233-260.
- [Base04] Basel Committee on Banking Supervision: International convergence of capital measurement and capital standards. http://www.riskglossary.com/articles/basle_committee.htm, 2005, accessed on: 2006-10-11.
- [BeFH03] Berman, Fran; Fox, Geoffrey C.; Hey Anthony J. G. (Hrsg.): Grid Computing; making the global infrastructure a reality. John Wiley & Sons Ltd., West Sussex 2003.
- [BFHH05] Buhl, Hans U.; Fridgen, Gilbert; Hackenbroch, Wolfgang; Henneberger, Matthias: An Economic Model for the Allocation of Grid Resources to Risk/Return Management. <http://www.wi-if.de>, 2005, accessed on: 2006-10-11.

- [BAGS02] Buyya, Rajkumar; Abramson, David; Giddy, Jonathan; Stockinger, Heinz: Economic Models for Resource Management and Scheduling in Grid Computing. In: The Journal of Concurrency and Computation: Practice and Experience 14 (2002) Special Issue on Grid Computing Environments, S. 1507-1542.
- [Doda06] Dodani, Mahesh: SOA in an enterprise architecture. <http://www-5.ibm.com/ch/events/soaas/>, 2006, accessed on: 2006-10-11.
- [DJMZ05] Dostal, Wolfgang; Jeckle, Mario; Melzer, Ingo; Zengler, Barbara: Serviceorientierte Architekturen mit Web Services. Elsevier Spektrum Akademischer Verlag, 2005.
- [DuPa97] Duffie, Darrel; Pan, Jun: An overview of value at risk. In: Journal of Derivatives (1997) 4 S. 7-49.
- [ElGr72] Elton, Edwin J.; Gruber, Martin J.: Earnings Estimates and the Accuracy of Expectational Data. In: Management Science 18 (1972) 8, S. B409-B424.
- [Engl00] Engle, Robert F.: The Econometrics of Ultra High Frequency Data. In: Econometrica 68 (2000) 1, S. 1-22.
- [ENRS06] Eymann, Torsten; Neumann, Dirk; Reinicke, Michael; Schnizler, Björn; Streitberger, Werner; Veit, Daniel: On the Design of a Two-Tiered Grid Market Structure. In: Lehner, Franz; Nösekabel, Holger; Kleinschmidt, Peter (Hrsg.): Multikonferenz Wirtschaftsinformatik 2006 – Creating and Improving Business with Information Technology. Passau 2006, S. 41-52.
- [FoKe98] Foster, Ian; Kesselman, Carl (Hrsg.): The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann Publishers, San Francisco 1998.
- [FoKT01] Foster, Ian; Kesselman, Carl; Tuecke, Steven: The Anatomy of the Grid – Enabling Scalable Virtual Organizations. In: International Journal of High Performance Computing Applications 15 (2001) 3, S. 220-222.
- [Frie03] Friedman, Matthew: Grid Computing: accelerating the search for revenue and profit for Financial Markets. In: Building an Edge (2003), S. 143-145.

- [Gru06] Grauer, Manfred: IT aus der Steckdose und Wirtschaftsinformatik. In: Wirtschaftsinformatik 48 (2001) 1, S. 71-72.
- [GRCB06] Grauer, Manfred; Reichwald, Julian; Christian, Dennis; Barth, Thomas: The potential of service oriented computing for small and medium enterprises – analysis of value chains. In: Lehner, Franz; Nösekabel, Holger; Kleinschmidt, Peter (Hrsg.): Multikonferenz Wirtschaftsinformatik 2006 – Creating and Improving Business with Information Technology. Passau 2006.
- [HuWh98] Hull, John; White, Alan: Incorporating Volatility Updating into the historical Simulation Method for Value-at-Risk. In: Journal of Risk 1 (1998) 1, S. 5-19.
- [Kupi95] Kupiec, Paul H.: Techniques for Verifying the Accuracy of Risk Management Models. In: Journal of Derivatives 3 (1995), S. 73-84.
- [Long04] Longworth, David: Grid lock-in on route to SOA. <http://www.looselycoupled.com/stories/2004/grid-lock-infr0831.html>, 2004, accessed on: 2006-07-21.
- [Midd04] Middlemiss, Jim: Gearing up for Grid. <http://www.wallstreetandtech.com/showArticle.jhtml?articleID=17603501>, 2004, accessed on: 2006-10-11.
- [NeHO06] Neumann, Dirk; Holtmann, Carsten; Orwat, Carsten: Grid-Economics. In: Wirtschaftsinformatik 48 (2006) 3, S. 206-209.
- [Risk06] RiskMetrics Group: Managed Services. <http://www.riskmetrics.com>, 2006, accessed on: 2006-07-21.
- [Schi97] Schierenbeck Henner: Ertragsorientiertes Bankmanagement – Risiko-Controlling und Bilanzstruktur-Management. 5. Aufl., Gabler, Wiesbaden 1997.
- [Völk01] Völker, Jörg: Value-at-Risk-Modelle in Banken. Berlin Verlag, Berlin 2001.
- [WPBB01] Wolski, Rich; Plank, James S.; Brevik, John; Bryan, Todd: Analyzing Market-Based Resource Allocation Strategies for the Computational Grid. In: International Journal of High Performance Computing Applications 15 (2001) 3, S. 258-281.

Native Code Security for Java Grid Services

Thomas Frieze

Siemens AG, Corporate Technology, Information and Communications,
Otto-Hahn-Ring 6, D-81539 Munich, Germany
E-Mail: th.frieze@siemens.com

Matthew Smith, Bernd Freisleben

Dept. of Mathematics and Computer Science,
University of Marburg, D-35032 Marburg, Germany
E-Mail: {matthew, freisleb}@informatik.uni-marburg.de

Abstract

In modern on demand grid computing scenarios, services from different organisations will potentially run on the same web service engine of a grid node. Secure isolation of data and code of different service instances is a vital requirement in such an environment, since mutual trust cannot be assumed between all involved parties. For Java based Grid applications the Java virtual machine offers sandboxing facilities, however the common occurrence of native code (e.g. C/C++, Fortran) in business and scientific Grid applications leads to a number of security issues which are not handled by the basic Java security mechanisms. In this paper, we analyze the threat scenarios that emanate from native code in a service-oriented Grid scenario. A novel security architecture is presented, which enables a fine grained confinement of native components of Grid applications into a secure environment for protecting the hosting system as well as other service instances. Although our work focuses on Grid services, it is also relevant for any hosting scenario in which multiple web services using native code components are deployed in the same service container.

1 Introduction

The service-oriented architecture (SOA) and especially the web service paradigm have been adopted in various fields of distributed systems development. With the advent of the service-oriented Grid computing paradigm (i.e. the Web Service Resource Framework - WSRF), the

web service standards have also been applied in the field of scientific and high performance computing, extending their adoption beyond the fields of enterprise application integration or general business computing.

Java based web service containers have been widely accepted for the provisioning of web services. Services become small self-contained units in this scenario. They encapsulate the business logic represented in the web service, the services use the infrastructure services provided by the service container in which they are deployed.

In previous work [1] the concept of an ad hoc Grid as a possible environment for on demand computing was introduced. A solution to the problem of hot service deployment, which is one of the basic requirements for such a flexible Grid computing environment [2] was presented. The fact that services can be deployed to a number of nodes and migrate freely between them during an applications lifecycle leads to a number of security issues that need to be dealt with. The security domain is divided in two parts related to pure Java implementations of Grid services and non-Java native code.

Current work on security issues in WSRF focuses on the enforcement of access restrictions and protection of message exchanges in transit. Implementations of the WSRF specifications do not address issues concerning intra-engine service security, since providing such mechanisms is not enforced or encouraged. Therefore, it is possible for various service implementations to directly access each other through simple method calls, bypassing the service security mechanisms established for access control. Our work on hot service deployment raised some of the security issues and attack scenarios that we have addressed and solved for pure Java services [3].

Apart from the Java based implementation of web service containers, the .NET environment offers the functionality to provide web services. For service-oriented Grid computing, the dominant middleware frameworks (e.g. Globus Toolkit 4 and Unicore GS) have chosen Java based implementations. Thus, in this paper we focus on a Java based solution.

Many business and scientific applications in Grid computing are based on legacy code bases which cannot be ported to Java for cost and efficiency reasons. The existing legacy solutions are usually wrapped with Java Grid service implementations to make them available to the service-oriented Grid environment. This creates a major security problem since the native code cannot be constrained by the standard Java security facilities.

In this paper, we analyze the threat model induced to the service-oriented Grid by native legacy code and present a solution for imposing fine grained security enforcement on such native

service components. Our solution allows to host different Grid services containing native code in the same hosting environment while ensuring that integrity of the hosting environment cannot be damaged by individual services; the services are kept in 'compartments', such that a service cannot attack another service. The proposed security mechanisms increase the resilience of the service hosting environment against both malicious attacks and erroneous code. Thus, our proposal paves the way for large scale hosting of Grid or web services in commercial scenarios. The paper is organized as follows. Section 2 presents the threats associated with native code in a Grid service environment. Section 3 discusses related work. Section 4 presents our approach to native code security. The changes required by our security approach to the development, deployment and execution of services are described in section 5. Section 6 presents the evaluation of our approach. Section 7 concludes the paper and outlines areas for future research.

2 Analysis of the Native Code Threat Model

For our threat analysis we consider a shared service hosting environment. The service-oriented ad hoc Grid is such an environment with the most demanding security requirements, since it allows the deployment of foreign services into a running service container. Threats may also arise in more statically configured Grid environments or even web service hosting scenarios. In both cases, a single service container can be used to host more than one service.

The focus of our investigation are the security threats arising from the native code contained within Grid or web services that are deployed into a Java based web service hosting environment. Security threats in the shared hosting environment arise from direct access to the underlying system or direct access to other service instances running in the hosting environment, that cannot be limited by Java security managers or sandboxing.

We distinguish between two types of attacks, the first focusing on data managed by the hosting environment or the other services, and the second one abusing other system resources, for instance network bandwidth or CPU cycles. Examples for each of the resulting attack scenarios are presented in the following:

- **Data attack against hosting environment:** A malicious native service may be used to extract or alter security critical data from the underlying operating system or hosting environment such as the system password files, certificate files or service container configuration. In a common Grid environment, the native part of a service is executed with the user rights of the hosting environment, enabling the malicious service to read the configuration of the hosting

container, and in many cases even allowing the alteration of configuration files (such as the container authorization lists).

- Data attack against other services: A malicious native service may be used to read temporary data or results produced by other services as well as input data used by those other services. If, for example, a pharmaceutical company uses a Grid node for computations in the design phase of a new drug, a competitor may deploy a malicious service that extracts the experimental data used as input of the computation or the resulting outputs.
- Resource attack against hosting environment: A malicious native service may implement a spam bot that is used to send unsolicited bulk emails from the Grid service hosting network node. A number of denial of service attacks also fall into this category. By using native code, an attacker can cause the underlying operating system or hosting environment to crash, evading type safety mechanisms and sandbox constraints of a pure Java environment, effectively performing an internal denial of service attack on the service host.
- Resource attack against other services: A malicious native code may invoke methods from other services directly or use software licenses for 3rd party software that belongs to other service instances.
- Both of the resource attacks can be subdivided into illegal access to resources and denial of service against the local system through excessive resource consumption. Note that participation in a distributed denial of service attack counts as illegal resource access since the host is only used to harm other systems by illegally using the network interface, whereas recursively starting new threads is a denial of service attack against the hosting system.

3 Related Work

Different mechanisms for the protection of UNIX like operating systems such as Linux or FreeBSD, OpenBSD and NetBSD with respect to untrusted applications have been proposed. A very popular mechanism is the virtualization of the entire hardware, allowing a guest operating system to run in a virtual machine environment created by the host operating systems. Such virtual machine systems include Usermode Linux [4] or Xen [5]. The latter system has seen a great increase in popularity for the small performance overhead caused by its virtualisation technology.

Chroot confines file system access of a process run in the chroot environment to a different base in the file system. Some well known mechanisms exist for processes to break out of the chroot

environment and access files outside of this chroot jail. Those vulnerabilities have been addressed by the BSD implementation of the jail system call. Jails partition a BSD environment into isolation areas. A jail guarantees that every process placed in it will stay in the jail as well as all of its descendant processes. The ability to manipulate system resources and perform privileged operations is limited by the jail environment. The accessible file name space is confined in the style of chroot (i.e. access is restricted to a configurable new root for the file system in the jail). Each jail is bound to use a single IP address for outgoing and incoming connections, it is also possible to control what network services a process within a jail may offer. Certain network operations associated with privileged calls are disabled to circumvent IP spoofing or generation of disruptive traffic. The ability to interact with other processes is limited to other processes in the same jail.

Systrace [6] has become a popular mechanism for call restriction as well as privilege elevation on a fine grained scale without the need for running entire processes in a privileged context namely in OpenBSD and NetBSD with ports being available for Linux and FreeBSD as well. It uses system call interposition to enforce security policies for processes run under the control of systrace. Systrace is implemented in two parts, an addition to the kernel that intercepts system calls, comparing them to a kernel level policy map, disabling the call if a negative entry or no entry at all is present. The kernel level implementation is assisted by a user-level part that reads and interpretes policy specifications to hand them to the kernel level policy map, report policy enforcement decisions to the user applications and even call GUI applications for interactive generation of policies.

Janus [7] is one of the first system call interception tools. It uses the ptrace and /proc mechanisms which are claimed not to be a suitable interface for system call interception, since for example race conditions in the interface allow an adversary to completely escape the sandbox [8]. Janus has evolved to use a hybrid approach similar to systrace to get direct control of system call processing in the operating system [9].

The ability to set the effective user id of CGI programs to another user than the user id the calling web server runs under was introduced as the suEXEC capability in Apache 1.2 [10]. Our approach also offers the possibility of using setuid on the native processes, in addition other more fine grained access restriction methods may be used in a mixed Java and native code environment.

Emerging proposals for isolation of different Java threads in the same JVM address security

threats arising from the sharing of a single JVM between different applications [11]. An approach to implement such a shared JVM is actively pursued by Sun microsystems in the form of the multitasking virtual machine (MVM) [12]. While MVM offers mapping of isolated Java threads to operating system processes, only a prototypical implementation for Sparc Solaris has recently been published and would require the availability of the MVM for the hosting system and a switch to this new JVM. Porting a service hosting environment like the Globus Toolkit and Tomcat to the new VM requires substantial efforts to changes those systems [13].

The possibility to decouple native processes from the process space of the Java virtual machine has been investigated in [14] in order to achieve better robustness of Java applications relying on native methods, not to enhance security of a shared Java environment or the underlying operating system. Systems like the Entropia Virtual Machine for Desktop Grids [15] or GridBox [16] propose the application of virtualisation and sandboxing technologies to achieve security for native Grid applications. They can only isolate entire native applications. Applied to the scenario of Java Grid services relying on native components, they cannot provide isolation of different services inside a single JVM.

4 An Approach to Native Code Security

Our security architecture addresses countermeasures for attacks stemming from native code used in Grid or web services, that fall into the different classes described in section 2. The technique used is process separation and confinement into secure sandboxes in order to allow for a flexible and fine grained definition of execution policies in an open multiple service environment. Our requirement does not require multiple instantiation of the JVM for isolation of different services, it is also independent of the JVM implementation allowing any JVM to be used to run the Grid or web service hosting environment.

The sandbox for Java classes within the JVM is defined by a security manager. Based on a given policy, the security manager controls access of Java classes to certain resources such as the file system or network interfaces. The Java security manager can block file system access for pure Java classes that must use the File classes of the Java IO packages for file system access. From an operating system perspective, all file accesses from the JVM is performed with the user rights of the owner of the JVM process. Child processes for native code also inherit the user ID of the JVM process. While a fine-grained and policy based restriction of resource access is possible for pure Java code by means of a custom security manager, this restriction of rights is impossible for native parts of a service. The JVM cannot keep the code from opening file handles with the permissions inherited from the JVM process.

Java offers two possible ways of using native code. The first is the creation of a new child process using `Runtime.exec` or `ProcessBuilder.start`, the second is the direct invocation of native method implementations through the Java Native Interface (JNI) [17]. The only means of protection against malicious native code in the standard Java language is the use of a security manager and a policy that prohibits execution of native code as a child process or loading of shared libraries (`System.load`, `System.loadLibrary`). This is not desirable in an environment like the service-oriented Grid, where reliance on native implementations can be expected to occur frequently.

4.1 Confinement of JNI Bound Implementations

The Java Native Interface specification defines the interface between the JVM and native methods implemented in C/C++. It enables invocation of native method implementations from Java classes and callbacks to Java methods from the native code. The JNI specifies a mapping from names of Java methods declared as native to C/C++ method names as well as mappings between Java types and native types. A sample method native `int intTest(int i)`; in the class `test.A` would be mapped to the native method: `Java_test_A_intTest`. The first two arguments of this method are used to pass pointers to the JNI interface and the objects self-reference (`this`) to the native implementation. Followed by other parameters defined in the Java class for the native method.

The JNI interface is organized like a C++ virtual function table. It is passed by reference to the native implementation and managed by the JVM per thread (i.e. a native method may be invoked from different threads and therefore receive different JNI interface pointers, invocations from the same thread are guaranteed to pass along the same pointer). The structure itself contains a reference to an array of function pointers to implementations of the JNI interface methods. Besides passing an invocation result with return, the native method must use those JNI interface functions for access to any method or field in Java classes and objects managed by the JVM. The native methods are compiled into shared libraries and Java code using native implementations loads those shared libraries using `System.loadLibrary`. Native code is then executed in the process space of the JVM which leads to the serious threats described before. The native code cannot be further constrained on a fine grained per-service level, only confinement based on the JVM process owner is possible. Figure 1 shows the relationship and confinement area using a standard approach for interfacing with the native code through the JNI.

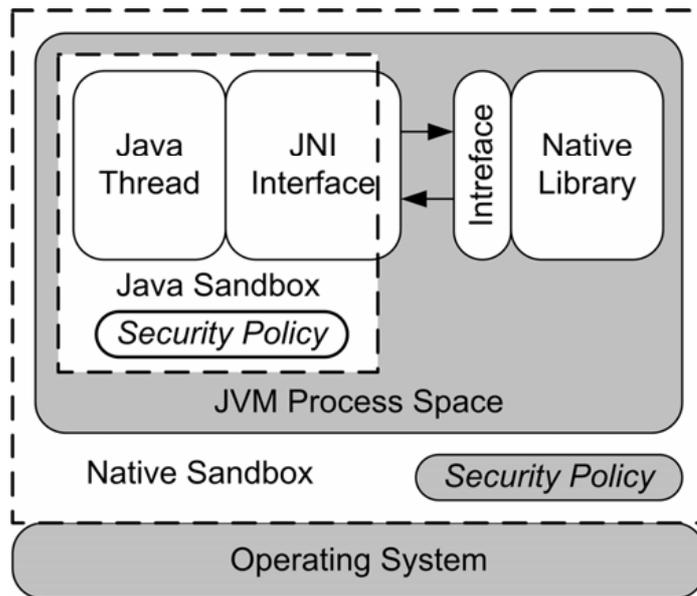


Figure 1: Standard access to native code through the JNI interface.

As a solution to this problem we propose the decoupling of the process spaces by use of an automatically generated transparent proxy intercepting all calls to the native implementation shown in figure 2. The native component of the service is replaced by a generated proxy that exposes exactly the interface of the original component. This proxy now receives all the calls to the native methods from the JVM. Such a call is passed on to the original native implementation that is instantiated in a different process than the JVM and managed by a process server. We refer to the wrapped and sandboxed native process as the I-Process. Creation of the I-Processes for the Java based Grid service hosting environment is managed by a custom process manager. The process server acts like the JVM to the native method implementations, it passes a reference to an altered JNI interface implementation to the original native code. Every reference to the JVM from the original native code is thereby intercepted by the custom JNI interface implementation. The transparent proxy and the process server communicate by means of standard IPC or RPC mechanisms, depending on the security and functionality requirements.

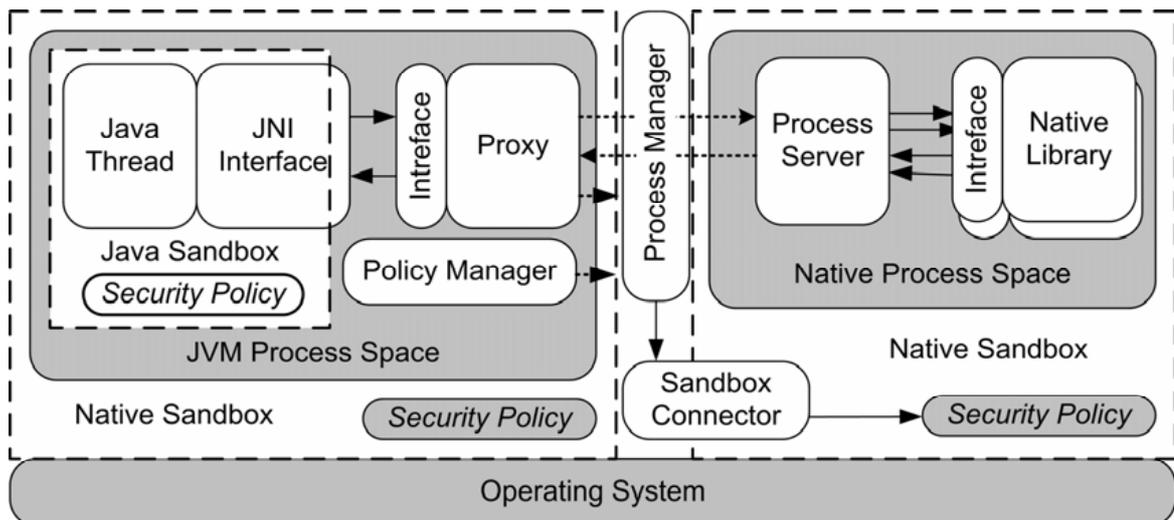


Figure 2: Decoupled process spaces for JNI attached native code, enabling secure isolation of native code.

4.2 Confinement of Execute Requests

In the previous section, native code was called via the JNI interface. A second possibility to execute native code from Java is to use the Java Runtime class to create a new shell environment where the native code is then run. The Runtime class uses the ProcessBuilder to create a new shell. The ProcessBuilder itself uses a statically linked JNI native create method to create the operating system shell. Unfortunately the proxy approach cannot be utilized in this case since the create method is integrated into the JVM. It is, however, possible to offer a custom ProcessBuilder which can sandbox a newly created shell on behalf of a service. This feature is currently not implemented in our proof-of-concept solution. To deal with this security issue in the meantime, we use the Java SecurityManager to forbid the ProcessBuilder to create new shells, thus restricting native code to be run via JNI.

4.3 Policy and Process Instance Management

In previous work [3], we have proposed a solution to intra-engine service security confining services or groups of services by use of a dedicated Java class loader. We extend this grouping scheme to the native parts of services, allowing the creation of sandboxes for I-Processes per service (or group). The hosting provider can attach a security policy to the service (group) that restricts resource access to the underlying operating system for all I-Processes created by services within this group. All processes started for services within the same service group also share the native sandbox. To be able to securely group different services, our solution to grouping Java services already provides an access control mechanism. A public-private-key pair is generated for a each newly created group. This key is obtained by the group owner (i.e. the creator of the group) who uses the private key to sign service archives containing the service

implementation (GAR files for the Globus environment). If the service implementation is signed using the correct private key, it is admitted to join the service group.

When a Java thread requests access to a native method, the transparent proxy implementation of this method will first of all check whether an I-Process has already been created for the Java thread. If a new I-Process has to be created for the current invocation, the group ID of the service is determined from the service class loader or its parent group class loader. This information can then be used to obtain the security policy that was specified for the service group by the hosting provider from the policy manager. This policy is passed along with the thread ID to the process manager that in turn creates either a new sandbox and an I-Process for the Java thread or only a new I-Process within an existing sandbox for the service group.

Creation of the sandbox environment is delegated to a sandbox connector. This component allows our isolation environment to be implemented for different underlying operating systems or different sandboxing techniques. It handles initial setup of the sandbox, inclusion of all necessary dependencies for a native library, installation and possible mediation of policies. The process manager also uses the sandbox connector for creation of new I-Processes as child processes of the sandbox connector, when such a new I-Process is requested by a Java thread calling into the transparent proxy. The hosting provider can specify the sandbox connector to use along with an access policy for a service group. Attachment of a security and sandbox policy to a service group may happen based on the user ID of the user that initially created the service group. Users may also be grouped allowing the application of a default policy to all service groups created by deployment of services by unknown users.

The Web Service Resource Framework introduces the notion of a so called Web Service Resource (WS-Resource) into the web service framework. A WS-Resource is the combination of a stateless web service and a state capturing Resource Property Document. A client receives a resource address upon creation of a WS-Resource for later reference in subsequent interactions with the WS-Resource. Service instances may attach to the resource property document in order to change the state data starting off from the current state of the WS-Resource. Isolation of the native service components connected to individual WS-Resource instances from each other would require the transparent proxy to identify the WS-Resource corresponding to the Java instance emitting the native call to the transparent proxy. There are many cases that prohibit the identification of the corresponding WS-Resource from the native side if no special precaution has been taken in the original library wrapper.

Such a fine granularity of confinement is only needed if the service implementation is untrusted and there is real concern that the service implementation could be used to exchange data between resource instances created for different users. A solution to this problem is the use of distinct service groups (and the corresponding class loaders) for different instances of the service. In this case, our proposal automatically isolates the I-Processes corresponding to different WS-Resources since they originate from different class loaders.

Our proposed solution can so far protect the hosting environment and other services against data attacks as well as illegal resource access. Depending on the sandbox capabilities, it is also possible for our process manager to monitor resource utilization and constrain e.g. CPU time as well as disk space or network bandwidth used by individual sandboxes. This still allows services to perform denial of service attacks against other service instances running in the same sandbox but leaves the option of shutting down entire sandboxes when they show behaviour that can cause harm to the hosting environment. A problem still remains in effective resource management in the JVM. No widespread solution to the problem of monitoring and managing resource consumption of individual Java threads (belonging to a service instance) is currently available.

4.4 Secure Process Spaces

A number of different options for the secure execution of native code (i.e. enforcement of access restrictions to the host operating system and isolation of processes against each other) can be used in our architecture. Those approaches have been introduced in section 3. A balanced decision needs to be made between the cost (i.e. instance creation time, computational overhead, increased resource consumption) incurred on the original Java service hosting environment and the strength of security offered by the chosen method. We will now discuss some of the implications of using the different techniques for native code isolation, with the cost and complexity imposed by such systems and the security provided by them.

Dedicated or Virtual Hosts can be used to achieve a very high level of confinement since they add a layer of abstraction for the entire hardware. Instantiation of the native process requires the shared object file to be present in the file system accessible by the new operating system instance that runs the I-Process. While startup times for the guest operating system instances of a virtual host environment can be somewhat leveled by pre-loading a number of guest operating systems that are used on demand, the memory consumption of this method is very high. We therefore favor a more fine grained and lightweight approach to resource virtualization or

compartmentalization of the I-Processes.

Changing the effective user ID of the I-Process provides security based on standard file and resource access control of the operating system. While this method requires no preparation of the sandbox and imposes virtually no performance overhead, it offers only very limited security against exploitable weaknesses in the operating system. A downside of this approach is the need for a pool of user accounts that I-Processes are mapped to by the sandbox manager.

BSD Jails are a way to partition a BSD environment into isolation areas. The jail system call has been developed to explicitly counter well known techniques to escape a similar chroot environment. The overhead created by running a process in a jail is very low. Jailed processes are tagged to belong to a certain jail, the system enforces security by identifying this tag for certain privileged operations, resulting in a very small overhead for only those calls limited by the jail environment.

Systrace offers even more fine grained policy enforcement over processes running under control of systrace. It has become a very popular tool for privilege control for a number of BSD variants. We experienced acceptable performance overhead with a Linux port of systrace. For a first prototypical implementation of the system, we employ a combination of chroot and an extended implementation of systrace as they offer the best balance between overhead and gained level of security. Shared objects and libraries they depend on are either copied mapped by hard linking into the system and can be protected by using systrace to intercept write attempts on the libraries. The (small) memory overhead is limited to the process and sandbox management components.

5 Development, Deployment and Execution

In this section, we discuss changes that are required by our security approach to the development and deployment process as well as special concerns during the execution of services.

5.1 Development

Using our system does not strictly require changes to the development process of the service containing native components. There is, however, added benefit for service developers and service users in taking a post-development step: specifying a requirement profile for their service. Such a requirement profile can also be derived automatically at deploy time, but specifying the profile beforehand helps the platform to compare requirements against the security policy and determine mismatches that lead to failure of service instantiation or

execution.

The requirement policy can either be specified manually or automatically generated in the following way: Our addition to the hosting environment supports a trace mode that can be used by the developer to install, instantiate and use the newly developed service. The process manager advises the sandbox manager to create the sandbox environment in trace mode that records any resource access of the service instance running inside this sandbox. Since we can assume the developer to trust his or her service implementation, no restrictions are enforced on the service instance. The recorded requirements stemming from a trace run can then be transformed into a generalized requirement profile to be used in the deployment descriptor of the service package. This consolidated policy contains generalizations of libraries that the service depends on, files that the service accessed and network addresses the service connected to. The consolidation operation is used to sort out typical calls that stem from the generated wrappers and standard library handling methods, in order to make the process of customization easy for the developer. Again, the requirement policy does not affect security but performance when searching for a compatible node that can be used to deploy and run the service.

5.2 Deployment

Deployment of a service containing native code consists of the following steps:

- Compare the service requirement policy with the security policy specified by the hosting provider (optional)
- Generate a custom transparent proxy for the native components of the service
- Create or join a service group and bind the security policy to the group
- Pre-load secure sandboxes (optional)

These steps must be executed on the target machine since the generated code runs in the privileged hosting environment and as such must not be supplied by the service.

The first step during deployment checks whether the environment into which the service is to be deployed offers sufficient access rights to successfully run the service. Since creating requirement specifications and publishing security policies creates additional costs and is not always required, this step is optional although it is recommended. This allows manual decisions to be made about which machines are suitable. If a service does not specify what requirements it

has, it can be deployed onto any machine, but will fail to execute if any operation is attempted that is not permitted by the secure hosting environment.

Next, for all shared objects a transparent proxy is generated by introspectively analyzing the shared objects to discover which methods contained therein are JNI compliant. Based on this information, the Java classes can be reflectively analyzed to retrieve the method signature since it is not contained in the shared object. From that, a proxy capable of accepting all pre-defined native calls in place of the original shared object is generated. The proxies are registered with the hosting environment to enable run-time monitoring of the sandbox integrity (see section 5.3). This step also includes resolution of dependencies (i.e. identification of other shared objects the service requires to run). These dependencies are recorded and attached as a sandbox descriptor to the transparent proxy.

If it is necessary that different services containing native code interact directly (i.e. not via their Grid service interfaces), the services must be deployed into the same group so they are not separated by a sandbox. The first service creates a group and the hosting environment binds a policy file to the group based on the user id of the service deployer. The service deployer also gets a public/private key pair with which all other services which should be permitted to enter the group are signed. Only properly signed services may enter the group.

The last step of the deployment process prepares the sandbox for operation. This can include the booting of a virtual hosting environment or making certain files accessible inside the sandbox (e.g. include copies or hardlinks of required shared objects in a chroot environment). This step is optional since it uses up system resources and should only be executed if the service requires quick first response times.

The steps required in the development and deployment process are visually summarized in figure 3.

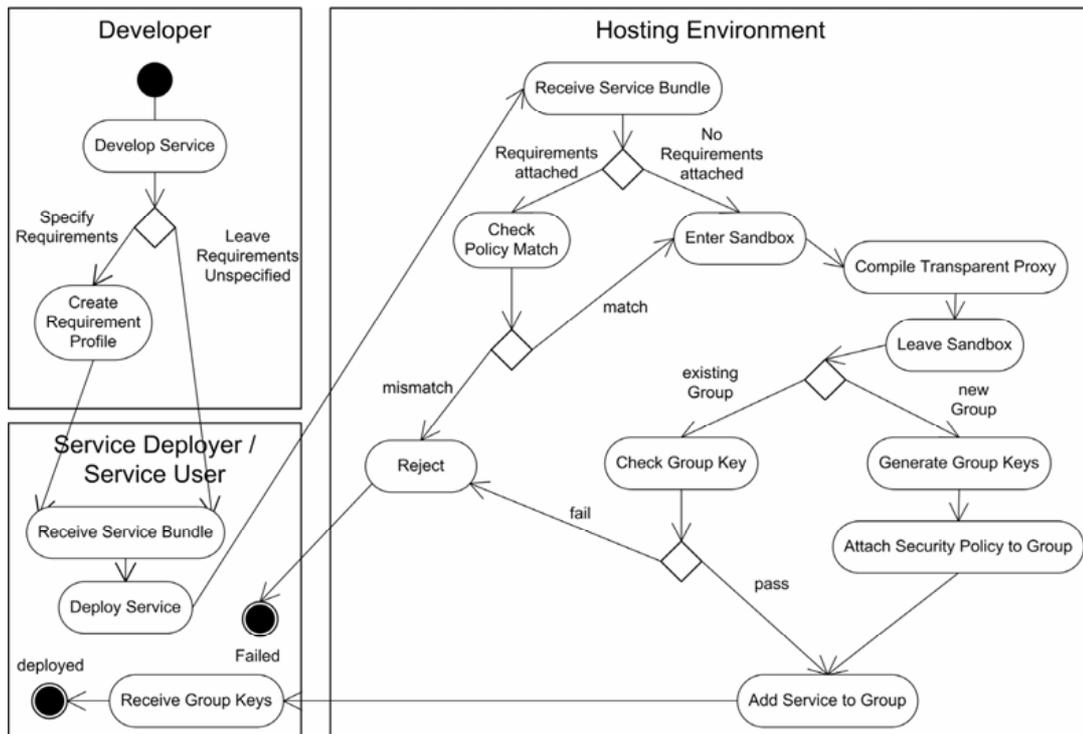


Figure 3: Steps in the development and deployment process. (The service bundle is signed with the group private key for deployment into an existing group.)

5.3 Execution

For most services containing native code, the environment is now fully configured and the service is sandboxed. As mentioned above, the proxies responsible for the sandboxing are created through code introspection. This creates the following security risk: If a service contains a shared object in an obfuscated form or generates code on the fly, the introspection process used during deployment will not be able to generate the sandboxing proxies and thus the sandbox would not be safe. To combat this, the Java SecurityManager is extended to check at runtime whether libraries which are to be loaded were processed during deployment and thus the needed proxies were generated. If there is no registered proxy for a given library, it is generated at run-time and substituted for the original shared object. Since only very few legitimate applications require the dynamic generation of code, a warning is sent to the hosting environment informing the administrator that code is being run in the sandbox which did not enter through the deployment process. This code is sandboxed but it should nevertheless be checked by the administrator.

6 Evaluation

Two main factors govern the performance of the presented security solution for individual calls to native methods: The overhead imposed by the native sandbox of the operating system and the

overhead imposed by the process dislocation technique that requires local inter process communication.

The cost analysis for native code sandboxing is subject to evaluation by the creators of the various techniques offering a solution in this area and is not covered by the scope of this paper.

In order to obtain an initial result on the performance of the presented solution, the following experiment was conducted to assess the overhead of transparent process dislocation for native code used in a Java service implementation through the JNI. A Java test class was created that uses a native method implemented using C. The absolute time before and after an invocation of the native method in the Java class was measured in a loop of 500 invocations. In the first case, the original native library compiled using gcc 4.0.1 was used. Then, the native library was replaced by a transparent proxy library that used RPC communication to perform method invocation on the original library in a process separate from the JVM. Again, the runtime of the native method invocation in the Java class was measured.

The regular call took 3 microseconds on the average while the call using separate process spaces took 566 microseconds. This increase in time required to perform a native call by a factor of 182 is only acceptable if relatively few native calls are required from the Java code into native libraries. Fortunately this is the normal situation when a Grid service implementation provides a front end to functionality provided by native business or scientific libraries. In this case methods provided by the library are used on a macroscopic level and the service will typically spend substantial amounts of its runtime in the native code alone instead of requiring many native method invocations in the Java implementation.

A benefit of the micro jailing technology lies in the small memory overhead created by the solution. The native process manager and all RPC related components require less than 1 megabyte of main memory. This is a small overhead compared to process isolation by use of dedicated Grid service container instances for different services and users. Using these dedicated Grid service containers, a complete JVM must be instantiated, requiring at least 20 megabytes of memory.

7 Conclusions

In this paper, we have analyzed the threat scenarios that emanate from native code in a service-

oriented ad hoc Grid environment and categorized them into four distinct types: data attack against the hosting environment, data attack against other services, resource attack against the hosting environment and resource attack against other services. To counter those threats, a novel security architecture was presented, which enables confinement of native components of Grid applications into a secure environment. The security framework is based on dynamically created JNI proxies which create a pipe between the secure Java environment and the secure native environment. The security solution is capable of protecting the hosting system as well as services from each other. While our work is focused on the Grid environment, our security solution also offers benefits in a regular shared web service hosting environment.

Future work includes the extension of the policy generation and matching mechanism and further usability tools. A custom ProcessBuilder will be provided to extend the sandboxing capabilities of the system to include `Runtime.exec()` commands as well as JNI calls.

Acknowledgements

This work is supported by the German Ministry of Education and Research (BMBF) (D-Grid Initiative, In-Grid Project)

References

- [1] Smith, M., Friese, T., Freisleben, B.: Towards a Service-Oriented Ad Hoc Grid. In: Proc. of the 3rd International Symposium on Parallel and Distributed Computing, Cork, Ireland (2004) 201–208
- [2] Friese, T., Smith, M., Freisleben, B.: Hot Service Deployment in an Ad Hoc Grid Environment. In: Proc. of the 2nd Int. Conference on Service Oriented Computing, New York, USA, ACM Press (2004) 75–83
- [3] Smith, M., Friese, T., Freisleben, B.: Intra-Engine Service Security for Grids Based on WSRF. In: Proc. of Cluster Computing and Grid, Cardiff, UK (2005)
- [4] Dike, J.: User-Mode Linux. In: Proc. of the 5th Annual Linux Showcase and Conference, Oakland, USA (2001)
- [5] Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neugebauer, R., Pratt, I., Warfield, A.: Xen and the Art of Virtualization. In: Proc. of the ACM Symposium on Operating Systems Principles (SOSP), Bolton Landing, USA, ACM Press (2003) 164–177
- [6] Provos, N.: Improving Host Security with System Call Policies. In: Proc. of the 12th USENIX Security Symposium, Washington, USA (2003) 257–272
- [7] Goldberg, I., Wagner, D., Thomas, R., Brewer, E.A.: A Secure Environment for Untrusted Helper Applications. In: Proc. of the 6th Usenix Security Symposium. (1996)

- [8] Wagner, D.A.: Janus: an Approach for Confinement of Untrusted Applications. Technical report, CSD-99-1056 (1999)
- [9] Garfinkle, T.: Traps and Pitfalls: Practical Problems in System Call Interposition Based Security Tools. In: Proc. of the ISOC Symposium on Network and Distributed System Security. (2003)
- [10] The Apache Software Foundation: (Apache HTTP Server Documentation: suEXEC) <http://httpd.apache.org/docs/suexec.html>.
- [11] Sun Microsystems, Inc.: Java Specification Request 121: Application Isolation API Specification (2003)
- [12] Czajkowski, G., Daynès, L., Titzer, B.: A Multi-User Virtual Machine. In: Proc. of the USENIX 2003 Annual Technical Conference, San Antonio, USA (2003) 85–98
- [13] Jordan, M., Daynès, L., Czajkowski, G., Jarzab, M., Bryce, C.: Scaling J2EE(TM) Application Servers with the Multi-Tasking Virtual Machine. Technical report, SMLI TR-2004-135, Sun Microsystems Laboratories (2004)
- [14] Czajkowski, G., Daynès, L., Wolczko, M.: Automated and Portable Native Code Isolation. Technical report, 2001-96, Sun Labs (2001)
- [15] Calder, B., Chien, A.A., Wang, J., Yang, D.: The Entropia Virtual Machine for Desktop Grids. In: Proc. of the First ACM/USENIX Conference on Virtual Execution Environments. (2005)
- [16] Dodonov, E., Sousa, J.Q., Guardia, H.C.: GridBox: Securing Hosts from Malicious and Greedy Applications. In: 2nd International Workshop on Middleware for Grid Computing. (2004)
- [17] Sun Microsystems, Inc.: Java Native Interface Specification (2003)

A Service-Oriented Grid-Based Infrastructure for Supporting Virtual Prototyping in Manufacturing

Manfred Grauer, Julian Reichwald and Thomas Barth

Information Systems Institute
University of Siegen
57076 Siegen
{ grauer,reichwald,barth@fb5.uni-siegen.de }

Abstract

Virtual prototyping in manufacturing gains more and more importance, since companies competing in industry noticed the increasing cost pressure alongside higher quality requirements. The utilization of virtual prototyping techniques provides a possibility to reduce the total development costs and the time to market, while simultaneously innovations of both products and processes increase the overall quality of the manufactured parts. The monetary effort and the lack of know-how especially in small and medium sized enterprises which is required to operate such systems lead to a disadvantage in the more and more globalized market. This paper states the problem of supporting virtual prototyping processes in the manufacturing industry and gives a solution approach by utilizing service-oriented architectures and concepts from grid computing. A pilot implementation of the proposed architecture is introduced and evaluated by case studies in various manufacturing domains.

1 Introduction

Companies competing in the manufacturing industry range from small and medium sized enterprises (SMEs) up to global players. They noticed the increasing quality requirements on their products (e.g. allowed tolerances in the product's geometry compared to the specification) alongside the pressure of lowering development and production cost to compete in the market.

Virtual product development, namely the utilization of numerical simulation and optimization techniques, can help to achieve the mentioned business goals. But since the utilization of virtual product development techniques requires dedicated hardware and expensive software systems as well as highly skilled system operators it is nearly unaffordable – especially for SMEs – to efficiently incorporate virtual prototyping in their product research and development. Besides, operating such systems goes far beyond the core competencies of such enterprises and should not be encouraged as an in-house solution.

The grid computing paradigm (s. [BeFH03] or [FoKe04]) has proven to utilize powerful distributed computing and storage resources in different – mainly scientific – scenarios efficiently. Current projects like the Large Hadron Collider Computing Grid [LCG06], dedicated to process and distribute the huge amount of data produced by the Large Hadron Collider which is currently being built at CERN, show the popularity and the acceptance of the approach. Alongside grid computing, service-oriented architectures (SOA) have been of research interest in the past (s. [DJMZ05] or [SiHu05]). The idea of loosely coupling software components while also ensuring a platform-independent and standardized way of communication (e.g. via web services), SOAs are far more flexible than monolithic software systems. Furthermore, process automation can be achieved by utilizing workflow engines which take care of invoking previously specified services in a defined execution order as well as passing necessary arguments to them and receiving their calculation results.

The coupling of service-oriented architectures and grid computing therefore can lead to a highly dynamic approach of interconnecting standardized grid services in an *on-demand* manner, lowering the need (especially for SMEs) to operate costly virtual prototyping resources in-house, but enabling them to access such resources whenever needed (s. www.migrad.de). The effect is a reduction of the total cost of ownership for both IT infrastructure as well as employment costs for the required specialists and a quality improvement for products and manufacturing processes.

This paper is organized as follows: In the next chapter, the process of simulation-based optimization in manufacturing as a basis for virtual prototyping is stated and the problems in supporting such processes adequately by IT systems are figured out. Afterwards, a solution approach based on service-oriented architectures and concepts from grid computing is given and the necessary services for the process support are identified. A pilot implementation of the proposed infrastructure and some of the identified services are given based on de-facto standard middle-

ware for grid computing. The functionality of the implemented system is then evaluated in a case study from the metal casting domain by optimizing a casting process of a gas turbine blade. Additionally, two more scenarios from virtual prototyping in industry are presented: The simulation of a sheet metal deep drawing process and the DMU (Digital Mock Up) kinematics simulation of a deep-drawing process including machinery for transferring parts between different stages of manufacturing. Afterwards, related work is discussed and finally several conclusions and topics for future work are presented.

2 Virtual Prototyping – The Problem Presentation

A virtual prototyping process in industrial practice is a lengthy course of action where physical prototypes are built, evaluated, and changed until product and process models are found which suit the predefined needs. Joint multistage optimization problems which commonly appear in the domain of sheet metal forming [GBGR05] or the highly complex tasks in metal casting process optimization [JBRG06] become arbitrarily difficult to solve.

Optimization algorithms coupled with numerical simulations provide a way to apply virtual prototyping on products and processes. The algorithm generates multiple alternative parameter sets of a given input model, thus generating a number of process designs, which are then evaluated. Since a typical model evaluation lasts from hours to days, the designs should be evaluated in parallel for efficiency reasons. At this point, powerful computing resources are required and efficient, scalable optimization algorithms have to be used to assure the highest degree of system load and by that the economic utilization of resources.

The complexity of a virtual prototyping process in manufacturing requires sophisticated know-how and therefore the involvement of various experts (see Fig. 1) and institutions from different problem domains, commonly geographically distributed and in different companies .

- The **domain expert** with the knowledge of the specific manufacturing problem (e.g. metal forming). This person resides in the enterprise and can be seen as an end user or service consumer respectively.
- Specialists from **computational engineering** (having knowledge about the simulation systems used for the numerical calculations as well as the CAD software suites which are used for the geometry models) as well as **experts in numerical optimization** (knowing about the mathematical algorithms). Typically, they are geographically dis-

tributed and do not necessarily belong to the same enterprise. Since these specialists offer their know-how to the end user, they can be seen as service providers in the overall process.

- A **telecommunication engineer** who operates and maintains the required IT infrastructure (hardware, software, local/global networking to handle communication between collaborating partners) and acts as a provider of telecommunication services.
- **Software integrators** which try to integrate domain-specific (e.g. CAD/CAE tools from computational engineering) and generic enterprise functionalities (e.g. Customer Relationship or Supply Chain Management systems) across organizational and technical boundaries in a distributed environment.

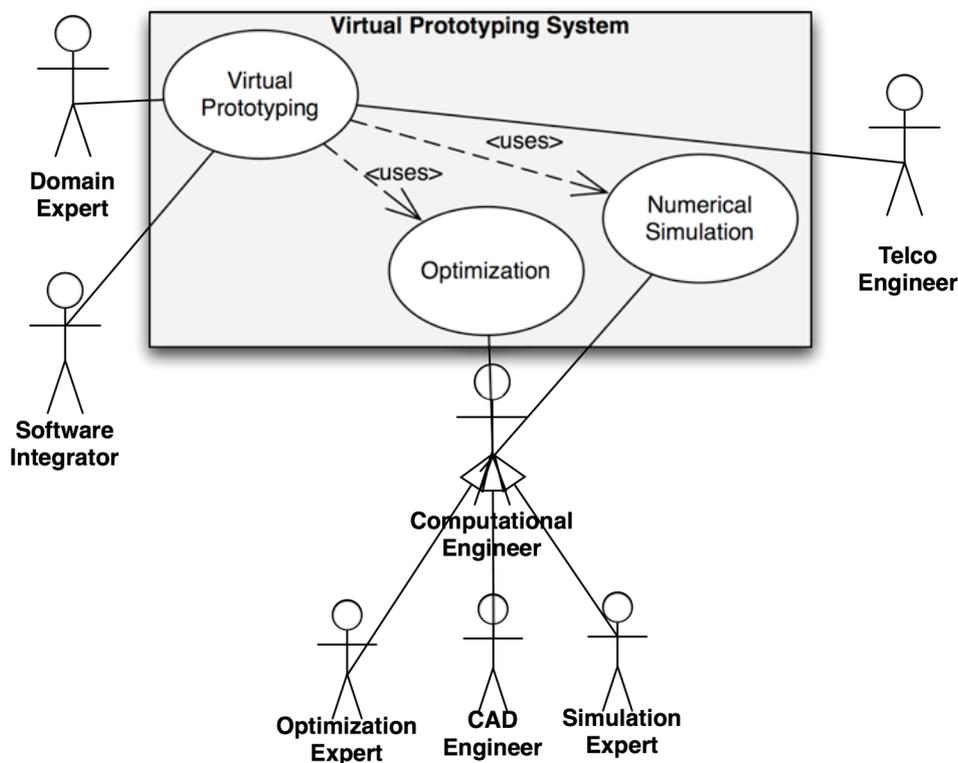


Figure 1: UML Use Case diagram of the incorporated actors in a virtual prototyping process.

As an example, a typical process in the product and process design in manufacturing, especially in the metal forming domain (e.g. sheet metal forming like deep drawing, metal casting), depicted in Fig. 2, will be used and analyzed here. The process can be divided into five phases: In the **first phase**, the real-life metal forming process has to be mapped to a process model, incorporating the specific details of the process (e.g. geometry, materials etc). For this task, not only the domain expert with the appropriate know-how of the real-life process has to be involved.

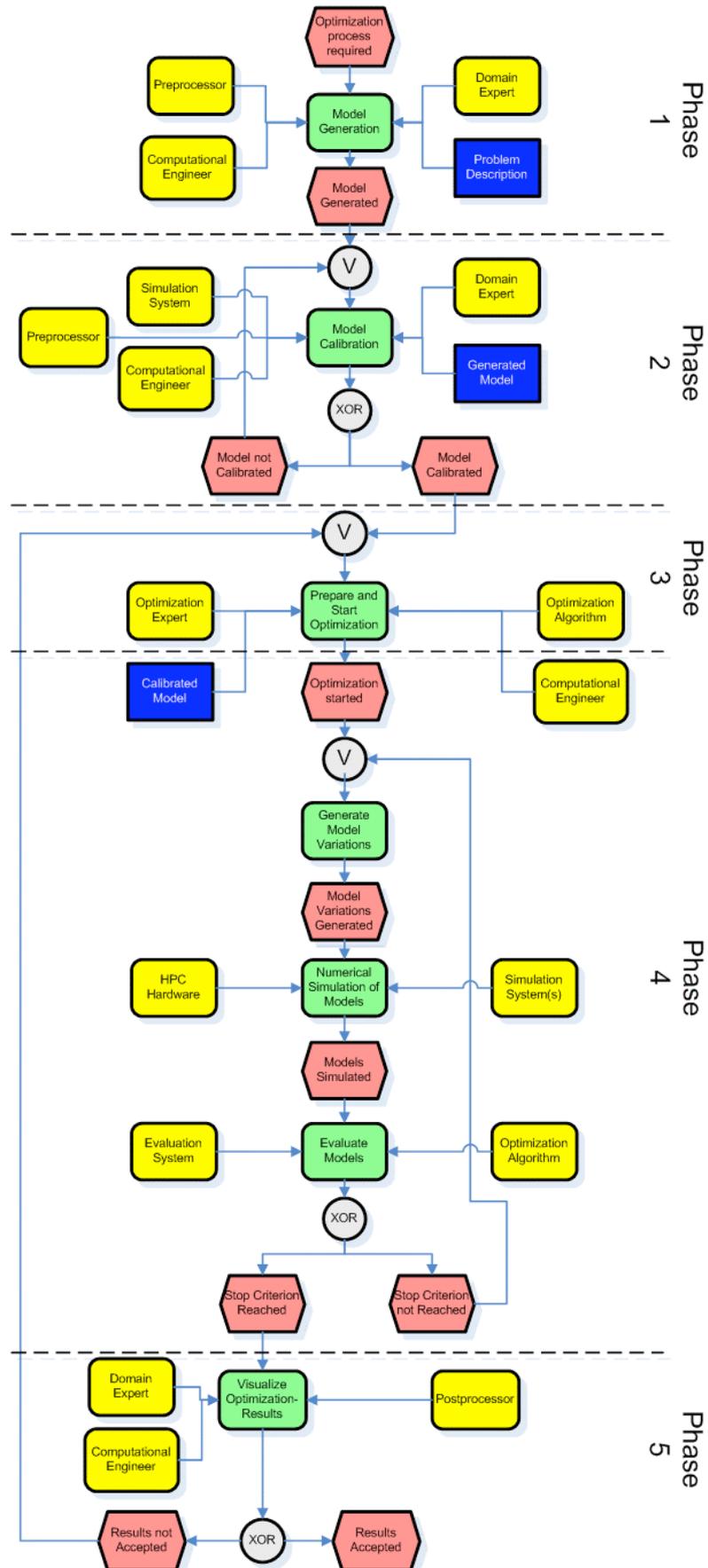


Figure 2: Core Phases of virtual prototyping presented for industrial forming by its business process

Moreover, a computational engineer having knowledge of the simulation software package and the preprocessing tools is needed. When the initial model creation is completed, the **second phase** is entered. The model needs to be calibrated, i.e. model parameters will be adjusted and single simulation runs are made to verify that the model reflects the reality. Again, the domain expert and computational engineer are required in close cooperation to calibrate the previously generated model. Since the calibration process is particularly difficult, the model is continuously checked if it suits the needs. Hence, the output of single simulation runs is permanently reviewed by all participants. The calibration process iterates as long as the model is not yet adequately calibrated. During the **third phase**, preliminary jobs (e.g. transferring the calibrated models to the compute nodes, setting algorithm parameters etc.) for preparing the optimization run are performed. The optimization sub-process in **phase four** is entered next. The optimization algorithm now takes care on generating a number of model parameter sets which need to be simulated and evaluated. As long as the stopping criteria of the optimization algorithm is not yet reached, this sub-process iterates until the model is sufficiently solved. **Phase five** consists of the post processing which incorporates the collaborative viewing between the domain expert and computational engineers. If the resulting model is not sufficient, the optimization cycle is re-entered.

The variety of prerequisites (software systems and competencies) in a geographically distributed scenario in each phase of the process is a challenging task for a sufficient software support. An infrastructure that adequately supports the whole process should be able to handle the tight cooperation among the involved participants. Collaboration aspects have to be respected as well as discovering and handling distributed data-, hard- and software resources. Furthermore, the model parameterization and generation needs to be supported in a generic way so optimization algorithms as well as metal forming simulation packages may be coupled together arbitrarily and in a flexible way without adjusting the code.

Since the whole process is geographically distributed and inter-organizational, heterogeneous resources and various technical (possibly proprietary) infrastructures are typically involved. The problem solving environment has to take notice of that and offer a way to interconnect the enterprises using open standards for communication, hence leading to an open, extensible system.

3 Solution Concept for the Virtual Prototyping Grid-Environment

The specific requirements for the supporting IT system are directly derived from the process shown in figure 2. First, the environment needs to be able to provide the computing capacities of high-end workstations and clusters over a limited period of time. This compute power must be available on demand without remarkable delays. Furthermore, the user must be able to monitor the current state and get intermediate results while the optimization is in progress. To prevent the general distribution of proprietary data and the implied risks, secure data transfer and remote file handling is mandatory. Alongside the computing infrastructure, the technology has to provide support for collaboration, so the geographically distributed experts are able to set up the initial models, calibrate them or discuss the results. Finally, it must be possible to adopt the overall process to changes (e.g. switch to a new simulation system or another optimization algorithm). In phases one, two and five, collaborative aspects between the involved parties dominate the process. Collaborative aspects and grid process generation is already discussed in [FSFR06], so infrastructural topics will be the main focus here.

The utilization of common hard- and software systems showed to be inapplicable to provide an acceptable solution strategy for the aforementioned scenario in the metal forming industry. The excessive runtime of numerical simulations require the utilization of high performance multi-processor machines or compute clusters, which exceed the financial possibilities of SME's. The transformation from fixed costs into variable costs by outsourcing such simulation jobs seem to be an adequate solution, alongside letting the enterprises concentrate on their core competencies. The spatially distributed process across different enterprises therefore requires a supporting IT infrastructure taking notice of the different hard- and software systems as well as the flexible combination of the components.

Service-oriented architectures built with standardized web services while also adopting concepts from grid computing provide a promising way out of the dilemma [BHMN04]. The loose coupling of components provides the high degree of flexibility needed to support the process. Services, which can be seen as single components in the overall software system, are provided by enterprises specialized in this field. These enterprises have to care about the proper functionality of the service they offer as well as service updates or availability. The business logic, knowledge and complexity is encapsulated and hidden by standardized, self-describing interfaces, which can be accessed by service consumers via standard internet connections and by using standard protocols and communication techniques.

The utilization of workflow engines can even more increase the degree of flexibility. By deploying whole workflows as a web service and providing them to others, complex service compositions and choreographies can be reused by other services or workflows. The main characteristics of a web service based service-oriented architecture adopting the grid paradigm lead to a network of services which provides the core services for the aforementioned simulation and optimization process, which are namely

- **File transfer services** which are potentially needed in all phases of the mentioned process, since input models, result files etc. have to be transferred to and from the compute nodes.
- **Services providing the appropriate optimization algorithms.** As stated in [GBGR05], direct search methods capable of constraint handling and with good scalability behaviour in distributed environments are a preferred choice here to ensure the highest possible degree of resource utilization.
- **Services providing the numerical simulation and evaluation.** Since most simulation packages consist of native, closed-source legacy code compiled for a predefined hardware architecture, a way of accessing such software packages in a service-oriented way according to the underlying technology has to be found.

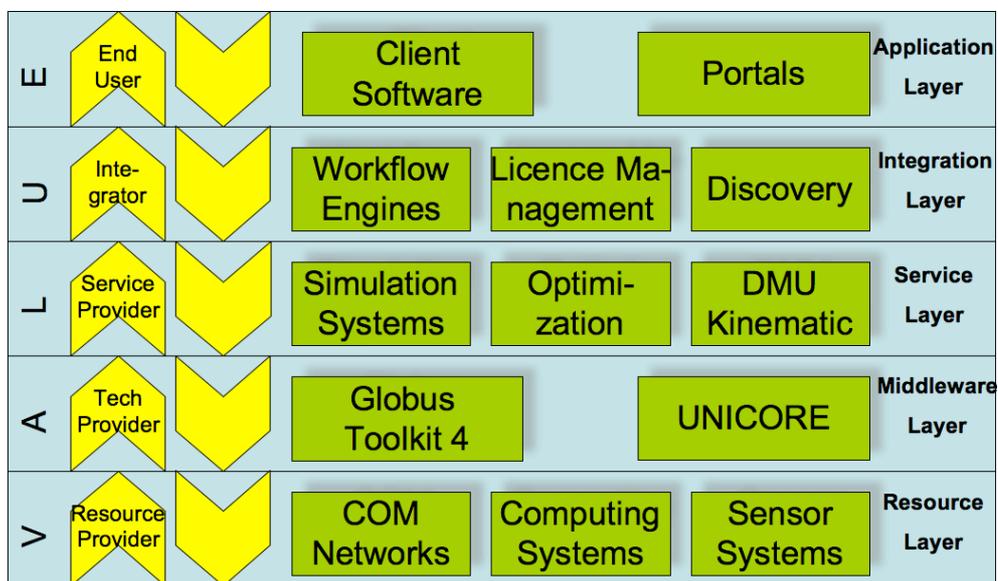


Figure 3: A service-oriented grid-based infrastructure and the corresponding value chain [Por85]

Figure 3 shows the proposed infrastructure split into five layers, according to the key personnel and institutions which are responsible for the system to work. At the same time the corresponding value chain [Por85] is indicated. In higher layers, the amount of value added for the end

user is increasing. Since the scope of this paper is focused on applications and the virtual prototyping process, infrastructural details like security aspects, billing, accounting, licensing or service-level agreements are not considered here.

4 Implementation of the Virtual Prototyping Grid-Environment

To adequately provide a support for the aforementioned process, a grid-based service-oriented infrastructure was designed and implemented. The Globus Toolkit (GT4) [Glob06] is the de-facto standard for building Web Service Resource Framework (WSRF) based grid infrastructures and was chosen as the underlying middleware. GT4 allows the utilization of standard grid services (e.g. the Monitoring and Discovery Service (MDS) or Reliable File Transfer (RFT)), but also the implementation of custom and – due to WSRF – stateful grid services. Since the file transfer and discovery services are already present in the infrastructure, the services for the simulations and the optimization algorithms have to be implemented. Since GT4 is (in its core parts) implemented in Java, the simulation and optimization services were also implemented in that language. First steps of the implementation were already presented in [GBGR05], where a test problem (representing the mixed-integer multistage optimization problem in sheet metal forming) could be optimized by using a client which realized both the optimization algorithm as well as the workflow implementation. A comparison of the problem solving time by using Dis-

Optimization method	# of CPUs	Optimal objective function value $f(x^*, \mathcal{A})$	Number of objective function evaluations	Wall-clock time [seconds]
SQP (Matlab [®])	1	26.9497	3166	47.16
DPS on SOA Architecture	1	26.9842	19868	927.63
	10			87.85
	50			17.32
	100			10.83
	200			6.24

Table 1: Comparison of DPS (SOA) and SQP (Matlab) when solving the test problem for a mixed-integer multistage sheet metal forming problem [GBGR05], using a service-oriented implementation of DPS

tributed Polytop Search (DPS) on the one hand and the sequential quadratic programming (SQP) implementation of Matlab[®] has been made (see table 1). The table indicates the feasibility of the SOA-implementation and a good scalability of the DPS-method in distributed sys-

tems. Nevertheless, results based on the aforementioned prototype should be seen as proof of concept, and not a comprehensive study on scalability or the quality of optimization algorithms. Next, numerical simulations for manufacturing problems were implemented, furthermore the optimization method has been deployed as a web service as well to allow arbitrary interconnections. Implementation details for the optimization service and for a metal casting service (as one example for a simulation service) are given in sections 5.1 and 5.2.

The whole infrastructure has been deployed on a 296-CPU cluster computer, consisting of 148 compute nodes with 2 CPUs, 2 GB main memory and 80 GB hard drive per node. The operating system for each node is SuSE SLES9.3. A redundant head node as well as a redundant storage node is used for cluster administration, control and storage access. The nodes are interconnected by Gigabit Ethernet.

4.1 The Optimization Service

This service (s. Fig. 3, “Service Layer”), is an implementation of the distributed polytop search (DPS) which belongs to the class of direct search method (s. [Wri95], [BeTs89]). The DPS was designed regarding efficiency and scalability in distributed systems. During its runtime, it requires an a priori unknown number of evaluations of both the objective function and corresponding constraint functions. The service has to save its state each time an evaluation request occurs, and it passes the data set which is to be evaluated immediately back to the service caller instead of directly invoking the simulation service. This behavior opens the possibility to use workflow engines to keep control over the whole workflow execution. Considering these conditions, the service makes use of the possibilities provided by WSRF. Being instantiated following the factory design pattern, the instance service operates on a set of resources (s. Fig. 3), allowing the service to keep its actual state, even when it is actually not in use. Furthermore, internal state variables checkpoint the calculation state, so the algorithm can be set on hold and resume at a given place in the code [GBGR05].

Besides a service operation which allows a client to set necessary parameters needed by the polytop search, the grid service operation `iterate(IterateRequest)` takes care of starting and resuming the algorithm at the appropriate position - according to its internal state and according to the input data inside the `IterateRequest` data structure. A resulting data set is returned immediately after invoking the operation, telling the client if further evaluations are needed or if the DPS reached a predefined stopping condition.

4.2 The Simulation Services for Metal Casting and Sheet Metal Forming

Since the implementation of the sheet metal forming service is analogous to the metal casting service, the casting service is described in detail as an example. The main purpose of this service (s. Fig. 3, “Service Layer”), is to wrap the metal casting legacy software CASTS[®] [LaNS98] as a Grid service. However, the CASTS Service does not only provide a service-wrapped version of CASTS, but it also takes care of the following operations: It is capable of modifying the input model of the casting process according to a set of parameters passed to the service. This parameter set can be the input received from optimization services (e.g. the DPS implementation mentioned above). The service executes the CASTS legacy application on a number of different execution platforms. Since a cluster computer is used in this case, a simulation request is leading the internal execution subsystem to incorporate the local resource manager Torque [TORQ06] and the scheduling system Maui [MAUI06]. The execution state of the cluster job is monitored and exposed by the Casts Service. The execution subsystem is highly modularized so that the service also works on single workstations without local queuing/scheduling. The service also provides functionality to evaluate the simulation result (which is done by CritCASTS, a legacy software system bundled with CASTS) and determining the objective function value as well as the constraint function values, needed for simulation-based optimizations. The utilization of WS-GRAM [Glob06] for the job execution was discouraged because of the aforementioned extra functionality which goes beyond the capabilities of WS-GRAM and leaving the process logic implementation up to the service consumer, which presupposes inexistent know-how.

In the next section, the aforementioned prototypical implementation is used to demonstrate the feasibility of the approach when applying it in the domain of metal casting. Furthermore, additional industrial scenarios from sheet metal forming are described focusing on optimal product (section 5.2) and process design (section 5.3).

5 Case Studies from Manufacturing and Results

5.1 Virtual Prototyping in Metal Casting

As an example, a casting process of a gas turbine blade was used to evaluate the implemented infrastructure (see figure 3). The evaluation of the simulation results have been done from the following points of view [JBRG06]:

- The probability of local freckle (shrinkhole) information at the surface of the turbine blade. Freckle probability was estimated based on the temperature gradients calculated by the simulator,
- the degree of curvature of the solidification front. It should be as horizontal as possible in order to achieve a high quality directional solidification,
- the ratio G/v (temperature gradient over solidification speed) must be greater than a critical value and
- the process time.

Goal of the optimization was an improved withdrawal profile (temperature gradient over solidification speed) for the casting process of the turbine blade. The withdrawal process is parameterized by eleven constrained design variables representing withdrawal velocities.

The optimization algorithm service has been parameterized to utilize 16 CPUs by generating 32 new withdrawal sets in each iteration.

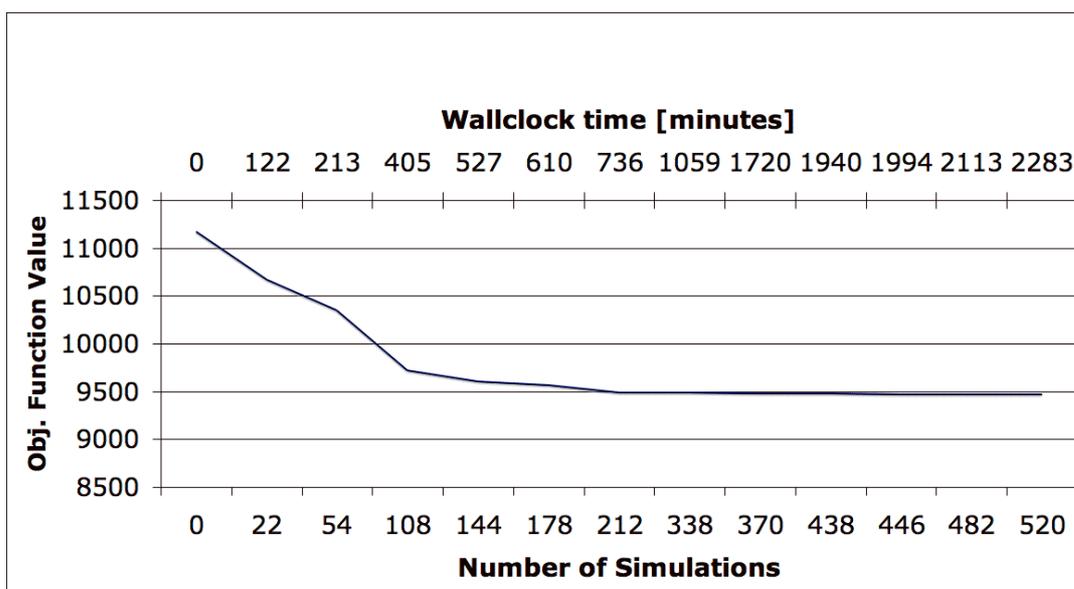


Figure 4: Objective function value in dependence to the total number of simulations and the total time in minutes

Figure 4 shows the objective function in dependence to the total number of simulations needed to yield the function value. Since the global minimum of the objective function is unknown, the calculated quality has to be interpreted by metal casting experts. Nevertheless, an optimization achievement can be noticed by reducing the objective function by 15.24%.

5.2 Virtual Prototyping in Sheet Metal Forming

Sheet metal forming copes with the forming process of solid state metal sheets, where the forming process aims on form, surface, and material properties. Risks in planning the manufacturing process can be predicted and reduced immediately and materials can be used in a more efficient and cost-effective way. The software system FETI-INDEED[®] [KWRD01] has been utilized to simulate deep-drawing processes with the ability of partitioning the assembly parts as well as friction observance and rezoning (mesh refinement during simulation runtime).

As an example from sheet metal forming, the calculation of a cylindric cup for the automotive industry was calculated with FETI-INDEED.

As can be seen in figure 5, the cup was partitioned into 16 parts by using the FETI method. A higher processor usage leads to a decreased processing time from 44:34 hours (2 CPUs) down to 06:41 hours (36 CPUs).

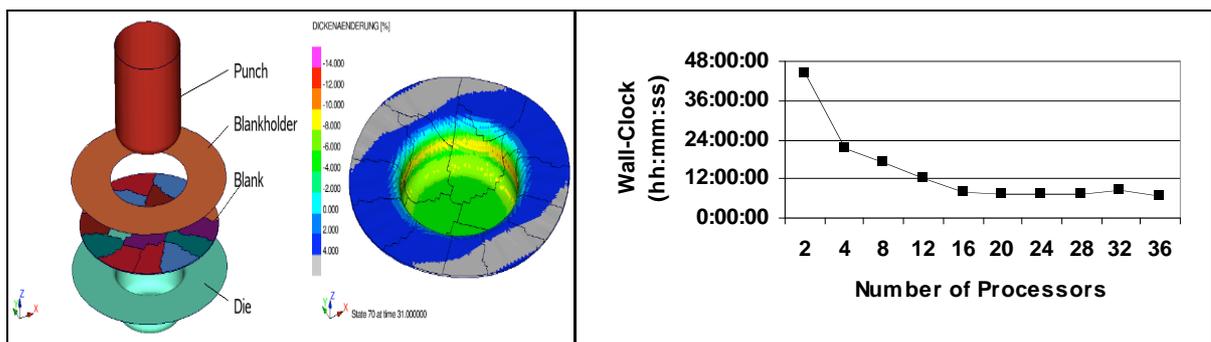


Figure 5: Model of a cylindric cup, subdivided into 16 parts, calculated with FETI-INDEED and according wall clock graph by varying the number of CPUs.

5.3 Virtual Prototyping of DMU Kinematics

Beyond sheet metal forming simulations, kinematics of the production process itself and the behavior of automatic transfer equipment can be simulated and visualized ([Cu05], [SchMe06]). The superposition of movements by transferring assembly parts to the next working stage and

the automatic transfer equipment can reduce the overall time needed for the whole process to avoid collisions, which depends on various parameters (e.g. transfer speed, stamp speed, angle etc). A trade-off between collision-free production processes and a maximum turnout has to be found. By utilizing such dynamic simulations, the component delivery to the next working station can be visualized and – if collisions appear – the parameters and tools can be adjusted early in the overall production process to achieve a collision-free manufacturing process again.

Figure 6 shows a simulation of a sheet metal forming press with a 630 tons maximum pressure, 4000 mm x 1700 mm table size and 500 mm maximum travel. Transfer equipment for assembly part movement between the working stations has been applied to the press.

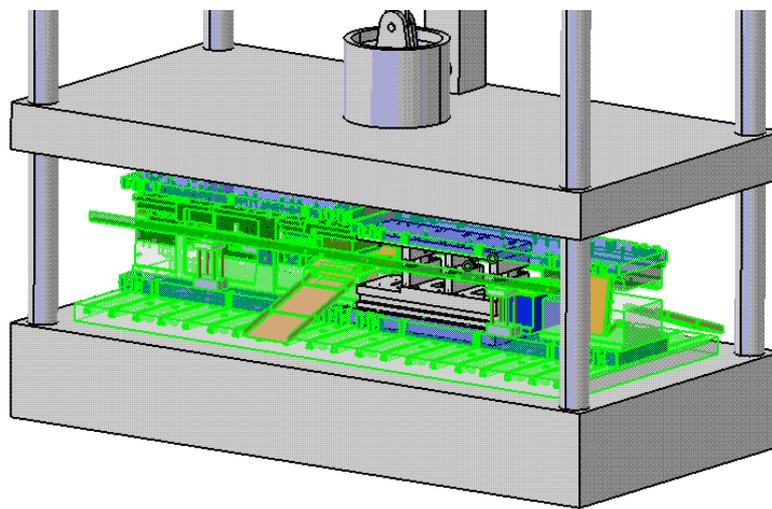


Figure 6: DMU kinematic-simulation of a sheet metal forming press and corresponding transfer equipment, visualized by CATIA V5 [SchMe06]

6 Related Work

Supporting business processes with software systems and especially service-oriented architectures realized with web services have received considerable attention in both academia and industry. Several other research projects try to cope with similar subjects in related fields.

The **Geodise** project [XSCK04; SONK04] focuses on optimization, design and fluid dynamics, especially in aerodynamics. Its main goal is to provide a distributed problem solving environment (PSE) for engineers working in the mentioned fields by utilizing e.g. MATLAB and adding Grid functionality to it. Although first Geodise implementations were based on Globus

Toolkit version 2, the core Geodise Toolbox is now part of the managed program of the Open Middleware Infrastructure Institute (OMII) [OMII06].

A **Grid-enabled problem solving environment for engineering** in design where distributed parties are able to collaborate has been introduced by Goodyer et al [GBJS06]. The system enables its users to start Grid jobs on Globus Toolkit based hosts, but the main focus is put on collaborative application steering and result visualization using the gViz library [BDGS04] instead on automatic or semiautomatic simulations (or virtual prototyping respectively).

The **P-GRADE** Portal [SiKa05] aims to be a workflow-oriented computational Grid portal, where multiple clients can collaboratively participate in design, development and execution of a workflow as well as multiple grids may be incorporated in the workflow execution. The P-GRADE Portal is based on Globus Toolkit version 2 for basic grid operations (such as file transfer and job execution), the workflow execution is done by a proprietary implementation. But since P-GRADE does not rely on web service based grid infrastructures, it can be stated that P-GRADE is a system based on proprietary standards and therefore will not provide the high degree of flexibility needed to adequately support the aforementioned virtual prototyping process in an adequate way.

Summarizing, the presented approaches emphasize either technical aspects from domain-specific numerical simulation or workflow management in engineering. Hence, the solution approach presented in this paper is focused on the integration of domain-specific applications from computational engineering, workflow management in engineering, and aspects from enterprise application systems as well. This integration provides the capability to consider technical as well as business aspects simultaneously.

7 Conclusions and Future Work

In this paper, the need for utilizing virtual prototyping techniques in manufacturing industry was shown and the complexity of a virtual prototyping process was indicated with all of the incorporated institutions, key actors and virtual prototyping resources (i.e. hard- and software). A solution approach of a layered service-oriented grid-based infrastructure and the according value chain was proposed to adequately support the virtual prototyping process in manufacturing enterprises. A prototype implementation was presented and computational results from the prob-

lem domains of metal casting, sheet metal forming, and digital mock-up kinematics were shown.

As future work, the single simulation- and optimization runs should be extended and coupled such that a complete production process chain can be simulated in by the system. Beyond the optimization of a process chain the design of the optimal layout of a complete manufacturing cell, a production line comprising multiple manufacturing cells, and finally optimizing the layout of the whole factory is envisaged.

Acknowledgements

Parts of the work presented in this paper are partly supported by a grant from the German Ministry of Education and Research (BMBF) (D-Grid initiative, InGrid Project) and by the Ministry of Science and Research of Northrhine-Westfalia out of the MiGrid-Project. Furthermore we would like to thank our industry partner ACCESS e.V., Aachen, GNS mbH Braunschweig and Co.Com Concurrent Computing GmbH, Siegen for providing both the simulation software systems and the dedicated know-how.

References

- [BeFH03] Berman, F.; Fox, G.; Hey, T. (eds.): Grid Computing: Making the Global Infrastructure a Reality, J. Wiley & Sons, Chichester 2003
- [BeTs89] Bertsekas, D., P.; Tsitsiklis, J. N.: Parallel and Distributed Computation – Numerical Methods, Prentice-Hall, Englewood Cliffs 1989
- [BDGS04] Brodlie, K., Duce, D., Gallop, J., Sagar, M., Walton, J., Wood, J.: Visualization in Grid Computing Environments. In: Proc. of IEEE Visual., pp. 155-162, 2004
- [BHMN04] Booth, D., Haas, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C., Orchard, D.: Web Services Architecture, W3C Working Group Note, <http://www.w3.org/TR/ws-arch> (visited 2006/07/24)

- [Cu05] Culha, B.: Automatic Arrangement of Product Parts, Assemblies and Modules, in: Bouras, A.; Gurumoorthy, B.; Sudarsan, R. (eds.): Product Lifecycle Management PLM 05, Inderscience, Geneve, 2005
- [DJMZ05] Dostal, W.; Jeckle, M.; Melzer, I.; Zengler, B.: Service-orientierte Architekturen mit Web Services – Konzepte-Standards-Praxis (in German), Elsevier Spektrum, München 2005
- [FoKe04] Foster, I.; Kesselman C. (eds.): The Grid 2: Blueprint for a New Computing Infrastructure, Morgan Kaufmann 2004
- [FSFR06] Friese, T., Smith, M., Freisleben, B., Reichwald, J., Barth, T., Grauer, M.: Collaborative Grid Process Creation in an Engineering Domain”. In: Proc. of the 13th IEEE Conf. on High Perf. Comp., Bangalore/India 2006 (to be published)
- [GBJS06] Goodyer, C., Berzins, M., Jimack, P., Scales, L.: A Grid-enabled problem solving environment for parallel Computational Engineering Design. In: Advances in Engineering Software, 37:439-449, 2006
- [GBGR05] Grauer, M., Barth, T., Gerdes, M., Reichwald, J.: On distributed virtual prototyping in the automotive supplier industry using service-oriented computing. In: Hahn, A. and Grauer, M. (Eds): Informations- und Wissensdrehscheibe Produktdatenmanagement (in German), GITO, 2005
- [Glob06] globus.org: The Globus Alliance, <http://www.globus.org> (visited 2006/07/24)
- [JBRG06] Jakumeit, J., Barth, T., Reichwald, J., Grauer, M., Friese, T., Smith, M., Freisleben, B.: A Grid-Based Parallel Optimization Algorithm Applied to a Problem in Metal Casting Industry, BIOMA 2006 (submitted)
- [KWRD01] Kessler, L., Weiher, J., Roux, F.-X., Diemer, J.: Forming Simulation of High and Ultrahigh Strength Steel using INDEED with the FETI Method on a Workstation Cluster. In: Mori, K. (ed.) Proc. of NUMIFORM'01, Balkeman Pub., pp. 399-404, 2001

- [LaNS98] Laschet, G., Neises, J., Steinbach, I.: Micro- Macrosimulation of casting processes. 4^{ième} école d'été de Modélisation numérique en thermique, 1998
- [LCG06] cern.ch, Large Hadron Collider Computing Grid Project, <http://lcg.web.cern.ch/LCG> (visited 2006/07/24)
- [MAUI06] clusterresources.com, Maui cluster scheduler, <http://www.clusterresources.com/pages/products/maui-cluster-scheduler.php> (visited 2006/07/24)
- [OMII06] omii.ac.uk. Open Middleware Infrastructure Institute (OMII). <http://www.omii.ac.uk>. (visited 2006/07/24)
- [Por85] Porter, M.: Competitive Advantage, New York, Free Press, 1985
- [SchMe06] Schuth, M., Meeth, J.: Bewegungssimulation mit CATIA V5. Hanser, 2006
- [SiHu05] Singh, M., P.; Huhns, M., N.: Service-Oriented Computing – Semantics, Processes, Agents, Wiley 2005
- [SiKa05] Sipos, G. and Kacsuk, P.: Collaborative workflow editing in the P-GRADE. Proceedings of the MicroCAD Conference, Miskolc, Hungary, 2005
- [SONK04] Song, W., Ong, Y.-S., Ng, H.-K., Keane, A., Cox, S., Lee, B.: A service-oriented approach for aerodynamic shape optimization across institutional boundaries. In: Proceedings of ICARCV, China, Kunming, 2004
- [TORQ06] clusterresources.com, Torque Resource Manager, <http://www.clusterresources.com/pages/products/torque-resource-manager.php>, (visited 2006/07/24)
- [XSCK04] Xue, G., Song, W., Cox, S., Keane, A.: Numerical Optimization as Grid Services for Engineering Design. In: Journal of Grid Computing, 2(3):223-238, 2004
- [Wri95] Wright, M.: Direct Search Methods: Once scorned, now respectable. In: Griffiths, D.; Watson, G. (eds.), Proc. Dundee Biennial Conf. in Num. Analysis, pp. 191 - 208, Addison-Wesley, 1995

Entwurf eines Enterprise Architecture Frameworks für serviceorientierte Architekturen

Betrachtung am Beispiel einer erweiterten UN/CEFACT Modeling Methodology

Philipp Offermann, Christian Schröpfer, Marten Schönherr

Fachgebiet Systemanalyse und EDV
Technische Universität Berlin
10623 Berlin

{Philipp.Offermann, Christian.Schroepfer, Marten.Schoenherr}@sysedv.tu-berlin.de

Maximilian Ahrens

Deutsche Telekom Laboratories
10587 Berlin
maximilian.ahrens@telekom.de

Abstract

Damit serviceorientierte Architekturen (SOAs) ihr Potential voll entfalten können, müssen neben technischen auch methodische und organisatorische Aspekte betrachtet werden. Enterprise Architecture Frameworks (EAFs) ermöglichen die dazu notwendige ganzheitliche Sicht auf das Unternehmen. Einzelne EAFs sind meist zu spezialisiert, um als Framework für die SOA angewendet werden zu können. Das Metaframework Generalised Enterprise Reference Architecture and Methodology (GERAM) kann jedoch genutzt werden, um die für eine SOA notwendigen Elemente zu identifizieren. Zusätzlich haben wir nachgewiesen, dass das Framework UN/CEFACT Modeling Methodology (UMM) für die Integration von innerbetrieblichen Prozessen geeignet ist. Daher wird die UMM unter Verwendung der GERAM in Hinblick auf die SOA analysiert, um Schwachstellen aufzudecken. Darauf aufbauend werden Erweiterungen der UMM für die SOA vorgeschlagen.

1 Einleitung

Während Technologien für serviceorientierte Architekturen (SOA) bereits standardisiert sind und eine gewisse Verbreitung gefunden haben, befinden sich Methoden zur Einführung einer SOA noch im Entwicklungsstadium.

1.1 Ziel des Artikels

Der Artikel hat zum Ziel, einen Beitrag zu methodischen Grundlagen für den Entwurf von SOAs zu leisten. Zur Einführung einer SOA gehört eine ganzheitliche Sicht auf das Unternehmen, wie sie gewöhnlich von Enterprise Architecture Frameworks (EAFs) eingenommen wird. Jedoch haben einzelne EAFs jeweils ihren eigenen Kontext und sind meist zu spezialisiert, um ohne weiteres für das Architekturkonzept der SOA eingesetzt werden zu können.

Die Generalised Enterprise Reference Architecture and Methodology (GERAM) stellt ein Metaframework für EAFs da, welches zum Vergleich von existierenden und zur Entwicklung von neuen Frameworks als Referenz benutzt werden kann. Ausgehend von GERAM wird in diesem Artikel die UN/CEFACT Modeling Methodology (UMM) daraufhin untersucht, inwieweit sie sich für den Entwurf einer SOA eignet. Die Untersuchung der UMM unter Bezugnahme auf GERAM soll aufzeigen, welche Probleme sich in Hinblick auf SOAs ergeben, welche zusätzlichen Artefakte benötigt werden und wie die UMM erweitert werden kann, um neben der Geschäftsprozessintegration auch SOAs zu unterstützen.

1.2 Methodik

Als wissenschaftliche Methodik wird die Deduktion verwendet [BrHW04, 5]. Ausgehend vom Metaframework GERAM wird abgeleitet, welche Anforderungen an ein Framework für SOAs zu stellen sind. Diese abgeleiteten Anforderungen werden dann mit dem existierenden Framework UMM verglichen, um daraus Erkenntnisse über die Anwendbarkeit der UMM für SOAs und mögliche Erweiterungen abzuleiten. Für die UMM wurde an unserem Fachgebiet bereits unter Anwendung der Methodik der Aktionsforschung [BrHW04, 16-17] nachgewiesen, dass sie für die Integration unternehmensinterner Geschäftsprozesse anwendbar ist [Diet06].

1.3 Gliederung

Zunächst werden EAFs eingeführt, um dann das Metaframework GERAM vorzustellen. Zusätzlich wird das Konzept der SOA vorgestellt und eingeordnet, welche Anforderungen an ein EAF für die SOA sich durch GERAM ergeben. Danach wird das konkrete Framework UMM vorgestellt und aufgezeigt, welche Erweiterungen für die innerbetriebliche Service-Orchestrierung notwendig sind. Schließlich wird die UMM in GERAM eingeordnet.

2 Enterprise Architecture Frameworks

Als erstes werden Enterprise Architecture Frameworks (EAFs) und die GERAM vorgestellt.

2.1 Das Ziel von Enterprise Architecture Frameworks

Unternehmen sind komplexe Gebilde, in denen Menschen in formalen und informalen Strukturen zusammenarbeiten, um ihre jeweiligen Ziele zu erreichen. Heutzutage erfolgt die Unterstützung dieser Strukturen vielfach durch EDV. EAFs helfen dabei, einen Überblick über das komplexe Zusammenspiel von Menschen, Organisationsstruktur und EDV zu erlangen. Für *Schallert und Rosemann* ist die Beherrschbarkeit von Komplexität und Management von Unternehmensintegration das oberste Ziel einer Architekturbetrachtung [ScRo03, 48]. Die International Organization for Standardization (ISO) definiert die Ziele einer Enterprise Architecture allgemein mit der Befähigung eines Teams, umfassend alle Ressourcen eines Unternehmen zu integrieren [ISO00, vii].

2.2 Existierende Frameworks

Seit den 80er Jahren wurden viele EAFs entwickelt. Eine Übersicht findet sich in [Sche04] und [Schö04]. Zu den bekanntesten Frameworks zählen CIMOSA, GIM – GRAI, PERA, Zachman, ARIS und DoDAF. Jedes dieser Frameworks hat auf Grund seiner Entstehungsgeschichte einen anderen Fokus.

Angesichts dieser Vielzahl von EAFs hat die *International Organization for Standardization* (ISO) einen Standard publiziert, in dem Anforderungen an EAFs definiert sind [ISO00]. Im Annex des ISO-Standards findet sich die *Generalised Enterprise Reference Architecture and Methodology* (GERAM) der IFAC/IFIP [IFIF99], ein generalisiertes Framework, um konkrete Frameworks vergleichen und bewerten zu können [BeNe97].

2.3 GERAM

GERAM besteht aus neun Elementen. Diese sind in Abbildung 1 dargestellt.

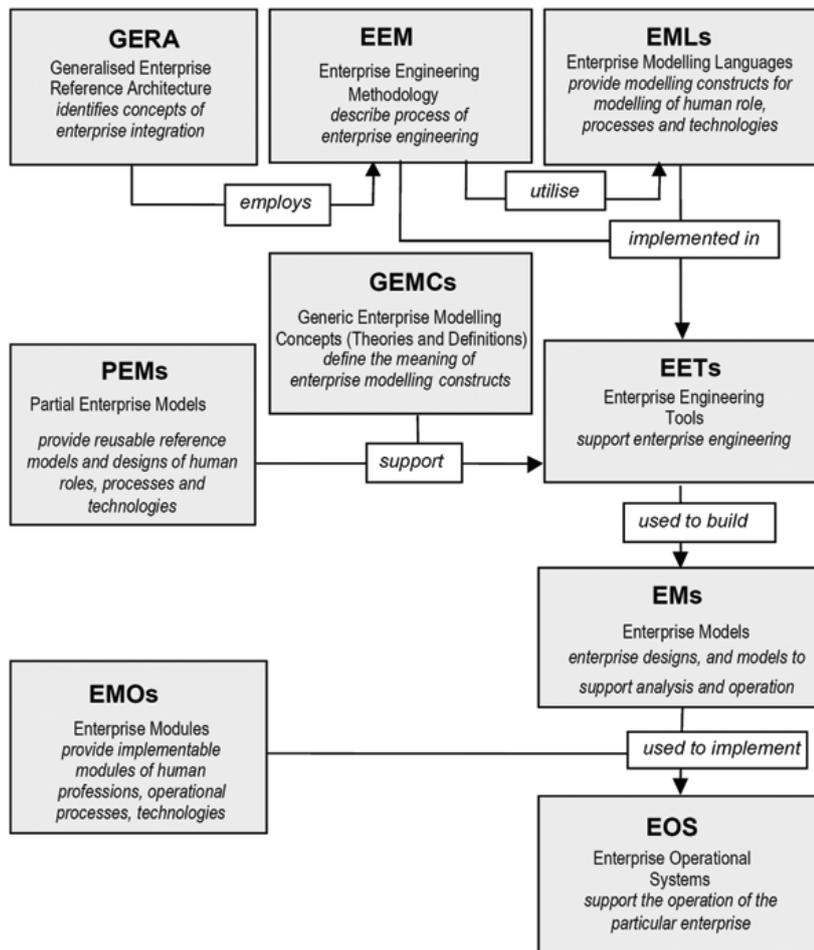


Abb. 1: GERAM Elemente [IFIF99, 5]

Das zentrale Element der GERAM ist die *Generalised Enterprise Reference Architecture* (GERA). In ihr werden allgemeine Konzepte zur Beschreibung eines Unternehmens empfohlen. Die Konzepte werden in drei Klassen unterteilt: mitarbeiterorientierte, prozessorientierte und technologieorientierte Konzepte. Die prozessorientierten Konzepte beinhalten ein Lebenszykluskonzept, welches acht Phasen unterscheidet. Die Phasen sind in Abbildung 2 zu sehen. GERA unterscheidet fünf Entity-Typen, welche Lebenszyklen besitzen können [IFIF99, 15]:

- Strategic Enterprise Management Entity,
- Enterprise Engineering/Integration Entity,
- Enterprise Entity,

- Product Entity und
- Methodology Entity.

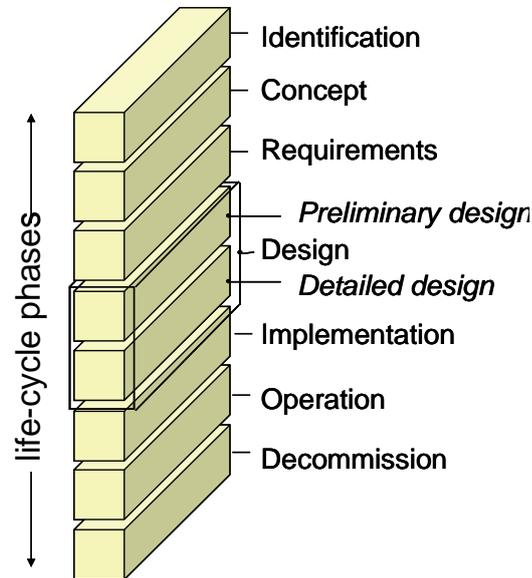


Abb. 2: GERA Lebenszyklusphasen [IFIF99, 10]

Die *Enterprise Engineering Methodology* (EEM) verwendet die abstrakten Konzepte der GERA und beschreibt allgemeine Methoden, die zur Erstellung von konkreten Unternehmensmodellen und dem Management von Unternehmenseinheiten entlang ihrer Lebenszyklen verwendet werden können.

Die Unternehmensmodelle, die durch Anwendung der EEM erstellt werden, müssen festgehalten werden. Hierzu sieht GERAM *Enterprise Modeling Languages* (EMLs) vor. Um die verschiedenen Aspekte eines Unternehmens modellieren zu können, sollte es mehrere unterschiedliche Modellierungssprachen geben. Durch die Integration der verschiedenen Sprachen wird es möglich, ein durchgängiges Unternehmensbild zu schaffen [IFIF99, 25].

Zum Erstellen und Verwalten von Unternehmensmodellen gibt es *Enterprise Engineering Tools* (EETs). Diese ermöglichen die Anwendung der EMLs und unterstützen damit die EEM. Weitere Anforderungen an EETs sind u. a. die Unterstützung von Kollaboration und ein Speicher für wiederverwendbare Modellteile.

Zur Unterstützung der Modellierung mit einem EET führt GERAM *Partial Enterprise Models* (PEMs) ein. Dabei handelt es sich um Referenzmodelle, die in vielen Unternehmen angewendet werden können. Der Vorteil der Benutzung von Referenzmodellen ist, dass die Modellierung

beschleunigt wird und unter Umständen die Referenzmodelle effizienter sind als die tatsächlich im Unternehmen eingesetzten Strukturen.

Für die Modellierung können weiterhin *Generic Enterprise Modeling Concepts* (GEMCs) eingesetzt werden. Hierbei handelt es sich um Glossare, Metamodelle oder Ontologien, welche die Bedeutung der verwendeten Begriffe definieren. Nur durch diese einheitliche Festlegung von Begriffen ist es möglich, konsistente Modelle aufzubauen.

Die EETs werden eingesetzt, um *Enterprise Models* (EMs) zu erstellen. Diese Modelle können als Entscheidungshilfe, gemeinsame Verständigungsgrundlage und Trainingsmaterial sowie zum Betrieb und zur Kontrolle verwendet werden.

Auf Grundlage der EMs kann schließlich die EDV, die *Enterprise Operational Systems* (EOS), implementiert werden. Zur Implementierung können *Enterprise Modules* (EMOs) eingesetzt werden, welche im Unternehmen oder am Markt verfügbare Module zur Umsetzung der EOS sind. Generell sind EMOs Implementierungen von PEMs. Wenn ein Teilunternehmensmodell (PEM) in einem Unternehmensmodell (EM) verwendet wird, kann das entsprechende Unternehmensmodul (EMO) für die Umsetzung des Unternehmensmodells (EMs) in den Unternehmenssystemen (EOS) benutzt werden.

2.4 GERAM und serviceorientierte Architekturen

Mit seinen neun Elementen spannt GERAM einen allgemeinen Rahmen auf, der von konkreten EAF gefüllt wird. Ein konkretes EAF fokussiert sich dabei oft auf bestimmte Aspekte eines Unternehmens, um diese beherrschbar zu machen und ggf. optimieren zu können. Ein solcher Aspekt sind Organisations- und EDV-Strukturen, die das Konzept der serviceorientierten Architektur (SOA) umsetzen.

Es wird zunächst das Konzept der SOA erläutert, um dann an einem konkreten Beispiel aufzuzeigen, welche Anforderungen an ein EAF für SOA im von GERAM aufgespannten Rahmen gestellt werden.

3 Serviceorientierte Architekturen

SOA sind ein aktueller Trend in der EDV-Industrie [Gome06]. Viele große Firmen wie IBM, Microsoft, BEA und SAP unterstützen Standards im Bereich von SOA oder stellen ihre Produkte auf eine SOA um [IBMC06; Micr06; BEAS06; SAPA06]. Auch Organisationen wie das

World Wide Web Consortium (W3C), OASIS und die Object Management Group (OMG) entwickeln und publizieren Standards im Zusammenhang mit SOA [WWWC04; OASI06; OMG06].

3.1 Ziel von serviceorientierten Architekturen

Die Gartner Group definiert eine SOA wie folgt: „SOA is a software architecture that builds a topology of interfaces, interface implementations and interface calls. SOA is a relationship of services and service consumers, both software modules large enough to represent a complete business function. So, SOA is about reuse, encapsulation, interfaces, and ultimately, agility.“ [McNa03]

Die am häufigsten angewendete Technologie zur Implementierung einer SOA sind Web services. Ein Web service ist ein Programm, welches seine Funktionen über definierte Schnittstellen und offene Protokolle anbietet [WWWC04]. Der Unterschied zur klassischen Modularisierung von Programmlogik ist, dass die Funktionalität, welche von den Web services angeboten wird, aus den Aktivitäten des Geschäftsprozesses und nicht aus den EDV-Systemen abgeleitet wird [Szyp98; Schm05; Food05].

Aus diesem Grund beschränkt sich das Konzept der SOA nicht auf die technische Seite, sondern reicht bis in das Geschäftsprozessmanagement hinein [LeRS02]. Indem Web services so entwickelt werden, dass sie Geschäftsprozessfunktionalität abbilden, ist es möglich, die Struktur der Geschäftsprozesse in Übereinstimmung mit der die Geschäftsprozesse unterstützenden EDV zu bringen.

Um zu ermöglichen, die EDV-Struktur an den Geschäftsprozessen auszurichten, ist es notwendig, die Reihenfolge, in der die Web services aufgerufen werden, zu beschreiben. Bei dieser Beschreibung muss sichergestellt werden, dass die Reihenfolge der Serviceaufrufe strukturell gleich ist zur Reihenfolge der Aktivitäten im Geschäftsprozess. Eine solche ausführbare Beschreibung heißt Orchestrierung, wenn die Aufrufe zentral koordiniert werden, und Choreographie, wenn die Koordination dezentral erfolgt [NeLo05, 246].

3.2 Anforderungen einer SOA an Elemente der GERAM

Um sowohl die Organisationsstruktur mit den Geschäftsprozessen als auch die EDV mit den Web services ganzheitlich modellieren zu können, eignen sich EAF sehr gut. Eine SOA stellt jedoch bestimmte Anforderungen an ein EAF, die am Rahmen von GERAM erläutert werden.

Mit dem Ziel, die EOS auf Basis von Services zu bauen, können Implementierungen von einzelnen Web services als EMOs angesehen werden. Das EAF sollte also als EMOs Web services vorsehen. Die Web services können, da sie Geschäftsprozessaktivitäten unterstützen, auf abstrakter Ebene auch als PEMs betrachtet werden.

Web services sollten strukturell in der gleichen Reihenfolge aufgerufen werden, in der auch die Geschäftsprozessaktivitäten ausgeführt werden. Deshalb müssen bei den EMs insbesondere die Geschäftsprozessmodelle beachtet werden. Diese sollten es ermöglichen, unter Verwendung der PEMs und ihrer Implementierung in EMOs automatisch eine Orchestrierung oder Choreographie von Web Services zu erzeugen.

Die EETs sollten es entsprechend unterstützen, die Geschäftsprozessmodelle so zu beschreiben, dass die Prozessmodelle in eine ausführbare Orchestrierung oder Choreographie übersetzt und Web services an die Aktivitäten gebunden werden können. Hierfür müssen wie bereits erwähnt die PEMs Web services und wenn möglich bereits vororchestrierte Prozesse enthalten. Weiterhin sollten, da für eine Bindung der Services an Aktivitäten eine semantische Beschreibung der funktionalen und nicht-funktionalen Eigenschaften der Aktivitäten sowie der Services nötig ist, die GEMCs eine solche Beschreibung unterstützen.

Für die EMLs ist neben den erwähnten Anforderungen an die Geschäftsprozessmodelle insbesondere vorzusehen, dass Rollen im Unternehmen für das Servicemanagement modelliert werden können. Entsprechend müssen in GERA die Konzepte ausgebaut werden, um die Modellierung des Servicemanagements zu ermöglichen.

Wir haben die UN/CEFACT Modeling Methodology (UMM) erweitert und werden an diesem Beispiel zeigen, wie eine konkrete Ausprägung der einzelnen Elemente für SOAs aussehen kann. Hierbei wird die GERAM als Rahmen verwendet, um die Elemente der UMM einordnen und damit die Vollständigkeit des Frameworks bewerten zu können.

4 UN/CEFACT Modeling Methodology für SOA

Um das Ziel der SOA, das heißt die Anpassung der EDV-Strukturen an die Geschäftsprozesse, zu erreichen, müssen die Prozesse aufgenommen und aus ihnen die Services abgeleitet werden. Die UN/CEFACT Modeling Methodology (UMM) stellt eine Methode und Modelle bereit, um Prozesse aufzunehmen, die für die Prozesse notwendige Kollaboration zwischen verschiedenen Geschäftspartnern zu identifizieren und eine Choreographie für Services zwischen den Ge-

schäftspartnern abzuleiten [UNCE06a]. Der Ursprung der UMM ist die Koordination von kollaborativen Geschäftsprozessen. Wie erwähnt wurde aber bereits nachgewiesen, dass sich die UMM auch zur Integration von unternehmensinternen Geschäftsprozessen eignet [Diet06].

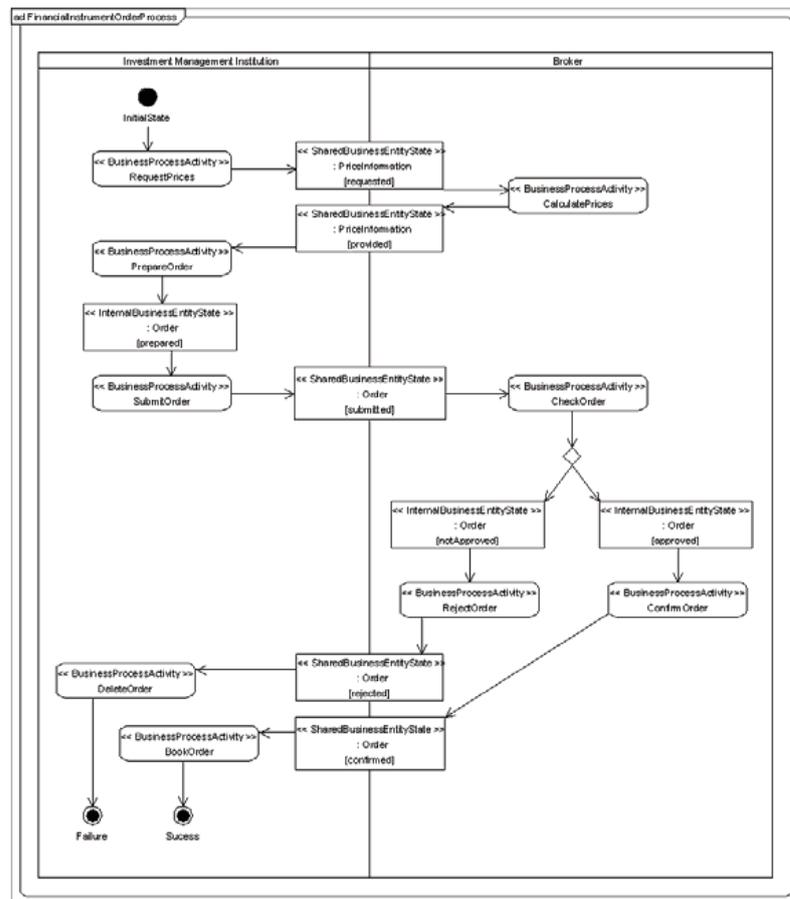


Abb. 3: Beispiel des UMM Business Process View [UNCE06a, 35]

4.1 Die Methodologie

Die UMM basiert auf drei Sichten: Business Domain View, Business Requirements View und Business Transaction View.

Die *Business Domain View* wird verwendet, um Geschäfts- und Prozessfelder zu identifizieren. Für jedes Feld werden dann Geschäftsprozesse und in diese involvierte Interessenten modelliert. Zum Einsatz kommen hierbei auf UML Anwendungsfalldiagrammen basierende Diagramme.

Im *Business Requirements View* werden im nächsten Schritt die im Business Domain View identifizierte Geschäftsprozesse modelliert. Dazu definiert die UMM den *Business Process View*, der auf UML Aktivitätsdiagrammen basiert. Ein Beispiel ist in Abbildung 3 zu sehen. Für die im Business Process View verwendeten Business Entities werden zusätzlich im *Business*

Entity View Lebenszyklen festgelegt. Aus dem Business Process View kann der Bedarf an Kollaboration zwischen Unternehmen abgelesen werden. Dieser Bedarf wird im dritten und letzten View des Business Domain Views, dem *Partnership Requirements View*, genauer modelliert.

Nachdem der Business Requirements View modelliert worden ist, ist der *Business Transaction View* der letzte Schritt der Methodologie. Er besteht aus drei Sichten: Business Choreography View, Business Interaction View und Business Information View. Im *Business Choreography View* wird der Ablauf von für eine Kollaboration zwischen Geschäftspartnern nötigen Aktivitäten festgehalten. Einzelne Kollaborationen werden dann im *Business Interaction View* modelliert. Die Informationseinheiten, die bei einer Kollaboration ausgetauscht werden, werden im *Business Information View* spezifiziert.

Vorteilhaft ist es, wenn die Informationseinheiten, die im Business Interaction View ausgetauscht werden, einheitlich verwendet und semantisch beschrieben werden. Hierzu können Core Components, wie auch UMM ein UN/CEFACT Standard, verwendet werden. Konkret können Business Information Entities, die von Core Components abgeleitet sind, die in UMM verwendeten Information Entities realisieren [UNCE06b].

Auf Basis der Business Interaction View und der Business Choreographie View in Verbindung mit den durch Core Components unterstützten Information Entities ist es möglich, eine Choreographie zu erstellen. Mit Hilfe des UMM Plug-Ins [ReUn06] für den Sparx System Enterprise Architect [Spar06] ist es zum Beispiel möglich, aus den Modellen eine Choreographiebeschreibung auf Basis von BPSS [UNCE03] zu erzeugen.

4.2 Kritik

Auf Grund ihres Ursprungs ist die UMM für die Integration kollaborativer Geschäftsprozesse gut geeignet. Wie dargestellt ist es sogar möglich, bei entsprechender Werkzeugunterstützung ausführbare Orchestrierungen zu generieren. Schwachstellen ergeben sich jedoch, wenn innerbetriebliche Prozesse betrachtet werden sollen. Diese sind entscheidend, um in einer SOA eine Prozessorchestrierung generieren zu können. Außerdem sind weiterreichende Beschreibungen der Prozessaktivitäten wünschenswert, um aus diesen Anforderungen an Services ableiten zu können. Lösungsvorschläge für diese Schwachstellen haben wir in Form von Erweiterungen der UMM vorgeschlagen.

4.3 Erweiterungen der UMM

Wir haben drei Erweiterungen der UMM vorgeschlagen, um neben der Erstellung von Choreographien auch die Erstellung von Orchestrierungen möglich zu machen. Der Vorteil hiervon ist, dass die Umsetzung einer SOA sowohl im Unternehmen als auch zwischen Unternehmen mit einer einzigen Methodologie und einem kohärenten Satz von Modellen möglich ist.

Die erste Erweiterung bezieht sich auf die Verwendung des Business Process View. Wir schlagen vor, die Aktivitäten auf der ersten Ebene nur mit jeweils einer Aktivität pro Geschäftspartner zu modellieren. Erst in den Verfeinerungen sollen die Details der internen Geschäftsprozesse modelliert werden. Hierdurch lässt sich aus der ersten Ebene der Bedarf an Kollaborationen deutlicher ablesen, während weitere Ebenen zur Erstellung von innerbetrieblichen Orchestrierungen verwendet werden können. Ein Beispiel hierfür ist in Abbildung 4 zu sehen.

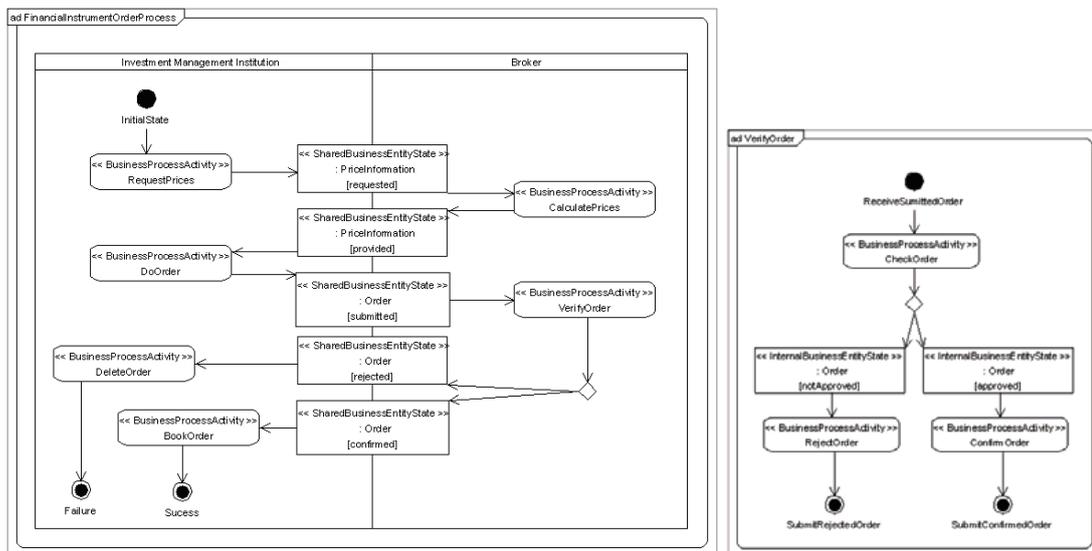


Abb. 4: Geschäftsprozessmodell mit Verfeinerung

Die zweite Erweiterung bezieht sich auf die semantische Hinterlegung der Business Entities, die im Business Process View benutzt werden. Wie beschrieben können die Information Entities, die im Business Transaction View benutzt werden, durch die Core Components mit einer einheitlichen Semantik hinterlegt werden. Für die Business Entities im Business Process View ist dies nicht vorgesehen. Für eine Anbindung von Aktivitäten an Services ist jedoch hauptsächlich der Informationsfluss interessant. Wir schlagen daher vor, eine Verbindung zwischen Information Entities und Business Entities zu schaffen. Da Information Entities an Business Information Entities gebunden werden können, wären damit auch die Business Entities semantisch hinterlegt und der Informationsfluss in der Business Process View semantisch beschrieben.

Als dritte Erweiterung schlagen wir vor, die Diagramme der Business Process View durch Diagramme in der Business Process Modeling Notation (BPMN) [OMG06] zu ersetzen. Ein Beispiel hierfür ist in Abbildung 5 zu sehen. Diese Variante hat folgende Vorteile:

- Während viele Konstrukte in BPMN Äquivalente in UML Aktivitätsdiagrammen haben, ist die BPMN, insbesondere was die Fehlerbehandlung und die Transaktionsverwaltung angeht, expliziter.
- BPMN wurde entwickelt, um Geschäftsprozessmodelle in die ausführbare Orchestrierungssprache BPEL [ACDG03] übersetzen zu können. Entsprechend gibt es eine im BPMN-Standard definierte Abbildung von BPMN auf BPEL.
- Während UML Aktivitätsdiagramme der OMG-Standard für die Modellierung von objektorientierter Software sind, ist die BPMN der OMG-Standard für die Geschäftsprozessmodellierung.
- Durch die speziellen Pfeile für den Nachrichtenfluss und die Markierung von Gruppen können notwendige Kollaborationen leichter identifiziert werden als im auf UML basierenden Business Process View.

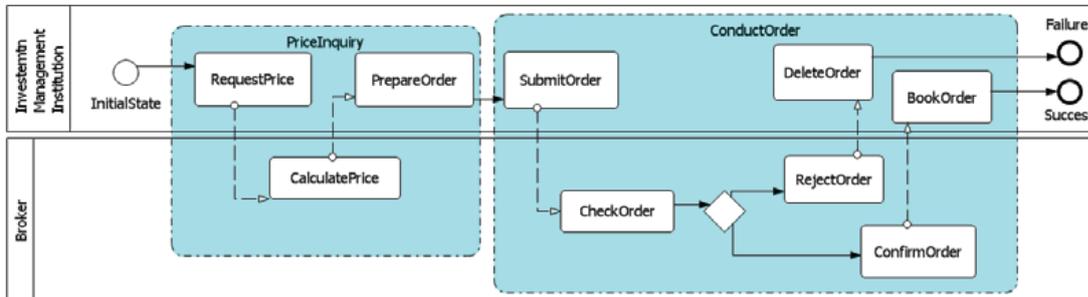


Abb. 5: Geschäftsprozessmodell in der BPMN

4.4 Einordnung der Erweiterung in GERAM

Es stellt sich die Frage, wie die erweiterte UMM die für eine SOA notwendigen Anforderungen der Elemente von GERAM umsetzt.

Für die EMOs ist es erforderlich, dass Web services als Module verwaltet werden können. Hierfür wird ein Web service Verzeichnis benötigt, das die Web services für die Benutzung in den EOS bereitstellt. Während ein solches Verzeichnis nicht Teil der beschriebenen Methodologie ist, gibt es Ansätze wie den UDDI-Standard [Newc02], solche Verzeichnisse zur Verfügung zu

stellen. Auch Verzeichnisse mit semantischen Erweiterungen wurden z. B. in [AVMM04] und [GrJH05] bereits vorgeschlagen.

GERAM-Elemente	Bestandteil	Umsetzung in erweiterter UMM
GERA	Mitarbeiterorientierte Konzepte	In Arbeit
	Prozessorientierte Konzepte	Fokus von erweiterter UMM
	Technologieorientierte Konzepte	Nicht unterstützt
EEM		Methodologie zur Erstellung der Modelle von UMM übernommen
EMLs	Jeder Bereich ist modellierbar	Wird durch UMM Modelle gewährleistet
	Modelle sind miteinander integrierbar	Wird durch UMM Modelle gewährleistet
EETs		Prototyp in Metamodellierungstool entwickelt
PEMs	Partial Human Role Models	Entsprechende Konzepte sind in Arbeit
	Partial Process Models	Referenzmodell-Verzeichnis wird erstellt
	Partial Technology Models	Nicht betrachtet
	Partial Models of IT systems	Insbesondere Web service Verzeichnis
GEMCs		Core Components
EMs		In der Methodologie erstellte Modelle
EOS		Das modellierte Unternehmen
EMOs		Web services

Tab. 1: Erweiterte UMM im Rahmen von GERAM

EMLs sollten insbesondere Geschäftsprozessmodelle vorsehen, die für eine automatische Orchestrierung und Choreographie verwendet werden können. Während mit der UMM die Erstellung einer in BPSS beschriebenen Choreographie möglich ist, gestattet die von uns vorgeschlagene Erweiterung auch die Erstellung von Orchestrierungen. In diesem Bereich werden die Anforderungen an EMLs für eine SOA gut erfüllt.

Im Bereich der EETs haben wir mit Hilfe eines Metamodellierungstools ein Modellierungstool entwickelt, welches alle für die erweiterte UMM notwendigen Modelle abbilden kann. Im Weiteren werden wir eine Core Components Bibliothek im Modellierungstool als GEMC anbinden, sodass der in den Geschäftsprozessmodellen dargestellte Informationsfluss semantisch hinterlegt werden kann. Dies ist notwendig, um eine automatische Bindung von Services an Aktivitäten zu ermöglichen.

Im Bereich der PEMs ist es möglich, Referenzprozesse zu hinterlegen, die bereits durch EMOs unterstützt werden können. Genauso wie für das Web Service Verzeichnis müsste auch für Referenzprozesse ein Verzeichnis geschaffen werden.

Die EEM wird von der UMM übernommen und ist bereits gut für SOA geeignet, da ein starker Fokus auf den Geschäftsprozessen und ihrer Unterstützung durch EDV liegt. Eine Erweiterung ist hier nicht notwendig.

Die Einordnung der Erweiterung in GERAM ist in Tabelle 1 zusammengefasst.

5 Schlussfolgerungen

Die Schlussfolgerungen beziehen sich zum einen auf die Erweiterung der UMM, zum anderen auf die Verwendung von GERAM als Referenzmodell für ein SOA Framework.

5.1 Erweiterung der UMM

Der Beitrag zeigt, dass die erweiterte UMM in vielen Bereichen der GERAM die durch eine SOA gestellten Anforderungen schon adressiert. Im Bereich von GERA und EML sollten zusätzlich für SOA notwendige Rollen und Lebenszykluskonzepte mit betrachtet werden. Wir arbeiten zurzeit an weiterreichende Modelle für Rollen und Technologien, die den gesamten Service-Lebenszyklus abdecken. Im Bereich der PEMs und EMOs erarbeiten wir im Zusammenhang mit einem Forschungsprojekt geeignete Lösungen in Form von Verzeichnissen, die auch in unser Modellierungstool eingebunden werden.

Durch die Einordnung der Methodologie in die GERAM ist es möglich, Schwachstellen und Stärken der Methodologie besser zu erkennen und somit gezielter zu einer ganzheitlichen Unternehmensmodellierung für eine SOA zu kommen.

5.2 Enterprise Architecture Frameworks für serviceorientierte Architekturen

GERAM stellt einen guten Rahmen zur Verfügung, in dem ein EAF für SOA entwickelt werden kann. Denn die für eine SOA notwendige ganzheitliche Sicht auf ein Unternehmen wird durch die von GERAM spezifizierten Elemente abgedeckt. Ein in diesem Rahmen entwickeltes Framework hat gute Voraussetzungen, die Einführung einer SOA und ihre Verwaltung zu erleichtern.

Wir werden im Rahmen der GERAM weiter an einem Framework arbeiten, dass die Belange einer SOA berücksichtigt. Die hier vorgestellte Erweiterung der UMM ist dabei nur ein erster Schritt zur Modellierung einer auf Services aufbauenden Unternehmensstruktur.

Literaturverzeichnis

- [AVMM04] *Aggarwal, R.; Verma, K.; Miller, J.; Milnor, W.*: Constraint driven Web service composition in METEOR-S. In: Services Computing, 2004. (SCC 2004). Proceedings. 2004 IEEE International Conference on (2004), S. 23-30.
- [ACDG03] *Andrews, Tony; Curbera, Francisco; Dholakia, Hitesh; Golland, Yaron; Klein, Johannes; Leymann, Frank; Liu, Kevin; Roller, Dieter; Smith, Doug; Trickovic, Ivana; Weerawarana, Sanjiva*: Business Process Execution Language for Web Services Version 1.1. <ftp://www6.software.ibm.com/software/developer/library/ws-bpel.pdf>, 2003, Abruf am 2006-06-12.
- [BEAS06] *BEA Systems*: Dev2Dev Online: Service-oriented Architecture. <http://dev2dev.bea.com/soa/>, 2006, Abruf am 2006-06-21.
- [BeNe97] *Bernus, Péter; Nemes, L.*: The Contribution of the Generalised Enterprise Reference Architecture to Consensus in the Area of Enterprise Integration. <http://www.cit.gu.edu.au/~bernus/publications/articles/iceimt97/iceimt97geram.pdf>, 1997, Abruf am 2004-06-12.
- [BrHW04] *Braun, Christian; Hafner, Martin; Wortmann, Felix*: Methodenkonstruktion als wissenschaftlicher Erkenntnisansatz. http://www.alexandria.unisg.ch/EXPORT/DL/Martin_Hafner/28306.pdf, 2004, Abruf am 2006-11-10.
- [Diet06] *Dietrich, Jens*: Nutzung von Modellierungssprachen und -methodologien standardisierter B2B-Architekturen für die Integration unternehmensinterner Geschäftsprozesse (unpublizierte Dissertation). TU Berlin, Fachgebiet Systemanalyse und EDV, 2006.

- [Food05] *Foody, Dan*: Getting web service granularity right. <http://www.soa-zone.com/index.php?/archives/11-Getting-web-service-granularity-right.html>, 2005, Abruf am 2006-06-12.
- [Gome06] *Gomes, Lee*: For business-software developers, flexibility can be spelled SOA. In: *The Wall Street Journal Europe* July 18, 2006 (2006), S. 32.
- [GrJH05] *Grønmo, R.; Jaeger, M. C.; Hoff, H.*: Transformations between UML and OWL-S. In: *European Conference on Model Driven Architecture–Foundations and Applications (ECMDA'05)* (2005).
- [IFIF99] *IFAC/IFIP: Integration, I. I. T. F. o. A. f. E.* (Hrsg.): Generalised Enterprise Reference Architecture and Methodology Version 1.6.3. <http://www.cit.gu.edu.au/~bernus/taskforce/geram/versions/geram1-6-3/v1.6.3.html>, 1999, Abruf am 2004-09-11.
- [IBMC06] *International Business Machines Corporation*: developerWorks : SOA and Web services. <http://www-128.ibm.com/developerworks/webservices>, 2006, Abruf am 2006-06-21.
- [ISO00] *ISO*: Industrial Automation Systems - Requirements for Enterprise Reference Architectures and Methodologies. In: *ISO 15704* (2000).
- [LeRS02] *Leymann, F.; Roller, D.; Schmidt, M. T.*: Web services and business process management. In: *IBM Systems Journal* 41 (2002) Nr. 2, S. 198-211.
- [McNa03] *McCoy, D.; Natis, Y.*: Service-Oriented Architecture: Mainstream Straight Ahead. In: *Gartner Research* (2003) Nr. LE-19-7652.
- [Micr06] *Microsoft Corporation*: .NET Architecture Center: Service Oriented Architecture. <http://msdn.microsoft.com/architecture/soa/>, 2006, Abruf am 2006-06-21.
- [Newc02] *Newcomer, Eric*: Understanding Web Services: XML, WSDL, SOAP and UDDI. Addison-Wesley Longman, Amsterdam 2002.

- [NeLo05] *Newcomer, Eric; Lomow, Greg*: Understanding SOA with Web Services. Addison-Wesley, Upper Saddle River, NJ 2005.
- [OASI06] *OASIS Open*: OASIS Committees by Category: SOA. http://www.oasis-open.org/committees/tc_cat.php?cat=soa, 2006, Abruf am 2006-06-21.
- [OMG06] *Object Management Group*: Business Process Modeling Notation Specification. <http://www.omg.org/cgi-bin/doc?dtc/2006-02-01>, 2006, Abruf am 2006-06-12.
- [ReUn06] *Research Studio Austria; University of Vienna*: UMM Add-In. <http://www.ifs.univie.ac.at/ummaddin/>, 2006, Abruf am 2006-06-12.
- [SAPA06] *SAP AG*: SAP - Enterprise Service-Oriented Architecture: Blueprint for Service-Based Business Solutions. <http://www.sap.com/platform/esa/index.epx>, 2006, Abruf am 2006-06-21.
- [ScRo03] *Schallert, M. ; Rosemann, M.*: Issues in the design of enterprise architectures. In: GI-Arbeitskreis EA Frühjahrskonferenz 2003. Universität St. Gallen, St. Gallen 2003, S. 42-49.
- [Sche04] *Schekkerman, Jaap*: How to survive in the jungle of Enterprise Architecture Frameworks. Trafford, Victoria, Canada 2004.
- [Schm05] *Schmelzer, Ronald*: Solving the Service Granularity Challenge. <http://www.zapthink.com/report.html?id=ZAPFLASH-200639>, 2005, Abruf am 2006-06-12.
- [Schö04] *Schönherr, Marten*: Enterprise Architecture Frameworks. In: Aier, S., Schönherr, M. (Hrsg.): Enterprise Application Integration - Serviceorientierung und nachhaltige Architekturen. Gito, Berlin 2004, S. 3-48.
- [Spar06] *Sparx Systems Pty Ltd.*: UML tools for software development and Modelling - Enterprise Architect Full Lifecycle UML modeling tool. <http://www.sparxsystems.com/>, 2006, Abruf am 2006-06-12.
- [Szyp98] *Szyperski, C.*: Component Oriented Programming. Springer, 1998.

- [UNCE03] *UN/CEFACT: UN/CEFACT – ebXML Business Process Specification Schema.*
http://www.untmg.org/index.php?option=com_docman&task=view_category&Itemid=137&subcat=3&catid=63&limitstart=0&limit=5, 2003, Abruf am 2006-06-12.
- [UNCE06a] *UN/CEFACT: UN/CEFACT's Modeling Methodology (UMM).*
http://www.untmg.org/index.php?option=com_docman&task=view_category&Itemid=137&subcat=1&catid=63&limitstart=0&limit=5, 2006, Abruf am 2006-06-12.
- [UNCE06b] *UN/CEFACT: ISO\DTS 15000-5: 2006 Core Components Technical Specification 2nd Edition UN/CEFACT Version 2.2.*
http://www.untmg.org/index.php?option=com_docman&task=docclick&Itemid=137&bid=43&limitstart=0&limit=5, 2006, Abruf am 2006-06-12.
- [WWWC04] *World Wide Web Consortium: Web Services Architecture.*
<http://www.w3.org/TR/ws-arch/>, 2004, Abruf am 2006-06-12.

Einführung in den Track

Business Intelligence

Prof. Dr. Peter Chamoni

Universität Duisburg-Essen

Prof. Dr. Bodo Rieger

Universität Osnabrück

Einreichungen sollen originäre Arbeiten beschreiben, entweder über innovative Methoden, Modelle und Konzepte des Business Intelligence (BI), oder deren praktische Anwendung im Rahmen des integrativen Szenario der Management Support Systeme (MSS). Dies umfasst insbesondere Forschungen über die effektive Anwendung von Mining- und anderer geeigneter, analytischer Technologien auf betriebliche Entscheidungsprobleme, die mehrdimensionale Überwachung der Unternehmens-Performance (Scorecard Modelle), die aktive Beratung von Entscheidern durch modellgestützte, bewertete Handlungsalternativen, und das Sammeln, Verteilen und kooperative Nutzen zugehöriger, qualitativer Informationen im Sinne des Wissensmanagement. Aus Sicht der praktischen Umsetzung sind Konzepte und bewährte, verteilte Multi-layer-Architekturen zur Integration von Data Warehouses, modellgestützten, analytischen Applikationen und individuell oder automatisch anpassbaren Anwenderportalen willkommen.

Programmkomitee:

Dr. Wolfgang Behme, Continental AG, Hannover
Prof. Dr. Peter Chamoni, Universität Duisburg-Essen (track chair)
Dr. Eitel von Maur, Universität St. Gallen
Prof. Dr. Roland Gabriel, Universität Bochum
PD Dr. Peter Gluchowski, Universität Düsseldorf
Prof. Dr. Norbert Gronau, Universität Potsdam
Prof. Dr. Dimitris Karagiannis, Universität Wien
Prof. Dr. Hans-Georg Kemper, Universität Stuttgart
Prof. Dr.-Ing. Peter Lehmann, Hochschule der Medien Stuttgart
Dr. Wolfgang Martin, Wolfgang Martin Team, Annecy, France
Dr. Harry Mucksch, IT-Beratung und Services, Apen
Prof. Dr. Bodo Rieger, Universität Osnabrück (track chair)
Prof. Dr. Wolfgang Uhr, TU Dresden

Kollaboratives Data Warehousing

Konzeption und prototypische Realisierung flexibler Schema- und Datenintegration

Thomas Matheis, Dirk Werth, Peter Loos

Institut für Wirtschaftsinformatik (IWi)
im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI)
66123 Saarbrücken
{thomas.matheis, dirk.werth, peter.loos}@iwi.dfki.de

Abstract

Die Bedeutung von Geschäftsmodellen des collaborative Business ist gewachsen. Kollaboratives Data Warehousing ermöglicht es den partizipierenden Unternehmen, sich in einem steigenden Wettbewerb strategische Wettbewerbsvorteile zu sichern, und sich damit in den stark umkämpften Kundenmärkten zu behaupten. Vor diesem Hintergrund werden Ansätze zum kollaborativen Data Warehousing vermehrt an Bedeutung gewinnen. Der vorliegende Beitrag stellt auf der Grundlage von föderierten Datenbanktechnologien einen Integrationsansatz zum kollaborativen Data Warehousing vor. Der Ansatz bezieht sowohl die Schemaebene, als auch die Datenebene in den Integrationsprozess ein, und erlaubt es insbesondere, dass ein beliebiges Unternehmen wieder aus der Kollaboration de-integriert werden kann. Über die Konzeption hinausgehend präsentiert der Beitrag auch einen Software-Prototypen, der für den Aufbau, den Betrieb und die Auflösung kollaborativer Data Warehouses eingesetzt werden kann.

1 Einleitung

Die fortschreitende Auflösung von Grenzen innerhalb und zwischen Unternehmen, sowie das Fortschreiten des technologischen Wandels, insbesondere der Internettechnologien, hat in den letzten Jahren zunehmend Einfluss auf die Gestaltung wertschöpfender Geschäftsprozesse ausgeübt [Sche+03]. Die dadurch entstandenen neuen Geschäftsmodelle werden unter dem

Schlagwort collaborative Business (c-Business) zusammengefasst und beinhalten Konzepte für die Zusammenarbeit über Unternehmensgrenzen hinweg [RöSc01, S. 289-292]. Im Mittelpunkt steht dabei die effiziente und effektive Gestaltung wertschöpfender Geschäftsprozesse, die nicht mehr nur unternehmensintern, sondern vor allem unternehmensübergreifend betrachtet werden. Die zunehmende Integration von einzelnen Unternehmen in global strategische Netzwerke ist eine globale Tendenz und eröffnet den beteiligten Unternehmen enorme Chancen. Nicht mehr einzelne Unternehmen werden künftig miteinander konkurrieren, sondern Netzwerke einzelner Unternehmen. Die Fähigkeit von Unternehmen, sich in kollaborative Netzwerke zu integrieren, stellt daher einen entscheidenden Schlüssel zu ihrem Geschäftserfolg dar [Öste+00]. Data-Warehouse-Systeme sind als Kern entscheidungs-unterstützender Informationssysteme für viele Unternehmen von strategischer Bedeutung [Lint01, S. 47f]. Da Informationen für Unternehmen immer wichtiger werden, um in Zeiten der Globalisierung auf dem Markt schnell reagieren zu können, ist das Vorhandensein aussagekräftiger Fakten für die strategische Entscheidungsfindung von großer Relevanz [AlÖs01]. Traditionelle Data-Warehouse-Lösungen sind in der Regel nur auf die Entscheidungsfindung einzelner Unternehmen ausgerichtet. Eine kollaborative Data-Warehouse-Lösung vereint die Daten mehrerer Unternehmen, die ihre Unternehmensdaten der Kollaboration über einzelne Data-Warehouse-Lösungen zur Verfügung stellen. Den an der Kollaboration beteiligten Unternehmen bietet eine kollaborative Data-Warehouse-Lösung bedeutende Möglichkeiten, um unternehmensübergreifende Entscheidungen zu treffen und strategische Wettbewerbsvorteile gegenüber ihren Konkurrenten zu erzielen [MaWe05].

Dieser Beitrag stellt auf der Grundlage von föderierten Datenbanktechnologien einen Ansatz für die flexible Integration von Data-Warehouse-Lösungen zu einer Data-Warehouse-Kollaboration vor. Kapitel 2 beschreibt die Methodik zur Integration multidimensionaler Datenmodelle, die die Grundlage des kollaborativen Data Warehousings bildet. Anschließend wird in Kapitel 3 die prototypische Realisierung des Integrationsansatzes vorgestellt. Der Beitrag schließt in Abschnitt 4 mit einer Zusammenfassung und einem Ausblick auf weitere Forschungsaufgaben.

2 Methodik zur Integration multidimensionaler Datenmodelle

Die Thematik der Schemaintegration ist bereits gut durchdrungen [Böhn01], so dass bei der Entwicklung einer Methode zur Integration multidimensionaler Datenmodelle ein kompletter

Neuentwurf nicht notwendig ist. Vielmehr kann ein bereits existierender Integrationsansatz um multidimensionale Konstrukte erweitert werden. Beispielhaft seien zur Schemaintegration Ansätze wie Upward Inheritance [Conr97], Correspondence Assertion [Spac+92] oder Generic Integration Model [Schm98] genannt. Für eine ausführliche Darstellung über Grundideen und Ansätze weiterer Integrationstechniken sei an dieser Stelle auf [RaBe01; Bati+86] verwiesen. Die zusicherungs-basierte Integration (Correspondence Assertion) ist unter den bestehenden Integrationsansätzen am ehesten geeignet, um als Grundlage für die Integration von multidimensionalen Datenmodellen zu dienen [MaWe05]. Daher wird in diesem Kapitel eine Methode für die Integration multidimensionaler Datenschemata vorgestellt, die auf dem Prinzip der zusicherungs-basierten Integration basiert. Bei diesem Ansatz werden auf der Grundlage eines generischen Datenmodells zwischen den zu integrierenden Schemata Inter-Schema-Korrespondenzen in Form von Zusicherungen definiert. Durch Anwendung von Integrationsregeln kann unter Einbezug der Zusicherungen das integrierte Schema konstruiert werden.

2.1 Vorüberlegungen und Voraussetzungen

Ein kollaboratives Data-Warehouse-System erfordert die Integration von mehreren Data Warehouses und somit die Integration von mehreren multidimensionalen Datenschemata. Ein wesentlicher Unterschied zwischen multidimensionalen Datenschemata und „einfachen“ Datenschemata besteht im Konzept der Summierbarkeit [LeSh97, S. 132-143]. Zentrales Ziel der Summierbarkeit ist es, die Korrektheit von Ergebnissen von Aggregatanfragen über multidimensionalen Daten zu garantieren. Beispielsweise kann die von den meisten multidimensionalen Datenmodellen verwendete Roll-Up-Beziehung, die eine Aggregation der Daten auf eine höhere Granularität beschreibt, nur dann sinnvoll eingesetzt werden, wenn eine korrekte Summierbarkeit vorausgesetzt werden kann [Lehn03]. Auch wenn vorausgesetzt wird, dass die lokalen Data-Warehouse-Datenquellen eine korrekte Summierbarkeit gewährleisten, das heißt die Bedingungen der Disjunktheit, Vollständigkeit und Typverträglichkeit erfüllen, so wird durch die Integration von mehreren multidimensionalen Schemata die Bedingung der Summierbarkeit in der Regel verletzt. Die dadurch entstehende Problematik im Rahmen der Anfragebearbeitung wird im Folgenden anhand des Beispiels der Abbildung 1 näher erläutert.

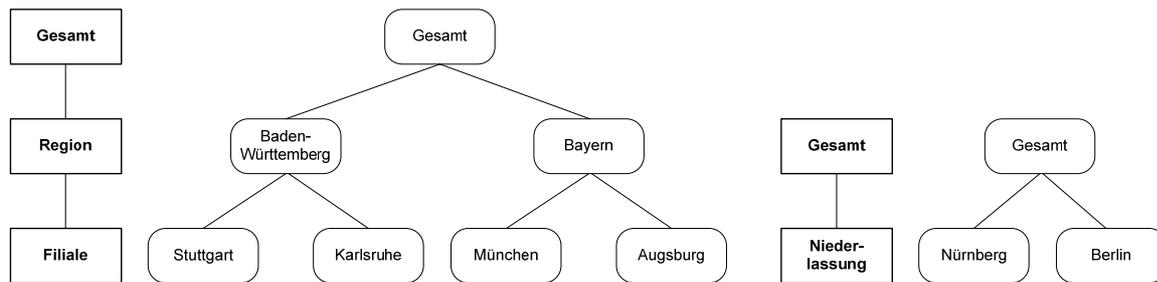


Abbildung 1: Zu integrierende multidimensionale Schemata

Die beiden Schemata der Abbildung 1 weisen die semantisch äquivalenten Schemaelemente 'Filiale' und 'Niederlassung' auf. Das Schemaelement 'Filiale' bzw. 'Niederlassung' wird somit in das integrierte Schema aufgenommen. Auf Instanzebene weist das integrierte Schema alle Instanzen der beiden Schemata auf der Ebene der Filialen bzw. Niederlassungen auf. Aufgrund der funktionalen Abhängigkeit [Lehn03] bestimmt beispielsweise die Instanz 'Stuttgart' die übergeordnete Instanz 'Baden Württemberg' funktional. Es ist allerdings unklar, welche Instanz des Attributes 'Region' von den Instanzen 'Nürnberg' und 'Berlin' funktional bestimmt werden, da das Schemaelement 'Region' im rechten Schema nicht enthalten ist. Diese Information muss bei der Integration erfasst werden, um das Konzept der Summierbarkeit, insbesondere der Vollständigkeit, zu gewährleisten. Die Vollständigkeit kann wieder hergestellt werden, indem die entsprechenden Instanzen den Instanzen der nächsthöheren Hierarchieebene direkt zugeordnet werden beziehungsweise entsprechende Instanzen der nächsthöheren Hierarchieebene neu definiert werden. In diesem Fall ergibt sich folgendes integriertes Schema mit den entsprechenden Instanzen, das in Abbildung 2 dargestellt ist.

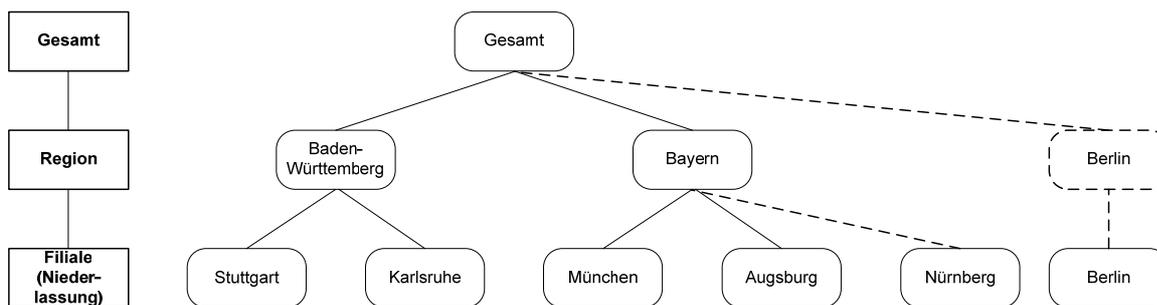


Abbildung 2: Integriertes Schema

Die Zuordnung einer Instanz zu einer Instanz der nächsthöheren Hierarchieebene muss meist manuell erfolgen, da hierfür semantisches Wissen erforderlich ist, das in der Regel nicht automatisch abgeleitet werden kann. Das wesentliche Ziel eines Data-Warehouse-Systems beziehungsweise eines kollaborativen Data-Warehouse-Systems, nämlich die Bereitstellung von aussagekräftigen Informationen zur Entscheidungsunterstützung, kann allerdings nur durch

diese Zuordnung der dimensionalen Instanzen erfüllt werden, da sonst keine aussagekräftigen Informationen bereitgestellt werden können.

Die oben dargestellte Problematik kommt besonders in einem kollaborativen Data-Warehouse-System zu tragen. Da an der Kollaboration eine Vielzahl von Unternehmen teilnehmen, die sich in die Kollaboration integrieren und auch wieder de-integrieren können, kann nicht davon ausgegangen werden, dass diese dieselben Schemaelemente in ihren multidimensionalen Schemata abgebildet haben. Um das Konzept der Summierbarkeit besser berücksichtigen zu können, bezieht der hier vorgeschlagene Integrationsansatz daher nicht nur die Schemaebene, sondern auch die Instanzebene der Dimensionshierarchien in den Integrationsprozess ein. Dabei spielt insbesondere das Konzept der funktionalen Abhängigkeit eine wesentliche Rolle. Durch Einbezug der Instanzebene in den Integrationsprozess erhöht sich zwar der Integrationsaufwand. Allerdings kann nur dadurch gewährleistet werden, dass im Rahmen einer Anfragebearbeitung aussagekräftige Anfrageergebnisse auf der Ebene der Kollaboration erzielt werden können. Durch Einschränkungen der zu integrierenden Daten kann der Integrationsaufwand wesentlich verringert und die praktische Anwendbarkeit des Ansatzes erhöht werden [MaWe05].

Die hier vorgestellte Methode beschränkt sich auf die Integration von multidimensionalen Schemata, die durch Fakten (Basiskennzahlen) und Dimensionen mit einfacher Hierarchie modelliert sind. Auf die Darstellung von abgeleiteten Kennzahlen, nicht dimensional Attributen und Dimensionen mit parallelen Hierarchiepfaden wurde bewusst verzichtet. Diese Konstrukte werden vom vorgestellten Integrationsansatz jedoch nicht ausgeschlossen, sondern können mit Hilfe weiterer Integrationsregeln ergänzt werden. Die Entscheidung für die Einschränkung liegt darin, dass sich mit Fakten und Dimensionen einfacher Hierarchie bereits die wesentlichen Konzepte zur Integration multidimensionaler Datenmodelle zeigen lassen.

2.2 Datenmodell

Um unabhängig von einem konkreten Datenmodell zu sein, werden die Zusicherungen bei der zusicherungs-basierten Integration auf der Grundlage des generischen Datenmodells formuliert. Das generische Datenmodell besitzt für multidimensionale Datenmodelle jedoch eine zu geringe Ausdrucksstärke. Daher wird im Folgenden in Anlehnung an die Arbeit von [Lehn03] ein multidimensionales Datenmodell beschrieben, das als Grundlage für die Formulierung der Zusicherungen und der Integrationsregeln dient.

An dieser Stelle muss zunächst das Prinzip der funktionalen Abhängigkeit erläutert werden, welches in den nachfolgenden Beschreibungen verwendet wird.

Definition Funktionale Abhängigkeit: Zwischen zwei Attributen A und B existiert eine funktionale Abhängigkeit ($A \rightarrow B$) genau dann, wenn für jede Instanz $a \in A$ genau eine Instanz $b \in B$ existiert. Das Attribut B wird damit von dem Attribut A funktional bestimmt.

Definition Schwache funktionale Abhängigkeit: Zwischen zwei Attributen A und B existiert eine schwache funktionale Abhängigkeit ($A \Rightarrow B$) genau dann, wenn für jede Instanz $a \in A$ höchstens eine Instanz $b \in B$ existiert.

Definition Datenwürfel: Das Schema eines Datenwürfels C besteht aus einer Menge von dimensionalen Schemata D und einer Menge von Fakten F. Dabei gilt: $C = (D, F) = (\{D_1, \dots, D_n\}, \{F_1, \dots, F_n\})$.

Definition Dimension: Das Schema einer Dimension D besteht aus einer geordneten Menge von dimensionalen Attributen ($\{A_1, \dots, A_n, \text{TopA}\}; \rightarrow$), wobei \rightarrow die funktionale Abhängigkeit bezeichnet und TopA ein generisches maximales Element in Bezug auf \rightarrow darstellt. Für zwei dimensionale Attribute A_i und A_{i+1} gilt, dass A_{i+1} von A_i funktional bestimmt wird. TopA wird von allen dimensionalen Attributen funktional bestimmt. Das dimensionale Attribut A_1 bestimmt alle anderen dimensionalen Attribute und stellt somit die feinste Granularität einer Dimension dar. Dabei gilt:

- $\forall i (1 \leq i < n): A_i \rightarrow A_{i+1}$
- $\forall i (1 \leq i \leq n): A_i \rightarrow \text{TopA}$
- $\exists i (1 \leq i \leq n) \forall j (1 \leq j \leq n), i \neq j: A_i \rightarrow A_j$

Die Definition einer Dimension bezieht neben der Schemaebene die Instanzebene nur indirekt über die funktionale Abhängigkeit ein. Um auch die Instanzebene besser berücksichtigen zu können, wird die Definition einer Dimension wie folgt erweitert. Jedes dimensionale Attribut A_i besteht aus einer Menge von Instanzen: $A_i = \{a_1, \dots, a_n\}$.

Zu beachten ist an dieser Stelle, dass die Definition eines Datenwürfels damit nicht nur die Schemaebene, sondern auch die Instanzebene der dimensionalen Attribute einbezieht.

2.3 Korrespondenz-Zusicherungen

Korrespondenz-Zusicherungen setzen die Bestandteile von Datenwürfeln zueinander in Beziehung. Es werden zwei Arten von Korrespondenz-Zusicherungen unterschieden.

Zusicherungen, die über Attribute Dimensionen in Beziehung zueinander setzen, sowie Zusicherungen, die Fakten in Beziehung zueinander setzen. Im Folgenden werden die verschiedenen Korrespondenz-Zusicherungen vorgestellt.

Dimension-Korrespondenz-Zusicherung: Seien A und B zwei Dimensionen, wobei A aus dem Datenwürfel C1 und B aus dem Datenwürfel C2 stammt. Es gilt: $A = (\{A_1, \dots, A_n, \text{TopA}\}; \rightarrow)$ und $B = (\{B_1, \dots, B_n, \text{TopB}\}; \rightarrow)$. Dann gibt es zwei Möglichkeiten, um Korrespondenzen zwischen A und B durch Zusicherungen zu formulieren.

- $A \leftrightarrow B$: Die Dimensionen A und B repräsentieren eine semantisch äquivalente Dimensionshierarchie.
- $A \uparrow$: Die Dimension A repräsentiert eine Dimensionshierarchie, die zu keiner Dimension des Datenwürfels C2 semantisch äquivalent ist. Diese Zusicherung ermöglicht es, neue Dimensionen in den integrierten Datenwürfel aufzunehmen beziehungsweise einzelne Attribute einer Dimension von A in eine neue Dimension aufzunehmen.

Attribut-Korrespondenz-Zusicherung: Seien A_i und B_j zwei beliebige Attribute, wobei A_i aus der Dimension A und B_j aus der Dimension B stammt. Dann gibt es die folgenden Möglichkeiten, um Korrespondenzen zwischen A_i und B_j durch Zusicherungen zu formulieren.

- $A_i \leftrightarrow B_j$: Die Attribute A_i und B_j sind semantisch äquivalent, das heißt sie repräsentieren die gleiche Menge von Instanzen. Die Menge D bezeichnet dabei alle Instanzen der Attribute A_i und B_j , die zueinander semantisch äquivalent sind. Es gilt: $D = \{(a,b) \mid a \in A_i, b \in B_j, a \text{ semantisch äquivalent zu } b\}$.
- $A_i \rightarrow B_j$: Das Attribut B_j wird von dem Attribut A_i direkt funktional bestimmt. Das Attribut A_i repräsentiert also eine Menge von Instanzen, die die Menge von Instanzen von B_j direkt funktional bestimmen.
- $A_i \rightarrow \text{TopB}$: TopB wird von dem Attribut A_i direkt funktional bestimmt. Das Attribut A_i repräsentiert also eine Menge von Instanzen, die von allen Mengen von Instanzen der Attribute aus B funktional bestimmt werden.
- $A_i \uparrow$: Das Attribut A_i lässt sich nicht durch eine der vorangegangenen Zusicherungen einem Attribut der Dimension B zuordnen.

Die Attribut-Zusicherungen können den Dimension-Zusicherungen wie folgt angefügt werden:

- $A \leftrightarrow B$ mit $A_i \leftrightarrow B_j$, $A_i \rightarrow B_j$, $A_i \rightarrow \text{Top}B$ oder $A_i \updownarrow$
- $A \updownarrow$ mit $A_i \leftrightarrow A_i$ oder $A_i \updownarrow$

Fakten-Korrespondenz-Zusicherung: Seien $F1$ und $F2$ zwei Fakten, wobei $F1$ aus dem Datenwürfel $C1$ und $F2$ aus dem Datenwürfel $C2$ stammt. Dann gibt es zwei Möglichkeiten, um Korrespondenzen zwischen $F1$ und $F2$ durch Zusicherungen zu formulieren.

- $F1 \leftrightarrow F2$: Die Fakten $F1$ und $F2$ repräsentieren eine semantisch äquivalente Basiskennzahl.
- $F1 \updownarrow$: Das Faktum $F1$ repräsentiert eine Basiskennzahl, die zu keinem Faktum des Datenwürfels $C2$ semantisch äquivalent ist.

2.4 Integration

Mit Hilfe der folgenden Integrationsregeln kann der integrierte Datenwürfel basierend auf den vorgestellten Zusicherungen schrittweise konstruiert werden. Dabei werden sowohl Veränderungen auf Schema- als auch auf Instanzebene angegeben. Die Dimensionen der zu integrierenden Datenwürfel weisen eine funktionale Abhängigkeit auf, während die Dimensionen des integrierten Datenwürfels durch das Anwenden der Integrationsregeln in der Regel nur eine schwache funktionale Abhängigkeit aufweisen. Deshalb wird auch insbesondere aufgezeigt, welche Integrationsregeln die funktionale Abhängigkeit verletzen können.

Seien die Datenwürfel $C1$ und $C2$ gegeben. Der Datenwürfel $C1$ wird in den Datenwürfel $C2$, der zum Beispiel den bestehenden Datenwürfel einer Kollaboration beschreibt, integriert. Die nachfolgend angegebenen Integrationsregeln beschreiben, wie sich der Datenwürfel $C2$ durch die Integration von $C1$ verändert. Die Dimension A repräsentiert dabei eine Dimension des Datenwürfels $C1$, die Dimension B eine Dimension des Datenwürfels $C2$.

Integrationsregel 1: Alle Fakten, Dimensionen und Attribute des Datenwürfels $C2$, zu denen keine Korrespondenz zu dem Datenwürfel $C1$ besteht, bleiben durch die Integration des Datenwürfels $C1$ in den Datenwürfel $C2$ unverändert. Auch die Instanzen der Attribute, die nicht mit dem Datenwürfel $C1$ korrespondieren, bleiben durch die Integration unverändert erhalten.

Integrationsregel 2: Sei die Zusicherung $A \leftrightarrow B$ mit $A_i \leftrightarrow B_j$ und der Menge D gegeben. Die Menge DIFF_A bezeichnet die Menge von Instanzen von A_i , die nicht semantisch äquivalent zu Instanzen von B_j sind. Es gilt: $\text{DIFF}_A = \{a \mid a \in A_i, (a_d, b) \in D, a \neq a_d\}$. Analog bezeichnet die

Menge DIFF_B die Menge von Instanzen von B_j , die nicht semantisch äquivalent zu Instanzen von A_i sind. Es gilt: $\text{DIFF}_B = \{b \mid b \in B_j, (a, b_d) \in D, b \neq b_d\}$. Das Schema der Dimension B bleibt durch das Anwenden der Integrationsregel unverändert. Auf Instanzebene wird das Attribut B_j um alle Instanzen der Menge DIFF_A erweitert. Die funktionale Abhängigkeit kann bei dieser Integrationsregel in folgenden Fällen verletzt werden.

- Existiert eine Zusicherung der Form $A_{i+1} \rightarrow B_{j+1}$, so existieren für alle Instanzen aus B_j , die zu der Menge DIFF_B gehören, keine funktional abhängigen Instanzen aus dem übergeordneten Attribut A_{i+1} . Dies trifft ebenfalls zu, falls B_{j+1} TopB entspricht.
- Existiert ein Attribut B_{j+1} und keine Zusicherung der Form $A_{i+1} \leftrightarrow B_{j+1}$ oder $A_{i+1} \rightarrow B_{j+1}$, so existieren zu allen Instanzen aus B_j , die zu der Menge DIFF_A gehören, keine funktional abhängigen Instanzen aus B_{j+1} .

Integrationsregel 3: Sei die Zusicherung $A \leftrightarrow B$ mit $A_i \rightarrow B_j$ gegeben. Die Dimension B wird um das Attribut A_i erweitert, so dass B_j von A_i direkt funktional bestimmt wird. Alle Instanzen des Attributes A_i werden übernommen. Die funktionale Abhängigkeit kann bei dieser Integrationsregel in folgenden Fällen verletzt werden.

- Existiert keine weitere Zusicherung der Form $A_{i+1} \leftrightarrow B_j$ oder $A_{i+1} \rightarrow B_j$, so existieren für alle Instanzen aus A_i keine funktional abhängigen Instanzen aus B_j .
- Existiert ein Attribut B_{j-1} und keine Zusicherung Form $A_{h,h<i} \leftrightarrow B_{j-1}$, so existieren für alle Instanzen aus B_{j-1} keine funktional abhängigen Instanzen aus A_i . Ausnahme: Existiert zusätzlich eine Zusicherung der Form $A_{h,h<i} \leftrightarrow B_{k,k<j-1}$ mit DIFF_A und DIFF_B , so existieren nur für die Instanzen aus B_{j-1} , keine funktional abhängigen Instanzen aus A_i , die von den Instanzen der Menge DIFF_B funktional bestimmt werden.

Integrationsregel 4: Sei die Zusicherung $A \leftrightarrow B$ mit $A_i \rightarrow \text{TopB}$ gegeben. Die Dimension B wird um das Attribut A_i erweitert, so dass TopB von A_i direkt funktional bestimmt wird. Alle Instanzen des Attributes A_i werden übernommen. Die funktionale Abhängigkeit kann bei dieser Integrationsregel in folgendem Fall verletzt werden.

- Existiert keine weitere Zusicherung der Form $A_{h,h<i} \rightarrow \text{TopB}$ und keine Zusicherung der Form $A_{h,h<i} \leftrightarrow B_n$ für das Attribut B_n , das TopB direkt funktional

bestimmt, so besitzen alle Instanzen des Attributes B_n keine funktional abhängigen Instanzen aus A_i . Ausnahme: Existiert zusätzlich eine Zusicherung der Form $A_{h,h<i} \leftrightarrow B_{k,k<n-1}$ mit DIFF_A und DIFF_B , so existieren nur für die Instanzen aus B_n keine funktional abhängigen Instanzen aus A_i , die von den Instanzen der Menge DIFF_B funktional bestimmt werden.

Integrationsregel 5: Sei die Zusicherung $A \leftrightarrow B$ mit $A_i \downarrow$ gegeben. Die Dimension B bleibt auf Schema- und Instanzebene unverändert. Die Dimension B wird nicht um das Attribut A_i erweitert, da nur Dimensionen mit einfachen Hierarchiepfaden und somit keine parallelen oder alternativen Hierarchiepfade entstehen sollen. Das Konzept der funktionalen Abhängigkeit bleibt in diesem Fall erhalten.

Integrationsregel 6: Sei die Zusicherung $A \uparrow$ mit $A_i \leftrightarrow A_i$ beziehungsweise $A_i \downarrow$ gegeben. Der Datenwürfel C2 wird um die Dimension A erweitert. Das Attribut A_i wird im Fall von $A_i \leftrightarrow A_i$ mit in die Dimension übernommen, im Fall von $A_i \downarrow$ nicht in die Dimension mit aufgenommen. Das Konzept der funktionalen Abhängigkeit bleibt in diesem Fall erhalten.

Integrationsregel 7: Sei die Zusicherung $F1 \leftrightarrow F2$ gegeben. Der Datenwürfel C2 bleibt unverändert.

Integrationsregel 8: Sei die Zusicherung $F \uparrow$ gegeben. Der Datenwürfel C2 wird um das Faktum F2 erweitert.

Im Rahmen der Integration müssen neben den Zusicherungen ferner noch Abbildungsinformationen zwischen dem integrierten Datenwürfel und den lokalen Datenwürfeln festgehalten werden, die notwendig sind, um Anfragen an den integrierten Datenwürfel korrekt auf die einzelnen lokalen Datenwürfel zu transformieren. Die De-Integration eines Datenwürfels aus einem integrierten Datenwürfel, der nach dem vorgestellten Integrationsansatz konstruiert wurde, kann mit Hilfe von De-Integrationsregeln durchgeführt werden. Auf die Darstellung der zu erfassenden Abbildungsinformationen und der Methodik zur De-Integration wird an dieser Stelle nicht näher eingegangen.

2.5 Beispiel

Die eingeführten Zusicherungen und Integrationsregeln werden anhand des folgenden Beispiels verdeutlicht. Die Abbildung 3 zeigt die Schemata zweier Datenwürfel, die im multidimensionalen E/R-Modell (MERM) [Dete02] modelliert sind, sowie beispielhaft einige dimensionale Instanzen. Der Datenwürfel C1 soll in den Datenwürfel C2 integriert werden.

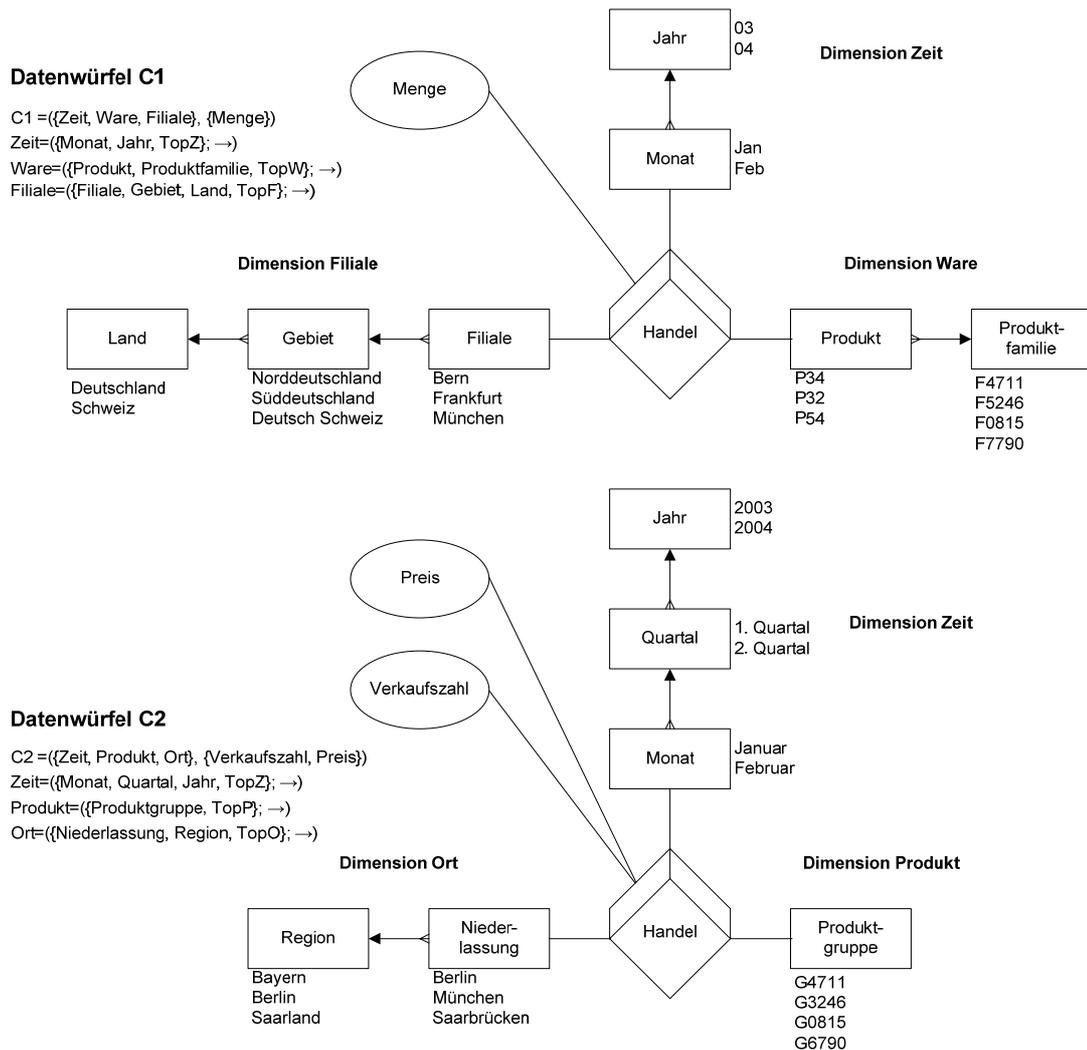


Abbildung 3: Zu integrierende Datenwürfel

Bevor die Integration durchgeführt werden kann, müssen zwischen den beiden Datenwürfeln die Korrespondenzen in Form von Zusicherungen bestimmt werden. Zwischen den beiden Datenwürfeln bestehen die folgenden Korrespondenz-Zusicherungen (siehe Abbildung 4).

- 1) $C1.Zeit \leftrightarrow C2.Zeit$ mit $C1.Monat \leftrightarrow C2.Monat$ und $D = \{(Jan, Januar), (Feb, Februar)\}$
- 2) $C1.Zeit \leftrightarrow C2.Zeit$ mit $C1.Jahr \leftrightarrow C2.Jahr$ und $D = \{(03, 2003), (04, 2004)\}$
- 3) $C1.Ware \leftrightarrow C2.Produkt$ mit $C1.Produkt \rightarrow C2.Produktgruppe$
- 4) $C1.Ware \leftrightarrow C2.Produkt$ mit $C1.Produktfamilie \leftrightarrow C2.Produktgruppe$ und $D = \{(F4711, G4711), (F0815, G0815)\}$
- 5) $C1.Filiale \leftrightarrow C2.Ort$ mit $C1.Filiale \leftrightarrow C2.Niederlassung$ und $D = \{(München, München)\}$
- 6) $C1.Filiale \leftrightarrow C2.Ort$ mit $C1.Gebiet \uparrow$
- 7) $C1.Filiale \leftrightarrow C2.Ort$ mit $C1.Land \rightarrow C2.TopO$
- 8) $C1.Menge \leftrightarrow C2.Verkaufszahl$

Abbildung 4: Korrespondenz-Zusicherungen

Mit Hilfe der Zusicherungen und unter Anwendung der Integrationsregeln kann nun der Datenwürfel C1 in den Datenwürfel C2 integriert werden. Die Integration wird hierbei am

Beispiel der Dimensionen Ort und Filiale verdeutlicht. Die Integrationsregel 2 wird auf die Zusicherung 5 angewandt. Das Schema der Dimension 'Ort' bleibt erhalten. Auf Instanzebene werden dem Attribut 'Niederlassung' die Instanzen 'Bern' und 'Frankfurt' zugefügt. Da das Attribut 'Region' existiert, aber keine Zusicherung der Form \leftrightarrow oder \rightarrow , die sich auf das Attribut 'Region' bezieht, existieren zu den Instanzen 'Frankfurt' und 'Bern' keine funktional abhängigen Instanzen des Attributes 'Region'. Durch Anwenden der Integrationsregel 5 auf die Zusicherung 6 bleibt der integrierte Datenwürfel auf Schema- und Instanzebene unverändert, da das Attribut 'Gebiet' nicht aufgenommen wird. Die Integrationsregel 4 verursacht bei Anwendung auf die Zusicherung 7 folgende Änderungen. Die Dimension 'Ort' wird um das Attribut 'Land' erweitert. 'TopO' wird nun von dem Attribut 'Land' direkt funktional bestimmt. Alle Instanzen des Attributes 'Land' werden übernommen. Da keine weitere Zusicherung für 'TopO' und keine Zusicherung der Form \leftrightarrow für das Attribut 'Region' existiert, besitzen alle Instanzen des Attributes 'Region' keine funktional abhängigen Instanzen aus dem Attribut 'Land'. Da aber zusätzlich die Zusicherung 5 besteht, tritt die Ausnahme der Integrationsregel 4 ein, so dass nicht alle Instanzen des Attributes 'Region', sondern nur die Instanzen 'Berlin' und 'Saarbrücken' keine funktional abhängigen Instanzen aus dem Attribut 'Land' besitzen'. Das Ergebnis der Integration ist in der Abbildung 5 dargestellt. Die in der Abbildung 5 kursiv dargestellten Bestandteile stellen synonyme Bezeichnungen dar, die im Rahmen der Integration durch die Zusicherungen aufgenommen werden.

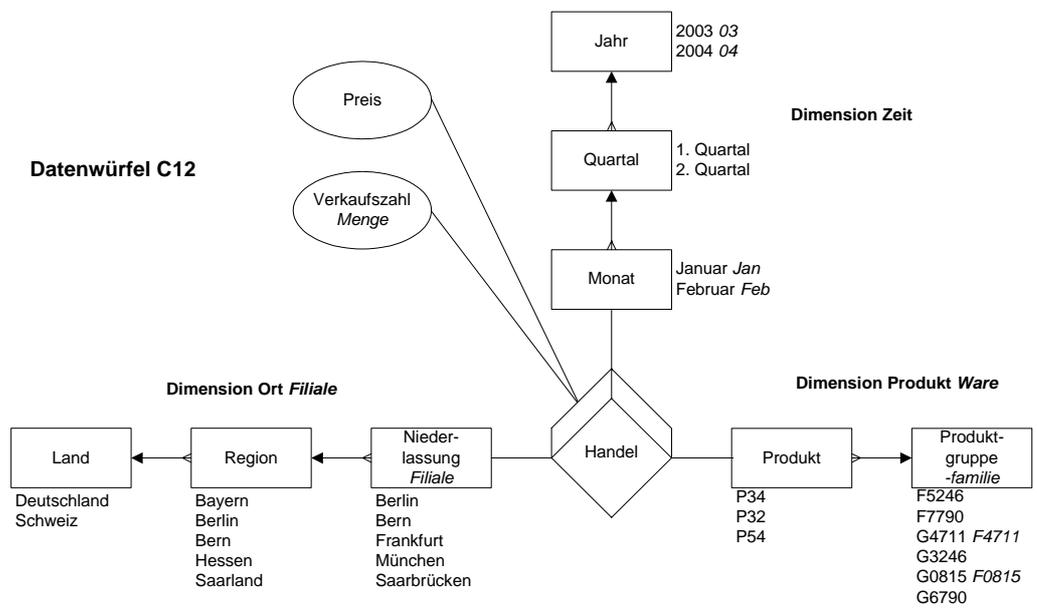


Abbildung 5: Integrierter Datenwürfel

3 Prototypische Implementierung

3.1 Architektur

Zur Evaluation der Integrationsmethode wurde das Konzept in einem Software-Prototyp umgesetzt. Dabei wurde von einer organisatorischen Umgebung ausgegangen, in der mehrere Unternehmen unabhängige und autonome Data Warehouses bzw. Data Marts betreiben. Systemseitig impliziert dies eine vollständige Kapselung der Datenquellen. Diese Kapselung sowie die Software-Funktionalitäten, die zum Aufbau, zum Betrieb und zur Auflösung kollaborativer Data Warehouses benötigt werden, sind die Hauptaufgaben des Prototyps. Einen Überblick über die Systemumgebung ist in der Abbildung 6 dargestellt.

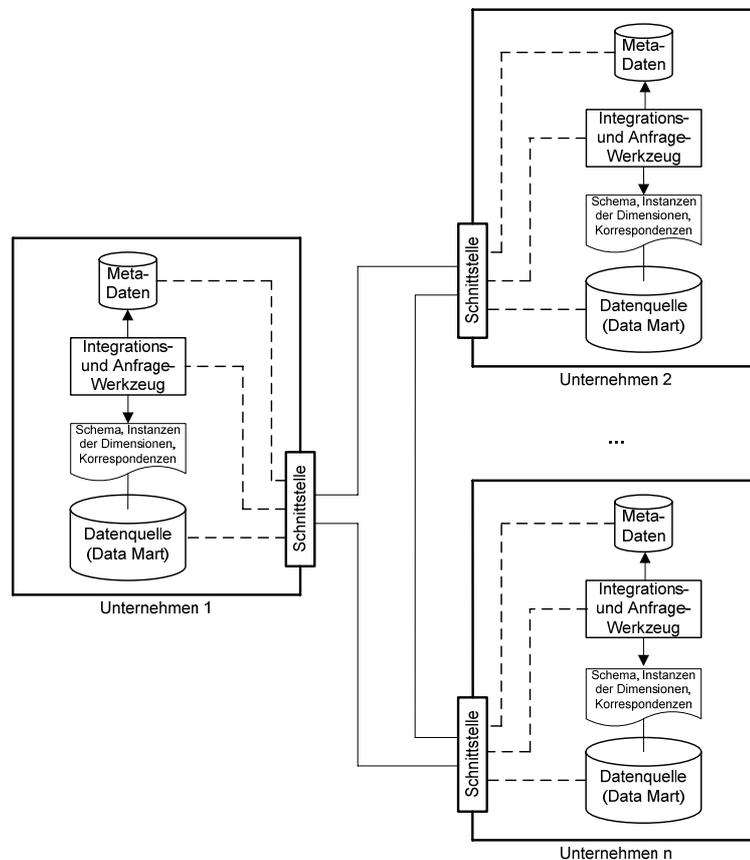


Abbildung 6: Systemumgebung

Jedes an der Kollaboration beteiligte Unternehmen besitzt als zentrale Komponente ein Werkzeug zur Unterstützung des Integrationsprozesses und der Anfragebearbeitung. Tritt ein Unternehmen der Kollaboration bei, so müssen zunächst der zu integrierende Datenwürfel und die zugehörigen Korrespondenzen in Form von XML-Dokumenten erstellt werden. Das Werkzeug liest aus den XML-Dokumenten den lokalen Datenwürfel, die Korrespondenzen sowie aus Metadaten den bereits bestehenden kollaborativen Datenwürfel ein. Unter

Anwendung von Integrationsregeln führt das Werkzeug anschließend die Integration durch. Der aus der Integration resultierende Datenwürfel wird in den Metadaten gespeichert. Entsprechend der Definition eines Datenwürfels werden nicht nur das Schema des kollaborativen Datenwürfels, sondern auch alle Instanzen der dimensional Attribute mit ihren funktionalen Abhängigkeiten in den Metadaten abgelegt. Jedes Unternehmen muss ferner eine Schnittstelle besitzen, über die es seine Daten der Kollaboration verfügbar macht. Auf der Grundlage des in den Metadaten gespeicherten kollaborativen Datenwürfels können im Werkzeug Anfragen an die Kollaboration gestellt werden. Das Werkzeug transformiert die globale Anfrage in Teilanfragen, liest über die entsprechenden Schnittstellen die lokalen Daten ein und transformiert diese zu einem globalen Ergebnis. Im Rahmen der prototypischen Realisierung wird ein Werkzeug zur Unterstützung des Integrationsprozesses und der Anfragebearbeitung realisiert.

Die Abbildung 7 gibt einen Überblick über die System-Architektur des entwickelten Werkzeuges zur Unterstützung des Integrationsprozesses und der Anfragebearbeitung. Die wesentlichen Komponenten der Architektur sind die Integrationskomponente und die Anfragekomponente. Die Integrationskomponente setzt sich aus dem Integrator und der Importkomponente zusammen, die Anfragekomponente aus dem Parser, der Anfragezerlegungskomponente und der Ausführungs- und Auswertungskomponente. Der integrierte Datenwürfel sowie alle benötigten Abbildungsinformationen werden in einer Datenbank abgelegt, so dass sie für die Anfragebearbeitung genutzt werden können. Über das Integrations-Interface kann der Administrator den Integrationsprozess steuern. Der Anwender kann über das Query-Interface SQL-Anfragen in Form von Star-Queries an die Kollaboration stellen. Bei der Entwicklung des Werkzeuges stehen die Anwendung der Integrationsmethode und die Behandlung von Konflikten auf Schema- und Instanzebene im Rahmen der Anfragebearbeitung im Mittelpunkt der Betrachtung. Auf die Entwicklung von Adaptern für unterschiedliche Datenmodelle und Datenbanksysteme wird daher verzichtet.

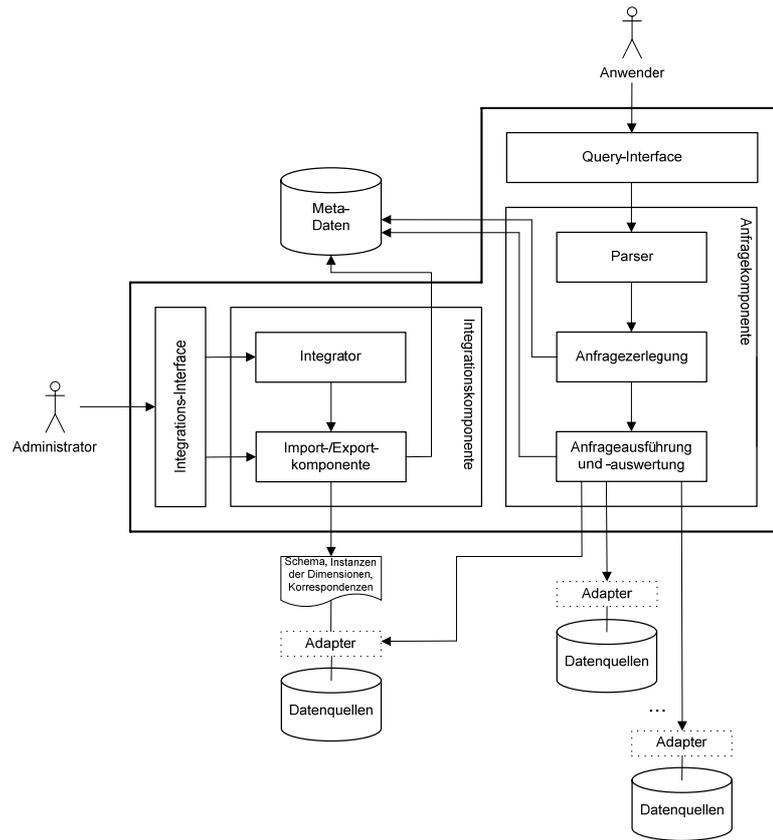


Abbildung 7: System-Architektur

3.2 Umsetzung

Die aktuelle Implementierung beinhaltet dieses Werkzeug, das den Anwender bzw. den Administrator bei der Durchführung des Integrations- und De-integrationsprozesses unterstützt sowie Abfragen auf dem Datenbestand des kollaborativen Data Warehouses ausführen kann. Die Korrespondenzen sowie die benötigten Informationen über das lokale Data Warehouse müssen vom Administrator über eine XML-Spezifikation in das Tool geladen werden. Das Tool führt anschließend durch Anwendung der Integrationsregeln die Integration durch. Dabei werden dem Administrator insbesondere auch die Stellen angezeigt, an welchen die funktionale Abhängigkeit innerhalb des Datenwürfels verletzt wird. Zur manuellen Behebung wird der Administrator durch Assistenten unterstützt, die ihn interviewartig durch den Prozess leiten. Die De-Integration kann im Tool mit Hilfe von De-Integrationsregeln durchgeführt werden. Des Weiteren wurde im Tool eine Anfragebearbeitung realisiert. Das Tool bietet dem Anwender dabei die Möglichkeit, Anfragen an die Data-Warehouse-Kollaboration zu stellen. Auf der Basis der Abbildung zwischen globalem Datenwürfel und lokalen Datenwürfeln, die im Integrationsprozess definiert wurde, wird die globale Anfrage des Anwenders in lokale

Teilanfragen zerlegt [Satt+00]. Die lokalen Anfragen werden auf den lokalen Data-Warehouse-Systemen ausgeführt. Die lokalen Ergebnisse werden anschließend anhand der Abbildungsinformationen zu dem globalen Ergebnis zusammengesetzt. Die Abbildung 8 zeigt einen Ausschnitt des Tools.

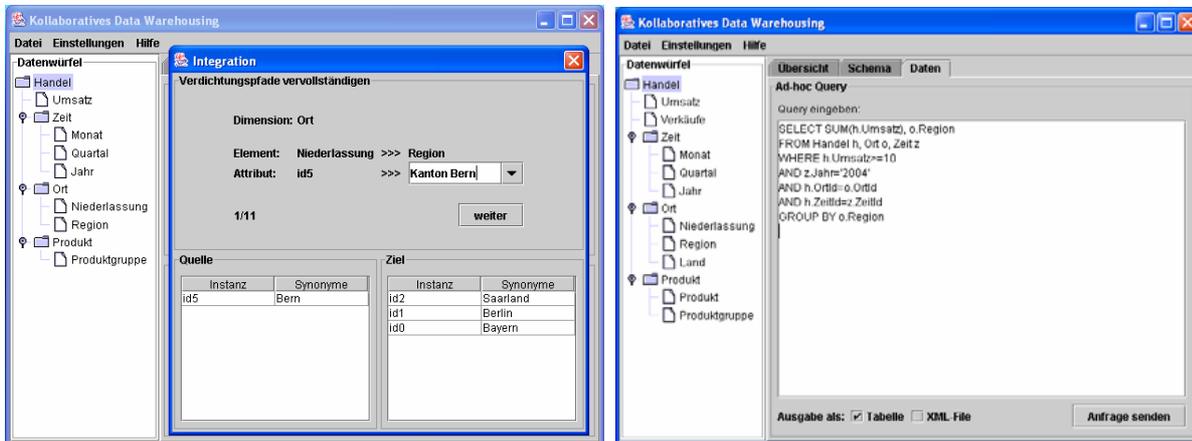


Abbildung 8: Screenshot

4 Zusammenfassung und Ausblick

In dem vorliegenden Beitrag wird auf der Grundlage von föderierten Datenbanktechnologien ein Ansatz für den Aufbau, den Betrieb und die Auflösung einer Data-Warehouse-Kollaboration vorgestellt. Im Mittelpunkt steht dabei der Integrationsansatz. Der Ansatz bietet Unternehmen die Möglichkeit sich flexibel in die Data-Warehouse-Kollaboration zu integrieren. Bestehende unternehmenseigene Data-Warehouse-Lösungen bleiben unverändert. Unternehmensnetzwerke können aus der Kollaboration strategische Informationen gewinnen und dadurch ihre Wettbewerbsposition verbessern.

Im Rahmen zukünftiger Arbeiten sollten noch weitere Aspekte Beachtung finden. Der vorgestellte Integrationsansatz erlaubt nur die Integration von Datenwürfeln, die durch Fakten und Dimensionen mit einfacher Hierarchie modelliert sind. Die Aufnahme weiterer Modellierungsmöglichkeiten bezüglich der Datenwürfel, wie zum Beispiel abgeleitete Kennzahlen oder Dimensionen mit parallelen oder alternativen Hierarchiepfaden, sollten daher noch weiter untersucht werden. Weiterhin sollte die vorgestellte Methodik in ein umfassendes Vorgehensmodell eingebettet werden, das weitere organisatorische und informationstechnische Aspekte des kollaborativen Data Warehousing aufgreift.

Literaturverzeichnis

- [AlÖs01] Alt, R.; Österle, H.: Real-Time Business: Lösungen, Bausteine und Potenziale des Business Networking. Springer, Berlin et al., 2004.
- [Bati+86] Batini, C.; Lenzerini, M., Navathe, S.: A Comparative Analysis of Methodologies for Database Schema Integration. In: ACM Computing Surveys, 18(4), 1986, S. 323-364.
- [Böhn01] Böhnlein, M: Konstruktion semantischer Data-Warehouse-Schemata. Forschungsbeiträge zur Wirtschaftsinformatik, 1.Auflage, Deutscher Universitäts-Verlag, Wiesbaden, 2001.
- [Conr97] Conrad, S.: Föderierte Datenbanksysteme. Konzepte der Datenintegration. Springer, Berlin et al., 1997.
- [Dete02] Determann, L.: Modellierung analytischer Informationssysteme – Ein Konzept zur multidimensionalen Datenstrukturierung. Deutscher Universitätsverlag, Wiesbaden, 2002.
- [Lehn03] Lehner, W.: Datenbanktechnologien für Data-Warehouse-Systeme. Konzepte und Methoden. dpunkt, Heidelberg, 2003.
- [LeSh97] Lenz, H.-J.; Shoshani, A.: Summarizability in OLAP and Statistical Data Bases. In: Proceedings of the 9th International Konferenz on Statistical and Scientific Database Management, 1997, S. 132-143.
- [Lint01] Linthicum, D.: B2B Application Integration: e-Business-Enable Your Enterprise. Addison-Wesley, Boston et al., 2001, S. 48-51.
- [MaWe05] Matheis, T.; Werth, D.: Konzeption und Potenzial eines kollaborativen Data Warehouse-Systems. In: Loos, P. (Hrsg.): Veröffentlichungen des Instituts für Wirtschaftsinformatik, Nr. 185, Universität des Saarlandes, Saarbrücken, 2005.
- [Öste+00] Österle, H.; Fleisch, E.; Alt, R.: Business Networking - Shaping Collaboration Between Enterprises. Springer, Berlin et al., 2000.

- [RaBe01] Rahm, E.; Bernstein, P.: A Survey of Approaches to Automatic Schema Matching. In: The VLDB Journal, 10(4), 2001, S. 334-350.
- [RöSc01] Röhrich, J.; Schlögel C.: cBusiness. Erfolgreiche Internetstrategien durch Collaborative Business am Beispiel der mySAP.com. Addison-Wesley, München et al., 2001.
- [Sche+03] Scheer A.-W.; Adam, O.; Hofer, A.; Zangl, F.: Nach Cost Cutting – Aufbruch durch Innovation. In: IM - Fachzeitschrift für Information Management & Consulting, (18), 2003, gleichzeitig Proceedings zur 24. Saarbrücker Arbeitstagung 2003.
- [Schm98] Schmitt, I.: Schemaintegration für den Entwurf föderierter Datenbanken. Dissertationen zu Datenbanken und Informationssystemen DISDBIS, Band 43, infix, Sankt Augustin, 1998.
- [Satt+00] Sattler, K.-U.; Conrad, S.; Saake, G.: Adding Conflict Resolution Features to a Query Language for Database Federations. In: Proceedings of the 3rd International Workshop on Engineering Federated Information Systems, 2000, S. 41-52.
- [Spac+92] Spaccapietra, S.; Parent, C.; Dupont, Y.: Model Independent Assertions for Integration of Heterogeneous Schemas. In: VLDB Journal, 1(1), 1992, S. 81-126.

Initialisierung des Dispositionssystems neu eröffneter Einzelhandelsmärkte mit Methoden des Data-Mining

Jochen Frank, Michael Kohl, Otto K. Ferstl

Lehrstuhl für Wirtschaftsinformatik, insbes. Industrielle Anwendungssysteme
Otto-Friedrich-Universität Bamberg
96045 Bamberg
jochen.frank@wiai.uni-bamberg.de

Thorsten Ruffer

GfD Gesellschaft für Datenverarbeitung mbH
42929 Wermelskirchen
thorsten.rueffer@gfd.de

Abstract

Bei der Neueröffnung von Einzelhandelsmärkten liegen naturgemäß noch keine historischen Abverkaufsdaten vor. Diese Datenbasis ist jedoch essentiell für eine verbrauchsgesteuerte Disposition, so dass bei der Bedarfsplanung in der Anfangsphase von Einzelhandelsmärkten häufig erhebliche Schätzfehler auftreten. Der vorliegende Beitrag stellt für dieses Problem der Initialisierung von Dispositionssystemen eine Lösung vor, die in Kooperation mit der GfD Gesellschaft für Datenverarbeitung mbH, einem Unternehmen der OBI[®]-Gruppe, entwickelt wurde. Auf Basis von Abverkaufsdaten ausgewählter Referenzmärkte wird mit Methoden des Data-Mining eine fiktive Abverkaufshistorie eines neuen Marktes konstruiert. Damit kann das Dispositionssystem schon bei Markteröffnung auf einer adäquaten Datenbasis arbeiten.

1 Problemstellung

Handelsunternehmen erfüllen beim Warenfluss zwischen Hersteller und Kunde eine zeitliche und räumliche Überbrückungsfunktion. Da im Einzelhandel die Kauftransaktionen seitens der Kunden in der Regel ohne zeitlich vorgelagerten Bestellprozess erfolgen, ist die verbrauchsgesteuerte Disposition eine zentrale Aufgabe des Handels. Unter Berücksichtigung der kon-

fliktären Ziele „Minimierung von Lager- bzw. Beschaffungskosten“ und „permanente Artikelverfügbarkeit“ umfasst die Dispositionsaufgabe alle Tätigkeiten, die mit der Wiederbeschaffung von Artikeln zur Befriedigung der zukünftig erwarteten Kundennachfrage verbunden sind [Götz99, S.22]. Angesichts der im Handel heute üblichen großen Artikelvielfalt wird diese Aufgabe weitestgehend automatisiert durch Dispositionssysteme durchgeführt.

Der Prozess der verbrauchsgesteuerten Disposition gliedert sich in drei Schritte [Hert99, S.268]: (1) Zunächst werden mit der *Bedarfsprognose* auf Grundlage historischer Abverkaufsdaten, die als Indikator für die tatsächliche historische Nachfrage dienen, und aktueller Bestandsdaten die kurzfristigen artikelindividuellen Bedarfe prognostiziert. Die Qualität der Bedarfsprognose besitzt daher wesentlichen Einfluss auf die Zielerreichung der Disposition. Hierbei ist aber zu beachten, dass ein Dispositionssystem auf der Basis von Vergangenheitsdaten arbeitet und folglich konkrete zukünftige Einflussgrößen nicht antizipieren kann. Durch Werbeaktionen veranlasste Nachfrageschwankungen sind hierfür ein Beispiel. Ihr Einfluss wird i. d. R. durch den Disponenten berücksichtigt. (2) Bei der *Generierung von Bestellungen* werden zur Optimierung der Beschaffungskosten die Ergebnisse der Bedarfsprognose mit der Bestellpolitik abgeglichen. (3) Im letzten Schritt erfolgt die *Bestellabwicklung*.

Bei der Neueröffnung von Märkten stoßen Dispositionssysteme an ihre Grenzen, da ohne eine Abverkaufshistorie eines neuen Marktes die Datenbasis zur Erstellung von Bedarfsprognosen fehlt. Insbesondere in den ersten Monaten nach Markteröffnung kann dies zu beträchtlichen systematischen Fehlern bei der Bedarfsprognose führen. Diese Prognosefehler bewirken sowohl zu hohe Bestände und damit Mehrkosten als auch zu niedrige Bestände, die in Out-of-Stock-Situationen von Artikeln resultieren und Fehlmengenkosten bzw. geringere Kundenzufriedenheit mit sich bringen [CoGr04, S.26ff.]. Angesichts des Trends von kleinen Märkten hin zu Großmärkten und der damit verbundenen raschen Expansion großer Einzelhandelsketten ist die beschriebene Problematik aktuell von hoher praktischer Relevanz.

Der vorliegende Beitrag stellt für das beschriebene Problem eine Lösung vor, die in Kooperation mit der *GfD Gesellschaft für Datenverarbeitung mbH*, einem Unternehmen der OBI®-Gruppe, entwickelt wurde. Kapitel 2 beschreibt zunächst die Rahmenbedingungen und Gestaltungsziele, die bei der Entwicklung einer Lösung zu berücksichtigen sind. In den beiden folgenden Kapiteln wird untersucht, wie die genannten Gestaltungsziele sich realisieren lassen. Kapitel 5 stellt die entwickelte Methodik im Einzelnen vor. Die erzielten Ergebnisse sowie sich daraus ergebende Forschungsanstöße sind schließlich Thema des letzten Kapitels 6.

2 Rahmenbedingungen

2.1 Abverkaufshistorien von Einzelhandelsmärkten

Ziel des zu entwickelnden Lösungsansatzes ist die Bereitstellung einer fiktiven Abverkaufshistorie (AH) für die Initialisierung des Dispositionssystems bei neuen Märkten. Hierbei bezeichnet eine AH die artikel- und tagesgenauen Abverkaufsmengendaten aller in einem Markt geführten Artikel bis hin zur Gegenwart. Da zur Erkennung saisonaler Muster generell Abverkaufszeitreihen über mindestens zwei Jahre nötig sind, muss eine AH mindestens diesen Zeitraum umfassen.

Eine Zeitreihe kann gemäß dem Multiplikativen Modell in *Zeitreihenstruktur (ZRS)* und *Niveauparameter (NP)* zerlegt werden [ScSt01, S.9f]. Der NP ist definiert als Integral einer Zeitreihe über ein endliches Zeitintervall. Bei einem Betrachtungszeitraum von 12 Monaten entspricht der NP der Jahresabverkaufsmenge (JAM) eines Artikels. Die ZRS ergibt sich durch Normierung der Abverkaufszeitreihe auf den Wert 1 mittels Division aller Werte der Abverkaufszeitreihe durch den zugehörigen NP. Bei Wahl gleicher Betrachtungszeiträume – hier normalerweise ein Jahr – lässt sich anhand der ZRS das Abverkaufsverhalten von Artikeln beschreiben und wegen der Normierung miteinander vergleichen. Angesichts im Handel üblicher Bestellzyklen reicht ein Vergleich von ZRS auf Basis des Rasters von Kalenderwochen aus. Abb. 1 zeigt einige markante ZRS, die bei Abverkaufsverläufen verschiedener Artikel identifiziert wurden.

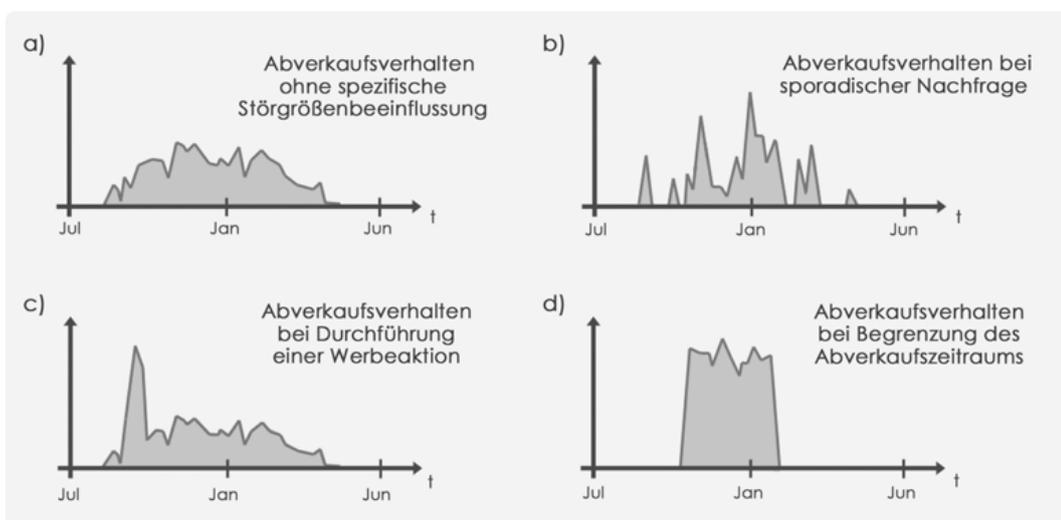


Abb. 1: Markante Zeitreihenstrukturen von Abverkaufszeitreihen wintersaisonaler Artikel eines Einzelhandelsmarktes

Grundsätzlich beeinflusst eine Vielzahl von Faktoren die täglichen Abverkäufe eines Artikels auf Marktebene. Diese Tatsache spiegelt sich in der ZRS wider, ohne dass die Kausalität jeder einzelnen Einflussgröße exakt quantifiziert werden könnte. Unter Berücksichtigung einer allgemeinen Zufallskomponente kann man dennoch häufig bei einem Artikel ein charakteristisches Abverkaufsverhalten wie z. B. ganzjähriger Abverkauf, Sommer- oder Wintersaisonalität erkennen (Abb. 1a). Aufgrund i. d. R. diskreter Verkaufseinheiten ist der Einfluss der Zufallskomponente für die Erkennung von Mustern der ZRS jedoch umso problematischer, je geringer die JAM eines Artikels ist (Abb. 1b). Als spezielle Einflussgrößen sind die Durchführung von Werbeaktionen sowie die angebotsseitige Begrenzung von Abverkaufszeiträumen von Artikeln hervorzuheben. Mit der Durchführung einer Werbeaktion wird eine bewusste Manipulation des Nachfrageverhaltens verfolgt, wodurch sich im Regelfall kurzfristig eine starke Erhöhung der Abverkaufszahlen ergibt, die aber zum Teil nur eine zeitliche Vorverlagerung der Nachfrage bedeutet (Abb. 1c) [Pepe01, S.88]. Eine Begrenzung des Abverkaufszeitraums ist vor allem bei saisonalen Artikeln beobachtbar. Diese werden z. B. aus Gründen der Produktivitätserhöhung der Verkaufsflächen unter Umständen nicht über den kompletten nachfragebedingten Saisonzeitraum angeboten. Die ZRS beschreibt in diesem Fall nur einen Ausschnitt des für den Artikel charakteristischen Abverkaufsverhaltens (Abb. 1d).

2.2 Gestaltungsziele für die Konstruktion der Abverkaufshistorie des neuen Marktes

Aus der eingangs beschriebenen Problemstellung lassen sich folgende drei Gestaltungsziele für die zu konstruierende fiktive AH ableiten:

(1) Zunächst muss die AH grundsätzlich den *Anforderungen des eingesetzten Dispositionssystems* genügen. Von Bedeutung sind hierbei Format und Zeitraum der AH. Hinsichtlich des *Formats* ist für jeden einzelnen Artikel eine tagesgenaue und lückenlose Historie der Abverkaufsdaten zu erstellen, die mit der Schnittstelle des Dispositionssystems kompatibel ist. Der *Beginn des Zeitraums* sollte mindestens zwei Jahre (vgl. 2.1) in der Vergangenheit liegen, wobei jedoch für Artikel mit kürzerem Listungszeitraum im Unternehmenssortiment die Abverkaufsdaten erst ab dem jeweiligen ersten Listungstag zu konstruieren sind. Das *Ende des Zeitraums* sollte für einen nahtlosen Übergang zu den zukünftig realen Abverkaufsdaten genau mit dem fiktiv letztmöglichen Verkaufstag vor Eröffnung des neuen Marktes zusammenfallen.

(2) Als zweites Gestaltungsziel ist die *Spezifität der Nachfragesituation des neuen Marktes* zu berücksichtigen. Die täglichen Abverkäufe von Artikeln unterliegen für jeden Markt charakte-

ristischen Einflussgrößen, die hinsichtlich ihrer Wirkung auf ZRS und NP differenziert zu betrachten sind. So spiegelt sich etwa der regional unterschiedliche Beginn der Wachstumsperiode deutlich in den ZRS für Gartenartikel wider. Dagegen hat beispielsweise die Größe eines Marktes (Verkaufsfläche, Sortimentsumfang) hauptsächlich Einfluss auf den NP.

(3) Schließlich sollten in der fiktiven AH *die Abverkäufe unbeeinflusst von Störgrößen* dargestellt werden. Wie in Kapitel 2.1 beschrieben, unterliegen die Abverkäufe eines Artikels marktindividuell einer Vielzahl von Faktoren. Als Störgrößen werden dabei jene Einflussfaktoren betrachtet, die die Indikatoreigenschaft der Abverkäufe für die tatsächliche Nachfrage verfälschen und somit dem Dispositionssystem des neuen Marktes, das für eine breite Anwendbarkeit des Ansatzes bewusst als Black Box aufgefasst wird, aus den realen Daten nicht ableitbare Muster vortäuschen könnten. Dazu zählen neben der allgemeinen Zufallskomponente insbesondere Werbeaktionen sowie angebotsseitige Begrenzungen des Abverkaufszeitraums. Die Bereinigung dieser besonderen Einflussgrößen, die in Kapitel 4 beschrieben wird, ist in Abb. 2 beispielhaft dargestellt. Eine direkte Konsequenz des letzten Gestaltungsziels ist die Begrenzung der Anwendbarkeit des Ansatzes auf Sortimentsartikel. Allerdings werden für reine Werbeartikel i. d. R. ohnehin eigene Dispositionsverfahren verwendet.

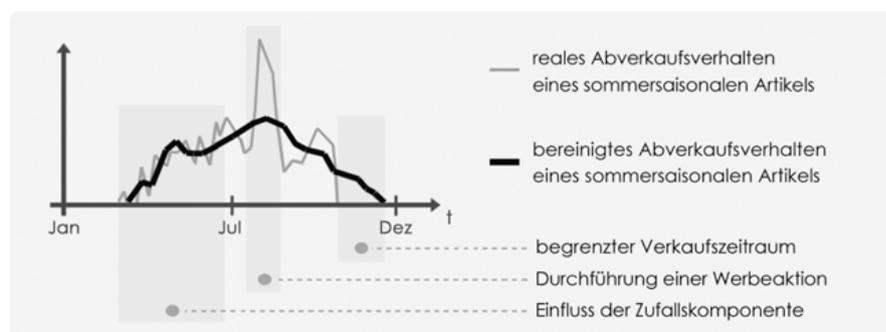


Abb. 2: Beispielhafte Darstellung eines um Störeffekte bereinigten Abverkaufsverhaltens eines sommersaisonalen Artikels

2.3 Anforderungen an die systemtechnische Umsetzung der Konstruktionsmethodik

In funktionaler Hinsicht wird aufgrund der Artikelvielfalt zunächst eine weitestgehende Automatisierung der Konstruktionsmethodik gefordert. Das System zur Konstruktion der fiktiven AH erhält dabei als Input die Parameter des neuen Marktes inklusive der Stammdaten aller im neuen Markt gelisteten Artikel sowie für die Konstruktionsmethodik notwendige Daten anderer Märkte. Der Output besteht aus einer fiktiven Abverkaufshistorie gemäß den Ge-

staltungszielen (vgl. 2.2). Die Ergebnisse eines Systemlaufs sollten bereits einige Zeit vor der Eröffnung des neuen Marktes verfügbar sein, damit noch genügend Zeit für organisatorische Vorarbeiten bleibt. In qualitativer Hinsicht sind insbesondere Performance, Zuverlässigkeit sowie – im Rahmen des Einsatzes zur Managementunterstützung – die Nachvollziehbarkeit von Ergebnissen von Bedeutung [KiK177, S.36ff.].

3 Klassifikation des Problems und des Lösungskonzeptes

Bei der Neueröffnung eines Einzelhandelsmarktes liegen naturgemäß noch keine historischen Abverkaufsdaten vor, sondern lediglich Strukturinformationen zu Marktaufbau (z. B. Verkaufsfläche, Sortiment) und Standort (z. B. Einzugsgebiet, Kaufkraft der Region) sowie Abverkaufsdaten weiterer, bereits existierender Märkte. Die Erstellung einer AH des neuen Marktes wird als Konstruktion des Verhaltens eines Systems auf der Grundlage festgelegter Strukturmerkmale des Systems sowie anhand der Verhaltensspektren ähnlicher Systeme interpretiert. Es wird postuliert, dass die Struktur und das Verhalten des neuen Systems den Strukturen und den Verhaltensspektren einer Referenzgruppe existierender Systeme ähnlich sein sollen, wobei die Mitglieder dieser Referenzgruppe ebenfalls untereinander Ähnlichkeit bezüglich Struktur und Verhalten aufweisen. Die Struktur des neuen Systems wird vom Marktgestalter ähnlich zur Struktur der Referenzgruppe festgelegt. Bei der Generierung der Abverkaufsdaten ist dann ein Verhalten zu konstruieren, das den Verhaltensspektren der Mitglieder der Referenzgruppe ebenfalls ähnlich ist. Da bei einer Einzelhandelskette Daten einer Vielzahl bereits bestehender Märkte verfügbar sind, ist zu untersuchen, welche Märkte in diese Referenzgruppe passen.

Für den Strukturvergleich muss eine möglichst feingranulare Aggregationsebene gewählt werden, für die aber hinreichend Strukturdaten verfügbar sind. Diese Ebene wurde beim betrachteten Unternehmen in Form der so genannten Sortimentshauptgruppen (SHG) gefunden. SHG sind unternehmensweit einheitliche Einteilungen des Sortiments, die in den Märkten in entsprechenden räumlich getrennten Verkaufsbereichen angeboten werden (z. B. Gartenartikel im Baumarkt). Märkte sind daher modular aus SHG aufgebaut. Strukturvergleiche werden daher nur auf Ebene der SHG – und nicht der Märkte insgesamt – angestellt.

Bei der Auswahl geeigneter Parameter für den Strukturvergleich ist deren möglicher Bezug zum Verhalten der SHG, also zur Gesamtheit aller Abverkaufszeitreihen, zu prüfen. Da das

Verhalten der SHG von Angebot und Nachfrage beeinflusst ist, wurden im Rahmen einer Kausalitätsanalyse systematisch Angebots- (z. B. Sortimentsumfang) und Nachfrageparameter (z. B. Kaufkraft des Einzugsgebiets) auf Plausibilität untersucht.

Im Rahmen einer umfassenden Analyse wurden Korrelationen zwischen Struktur und Verhalten der SHG erkannt und als Hypothese für die weitere Untersuchung zugrunde gelegt. Bei Zerlegung der Zeitreihen in ZRS und NP zeigte die Analyse beispielsweise, dass Lageparameter (z. B. Region, Wirtschaftsraum) vor allem die ZRS beeinflussen, während Größenparameter (z. B. Verkaufsfläche, Sortimentsumfang) sich primär auf den NP auswirken. Beim Konstruktionsverfahren für das Systemverhalten, d. h. für die zu generierenden Abverkäufe, wurde daher angenommen, dass die Korrelationsbeziehungen der existierenden SHG auch in den SHG des neuen Marktes zumindest näherungsweise bestehen. Um die Referenzgruppe zu bestimmen, sind bereits bestehende SHG zu finden, die den SHG des neuen Marktes in der Struktur ähnlich sind. Aus dem Verhalten dieser SHG in der Vergangenheit wird dann die fiktive Abverkaufshistorie des neuen Marktes konstruiert.

Da im Vorfeld der Ähnlichkeitsanalyse keine Hypothesen über die Ähnlichkeit bestimmenden Merkmale von Märkten und ihren SHG bestehen, liegt ein hypothesenfreies und somit datengetriebenes Datenanalyseproblem vor. Als Analysemethoden für diese Art von Problemen bieten sich insbesondere Data-Mining-Verfahren an, deren Zielausrichtung, die Aufdeckung unbekannter, komplexer Muster in großen Datenbeständen, zu den beschriebenen Eigenschaften der vorliegenden Untersuchungssituation passt [Fers79, S.43; NeKn05, S.81f]. Die im Rahmen der Lösung eingesetzten Data-Mining-Verfahren werden in den nächsten beiden Kapiteln beschrieben.

4 Behandlung von Störgrößen

Für das Gestaltungsziel einer von Störgrößen möglichst unbeeinflussten AH (vgl. 2.2) sind noch Vorüberlegungen zur Extraktion und Transformation der verfügbaren Abverkaufszeitreihen anzustellen.

4.1 Eröffnungseffekt

Nach der Eröffnung eines Marktes zeigen die Abverkäufe ein bestimmtes Verhalten, das hier als Eröffnungseffekt bezeichnet werden soll. Bei diesem Phänomen (Abb. 3) können idealty-

pisch drei Phasen abgegrenzt werden. Anfangs besteht in Phase I aufgrund von Werbemaßnahmen, Verkaufsaktionen und Unterhaltungsangeboten ein besonderes Kundeninteresse, das in erhöhten Verkäufen von Sortiments- und vor allem Werbeartikeln resultiert. Phase II zeichnet sich durch einen Rückgang der Abverkäufe aus, der sich zum Teil durch eine Nachfrageverlagerung in Phase I erklären lässt. Bei einem Teil des Sortiments machen sich aber auch Out-of-Stock-Situationen bemerkbar, deren Ursachen in Kapitel 1 beschrieben wurden. Phase 3 ist vom Ausklingen des Eröffnungseffektes, d. h. einer allmählichen Normalisierung der Abverkäufe gekennzeichnet. Untersuchungen im Unternehmen haben gezeigt, dass für die ersten beiden Phasen ein Zeitraum von maximal sechs Monaten anzusetzen ist.

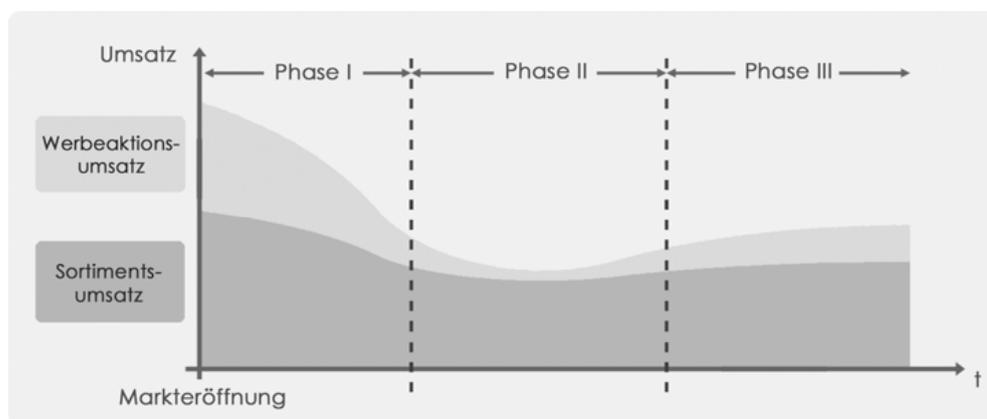


Abb. 3: Phasen des Eröffnungseffekts

Der Eröffnungseffekt verfälscht den Indikator Abverkäufe, so dass die Abverkaufsdaten kürzlich eröffneter Märkte unbrauchbar sind. Für den Lösungsansatz ist er im Weiteren nicht relevant, da er in Phase I primär auf die hier nicht berücksichtigten reinen Werbeartikel wirkt (vgl. 2.2) und die Probleme in Phase II gerade durch den Lösungsansatz verhindert werden sollen. Im Übrigen können Dispositionssysteme mit einer geeigneten AH – wie sie hier konstruiert werden soll – den Eröffnungseffekt als spezielle Werbemaßnahme berücksichtigen.

4.2 Glättung der Zufallskomponente

Die Zufallskomponente beeinflusst vor allem die ZRS, während beim NP durch die Aggregation von Tagesabverkäufen gemäß dem Gesetz der großen Zahlen automatisch ein Glättungseffekt eintritt [Sach04, S.129]. Die Glättung der Zufallskomponente für Artikel bestehender Märkte kann daher prinzipiell durch Mittelwertbildung über mehrere ZRS erreicht werden. Um aber den marktspezifischen Charakter der bereinigten ZRS zu erhalten, kann die Mittel-

wertbildung nicht über gleiche Artikel verschiedener Märkte, sondern nur über Artikel desselben Marktes erfolgen.

Zur Gruppierung der Artikel, deren Zeitreihen durch Aggregation geglättet werden sollen, liegt angesichts der Sortimentsstrukturhierarchie des Unternehmens zunächst die oberhalb der Articlebene angesiedelte Ebene der Warengruppe (WG) nahe. Diese böte die Möglichkeit, die Glättung über eine größere Anzahl von Artikeln mit gemeinsamen Merkmalen durchzuführen. Die Artikel einer WG weisen jedoch teilweise ein sehr unterschiedliches Abverkaufsverhalten auf. Dies ist erklärbar, da in einer WG beispielsweise Gebrauchsgüter (z. B. Rasenmäher) mit zugehörigen Verbrauchsgütern (z. B. Klingen) eingeordnet sein können. Eine Glättung auf Ebene der WG würde daher auch zu einer nicht gewünschten Eliminierung des charakteristischen Abverkaufsverhaltens der Artikel führen. Für eine zweckmäßigere Gruppierung der Artikel wird daher die Ebene der virtuellen Warengruppen (vWG) zwischen WG und Artikel eingeführt.

Die Artikel innerhalb einer WG werden mittels eines Clusterverfahrens, das die aus normierten Werten bestehende ZRS der Abverkäufe auf Unternehmensebene und damit unabhängig von individuellen Einflussfaktoren einzelner Märkte vergleicht, in vWG eingeteilt. Gegen eine Clusterung aller Artikel auf einer höheren Sortimentshierarchieebene (z. B. SHG) spricht der damit einhergehende Informationsverlust, da gerade die Einordnung von Artikeln in WG aufgrund gemeinsamer Merkmale dieser Artikel erfolgt und die WG-Zugehörigkeit anderenfalls nur noch als nominal skaliertes Merkmal in die Clusterung einginge.

Zur Erkennung saisonaler Muster bei der Bildung von vWG wäre eine möglichst große Zeitspanne wünschenswert. Die Ähnlichkeitsanalyse von ZRS setzt jedoch deren einheitliche Länge voraus, so dass Artikel, die innerhalb des Vergleichszeitraums neu gelistet wurden, ausselektiert werden müssen. Je länger der Zeitraum gewählt wird, desto geringer ist folglich die Datenbasis. Unter der empirisch überprüften Annahme, dass das jahressaisonale Abverkaufsverhalten der Artikel hinreichend stabil ist, werden für den Input des Clusterverfahrens die letzten 52 Kalenderwochen (Rolljahr) vor dem Stichtag des Datenabrufs verwendet. Als der Analysesituation angemessenes Clusterverfahren wurde die Self Organizing Map (SOM) [Koho01; Pete05, S.26] gewählt. Die SOM benötigt keine Vorgabe hinsichtlich der Anzahl der zu bildenden Cluster, welche lediglich durch die vorher festgelegte Kartengröße nach oben beschränkt wird, und toleriert bei entsprechender Algorithmik auch fehlende Werte im Inputvektor [CoLe05]. Falls Cluster mit einer für die Glättung der Zufallskomponente nicht

ausreichenden Zahl an Artikeln entstehen und somit als vWG ungeeignet sind, werden diese nachfolgend mit einem bestehenden Cluster mit minimaler Distanz zwischen den zugehörigen Kartenneuronen zusammengefasst.

Jeweils individuell für jeden Markt erhalten alle Artikel einer vWG die gleiche geglättete ZRS. Den vor der Clusterung ausselektierten neu gelisteten Artikeln, die folglich keiner vWG zuordenbar sind, wird ersatzweise die über alle Artikel ihrer WG geglättete ZRS zugewiesen.

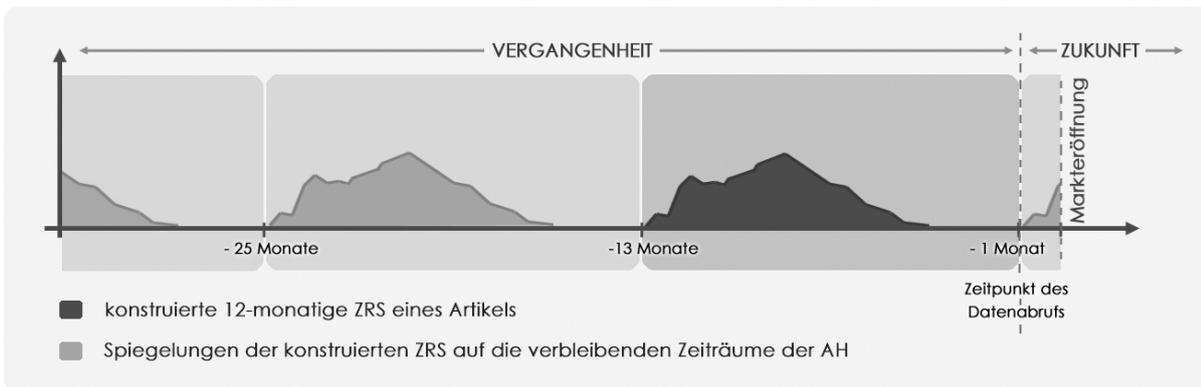


Abb. 4: Konzept der Spiegelung der ZRS

Um die geforderte Abverkaufshistorie von mehr als zwei Jahren zu erhalten, wird die bisher nur für das aktuelle Rolljahr ermittelte ZRS gespiegelt (Abb. 4). Diese Spiegelung ist sowohl in die Vergangenheit als auch in die Zukunft möglich und bietet somit zusätzlich den Vorteil, dass für die unvermeidbare zeitliche Lücke zwischen dem Datenabruf bei den bestehenden Märkten und der Eröffnung des neuen Marktes Abverkaufszeitreihen konstruiert werden können. Als nachteilig ist bei diesem Vorgehen anzusehen, dass durch die Spiegelung mehrjährige Trends vom Dispositionssystem anhand der Datenbasis nicht mehr erkannt werden können. Zudem bleiben bewegliche Feiertage der Vorjahre unberücksichtigt, und einmalige, unternehmensweit auftretende Einflussfaktoren innerhalb des 12-monatigen Abfragezeitraums, wie z. B. anhaltende, extreme Wetterlagen, werden eventuell generalisiert. Eine Lösung dieser Probleme erscheint möglich, jedoch müsste man die Prämisse von der Unabhängigkeit des hier vorgestellten Verfahrens vom eingesetzten Dispositionssystem aufgeben.

4.3 Begrenzte Abverkaufszeiträume

Die auf Marktebene vor allem bei saisonalen Artikeln beobachtete angebotsseitige Verkürzung des Verkaufszeitraums (vgl. 2.1) verfälscht grundsätzlich die Indikatoreigenschaft der Abverkäufe für die Nachfrage. Die Auswirkungen dieses Vorgehens können durch einen Ver-

gleich der ZRS eines betroffenen Artikels auf Markt- und Unternehmensebene untersucht werden. Dabei zeigt sich, dass in den meisten Fällen der damit verbundene Prognosefehler vernachlässigbar ist. Dies ist nachvollziehbar, da die Verkürzung meist ganz bewusst auf den umsatzschwachen Anfangs- bzw. Endzeitraum eines Saisonartikels gerichtet ist. Die relativ kleine Anzahl an Artikeln mit merklichen Abweichungen wird bei diesem Vergleich erkannt und ausselektiert.

4.4 Bereinigung der Effekte von Werbemaßnahmen

Bei der Bereinigung der Effekte von Werbemaßnahmen wird versucht, deren direkte Auswirkungen auf die Abverkäufe des beworbenen Artikels bzgl. ZRS und NP näherungsweise zu glätten. Mögliche positive oder negative Rückkopplungen auf die Verkäufe anderer Artikel werden dabei nicht berücksichtigt.

Zunächst sind jene wenigen Artikel zu identifizieren, bei denen auf Marktebene der Anteil der Abverkäufe während einer Werbeaktion an der Jahresabverkaufsmenge (JAM) wesentlich höher ist als auf Unternehmensebene. Die Daten dieser Artikel sind wegen ihrer durch die Werbemaßnahmen induzierten extremen Abverkaufsspitzen zur Glättung der ZRS innerhalb ihrer vWG ungeeignet und werden ausselektiert. Für die übrigen Artikel erweist sich das in Kapitel 4.2 beschriebene Verfahren zur Bereinigung der Zufallskomponente auch für die Werbemaßnahmen als ausreichend.

Beim NP gestaltet sich die Bereinigung um Werbeeffekte aufwendiger, weil abzuschätzen ist, welcher Anteil an der JAM eines Artikels dem Einfluss von Werbemaßnahmen zuzuschreiben ist und somit nicht zu den gewünschten nachfragebedingten Abverkäufen gehört. Zur Lösung dieses Problems bietet es sich an, auf die ZRS der vWG zurückzugreifen, da auf dieser Ebene die Störeffekte der Werbemaßnahmen bereits eliminiert sind, wie der Vergleich mit der ZRS eines beworbenen Artikels zeigt (Abb. 5a). Diese Erkenntnis wird genutzt, indem man für den Zeitraum der Werbeaktion die ZRS des Artikels durch diejenige der zugehörigen vWG ersetzt und so näherungsweise eine um die Werbeabverkäufe bereinigte ZRS konstruiert (Abb. 5b). Das Verhältnis der Integrale von modifizierter und originärer ZRS des beworbenen Artikels entspricht in diesem Sinne dem Anteil der reinen Sortimentsabverkäufe ohne Werbeabverkäufe an der JAM und kann als Faktor zur Bereinigung des NP um Werbeeffekte verwendet werden.

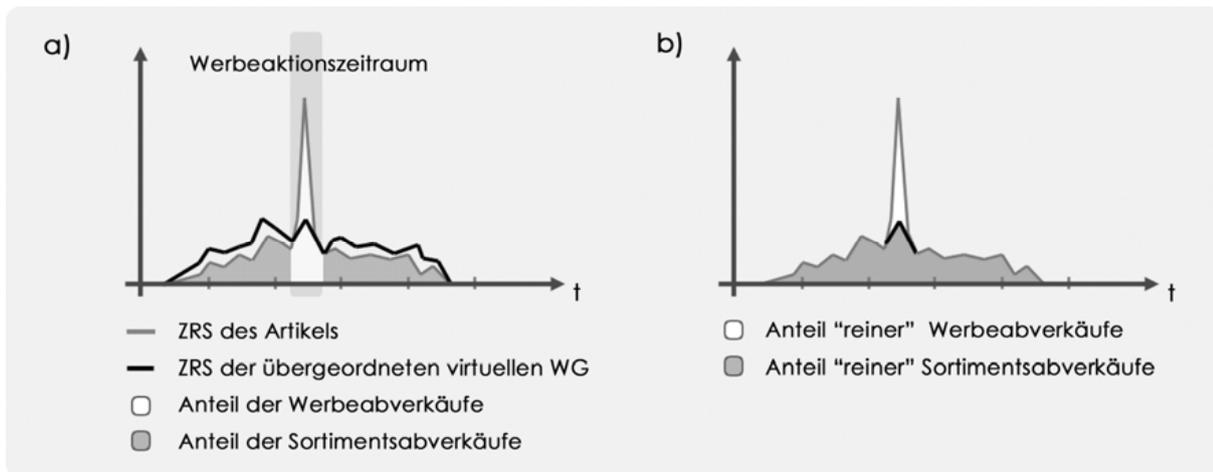


Abb. 5: Bereinigung des NP um Werbeeffekte

5 Konstruktionsverfahren der Abverkaufshistorie

Auf Basis des in Kapitel 3 beschriebenen Lösungskonzeptes und der Überlegungen zur Behandlung von Störgrößen aus Kapitel 4 wird nun das aus drei Schritten bestehende Verfahren zur Konstruktion einer fiktiven AH (Abb. 6) vorgestellt. (1) Zunächst erfolgt die *Generierung der Zeitreihenstruktur* aller Artikel des neuen Marktes auf Grundlage der Abverkaufsdaten möglichst ähnlicher Referenzmärkte. Dies sind Märkte mit einer AH von mindestens 18 Monaten, so dass sie über eine Datenbasis von mindestens einem Rolljahr verfügen, die nicht mehr vom Eröffnungseffekt (vgl. 4.1) beeinflusst ist.

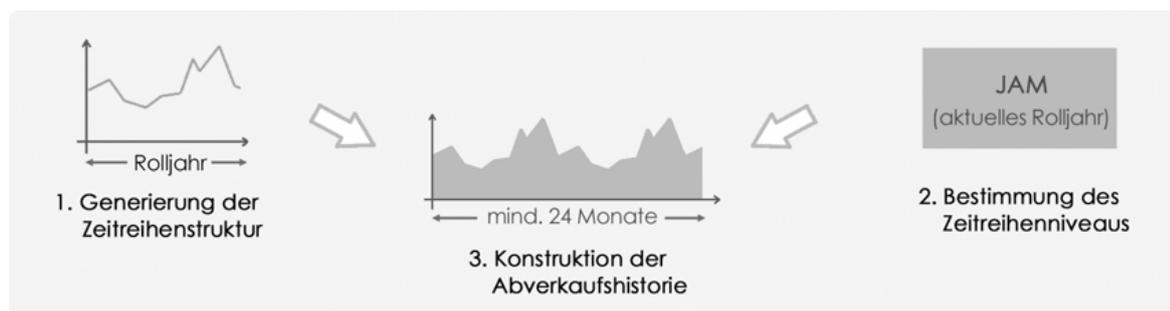


Abb. 6: Teilschritte des Konstruktionsverfahrens

(2) Danach werden mit der *Bestimmung des Niveauparameters* die fiktiven JAM dieser Artikel ermittelt. (3) Schließlich werden bei der *Konstruktion der fiktiven Abverkaufshistorie* die Ergebnisse der ersten beiden Schritte zusammengeführt und mit Hilfe eines Historienkalenders in die gewünschte Abverkaufshistorie transformiert. Die ersten beiden Teilschritte wer-

den dabei für den Zeitraum eines Rolljahres durchgeführt, während durch Spiegelung (vgl. 4.2) die Abverkaufshistorie für einen beliebig langen Zeitraum konstruiert werden kann.

5.1 Generierung der fiktiven Zeitreihenstruktur

Die Generierung der fiktiven ZRS (Abb. 7) beginnt mit der *Bereinigung der ZRS auf Basis der Abverkaufsdaten der Referenzmärkte* (Abb. 7a). Dabei werden von Störgrößen bereinigte ZRS aller vWG erzeugt (vgl. 4.2), und zwar sowohl auf Markt- als auch auf Unternehmens-ebene. Letztere werden immer dann verwendet, wenn auf Marktebene sporadisches Abverkaufsverhalten [Nowa05, S.61ff.] vorliegt.

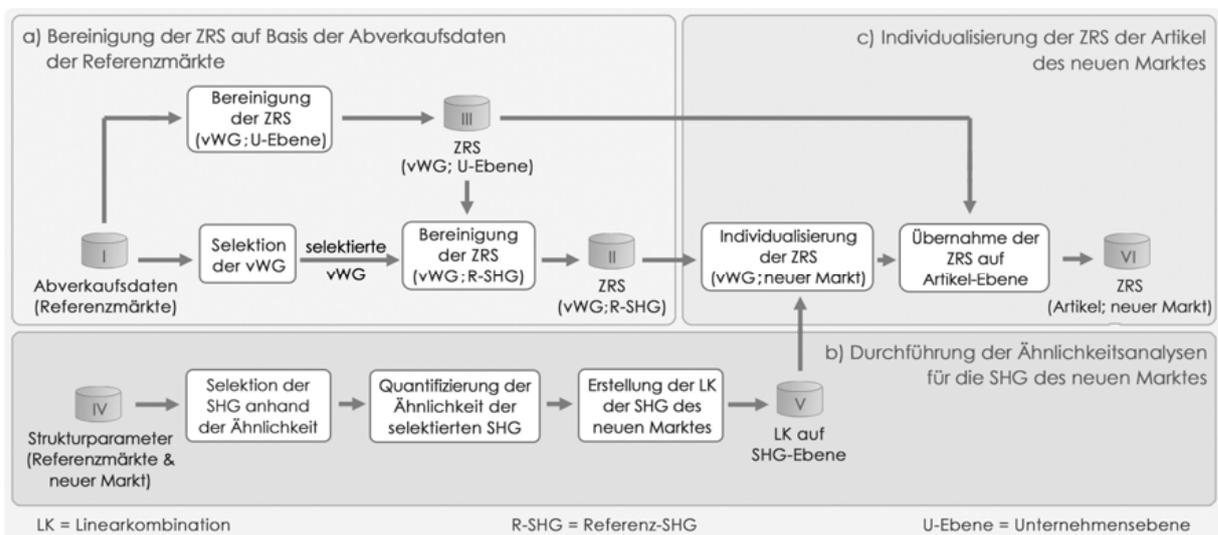


Abb. 7: Generierung der fiktiven Zeitreihenstruktur

Gemäß dem in Kapitel 3 vorgestellten Ansatz basiert die Konstruktion der fiktiven AH des neuen Marktes auf historischen Abverkaufsdaten von Märkten, die in ihrer Struktur möglichst ähnlich sind. Dazu wird eine dreistufige *Ähnlichkeitsanalyse der Märkte auf Ebene ihrer SHG* (Abb. 7b) durchgeführt. (1) Zunächst erfolgt mittels Clusteranalyse die *Selektion der SHG* jener Referenzmärkte, welche zur korrespondierenden SHG des neuen Marktes eine möglichst hohe strukturelle Ähnlichkeit aufweisen. Analog zu den Überlegungen in Kapitel 4.2 wird hier erneut SOM als Clusterverfahren gewählt. Angesichts der zahlreichen verwendeten Inputvariablen verschiedenartigen Charakters – z. B. als Lage- oder Größenparameter (vgl. Kapitel 3) – bietet sich hier jedoch die vorherige Anwendung eines dimensionsreduzierenden statistischen Verfahrens an. Bei einer Clusteranalyse führt ansonsten die Existenz von Korrelationen und Redundanzen innerhalb des Inputvektors dazu, dass die durch mehrere Variablen

repräsentierten Informationen eine unverhältnismäßig hohe Wirkung auf die Clusterbildung besitzen [HiSc01, S.613ff.]. Dieser Effekt lässt sich durch Vorschalten einer Faktorenanalyse vermeiden. Dabei werden die metrischen Inputvariablen standardisiert, über eine Analyse der Korrelationen in Gruppen zusammengefasst und durch Faktoren repräsentiert. Als Extraktionsverfahren wird dabei eine Hauptkomponentenanalyse in Verbindung mit einer Rotation der extrahierten Faktoren eingesetzt [KüKa01, S.186]. (2) Nach der Clusteranalyse wird mit einem adäquaten Proximitätsmaß – hier aufgrund metrischer Merkmale die Euklidische Distanz – die *Ähnlichkeit der selektierten SHG* zu den SHG des neuen Marktes mit Distanzwerten *quantifiziert*. (3) Zuletzt werden *Linearkombinationen (LK) zur Beschreibung der SHG* des neuen Marktes erstellt (Abb. 8). Hierfür transformiert man die Distanzwerte in Gewichte, die zu den Distanzwerten in einem reziproken Verhältnis stehen und in ihrer Summe auf 1 normiert sind.

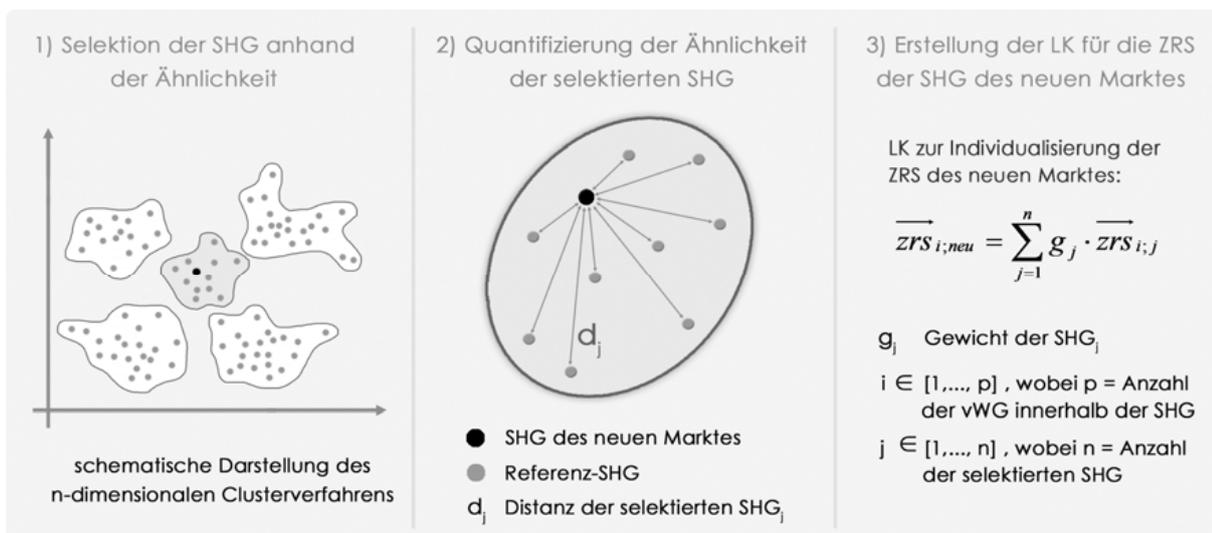


Abb. 8: Teilschritte der Ähnlichkeitsanalyse

Zur *Individualisierung der ZRS von Artikeln des neuen Marktes* (Abb. 7c) werden die für die vWG der Referenz-SHG ermittelten ZRS gemäß der durch die LK bestimmten Gewichte zu den fiktiven ZRS der vWG des neuen Marktes kombiniert. Bei vWG ohne marktindividuelle ZRS aufgrund sporadischen Abverkaufsverhaltens erhalten die vWG des neuen Marktes die auf Unternehmensebene bestimmten ZRS. Im letzten Schritt wird jeweils die ZRS der vWG auf die zugehörigen Artikel übertragen, so dass alle Artikel einer vWG des neuen Marktes die gleiche ZRS besitzen.

5.2 Bestimmung des fiktiven Niveauparameters

Analog zur ZRS-Generierung werden auch bei der Bestimmung des fiktiven NP relevante Störgrößeneffekte auf die Abverkaufsdaten bereinigt (Abb. 9). Hierfür erfolgt zunächst für jeden Referenzmarkt nach Artikeln getrennt eine *Selektion der JAM-Daten*, die für die Bestimmung der JAM des neuen Marktes geeignet sind. Nicht ausgewählt werden insbesondere Abverkaufsdaten von Artikeln, bei denen der Abverkaufszeitraum eingeschränkt (vgl. 4.3) oder der Anteil der Werbeabverkäufe zu hoch ist (vgl. 4.4). Anschließend werden die selektierten JAM-Daten – sofern es sich um beworbene Artikel handelt – durch die *Bereinigung der JAM-Daten* um den Anteil der „reinen“ Werbeabverkaufsmengen reduziert (vgl. 4.4).

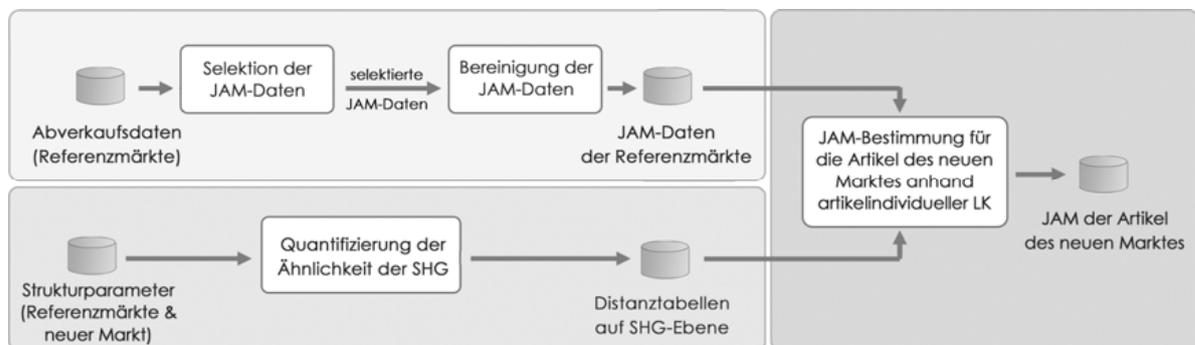


Abb. 9: Bestimmung des fiktiven Niveauparameters

Im Gegensatz zur ZRS-Generierung können die JAM-Daten nicht auf Basis einer SHG-einheitlichen, sondern ausschließlich auf einer artikelindividuellen Auswahl ähnlicher Referenzmärkte ermittelt werden. Angesichts des Sortimentsumfangs bei Einzelhandelsketten ist jedoch für jeden einzelnen Artikel eine Ähnlichkeitsanalyse mit einem Clusterverfahren nicht praktikabel. Statt dessen wird mit Hilfe der Euklidischen Distanz die Ähnlichkeit der Referenz-SHG für jede SHG des neuen Marktes quantifiziert. Pro SHG des neuen Marktes erhält man somit eine Distanztabelle, welche die SHG der Referenzmärkte mit absteigender Ähnlichkeit auflistet.

Für die artikelindividuelle Erstellung der LK kommen grundsätzlich diejenigen SHG in der Distanztabelle in Frage, deren JAM-Daten für den betreffenden Artikel die eingangs beschriebenen Selektionskriterien erfüllen. Von diesen werden – ähnlich zu der im Rahmen der ZRS-Generierung ermittelten Anzahl von SHG je Cluster – die jeweils n ähnlichsten ausgewählt und ihre Distanzen in Gewichte transformiert. Diese Gewichte fließen schließlich in die LK für jeden Artikel zur Bestimmung der fiktiven JAM des neuen Marktes ein.

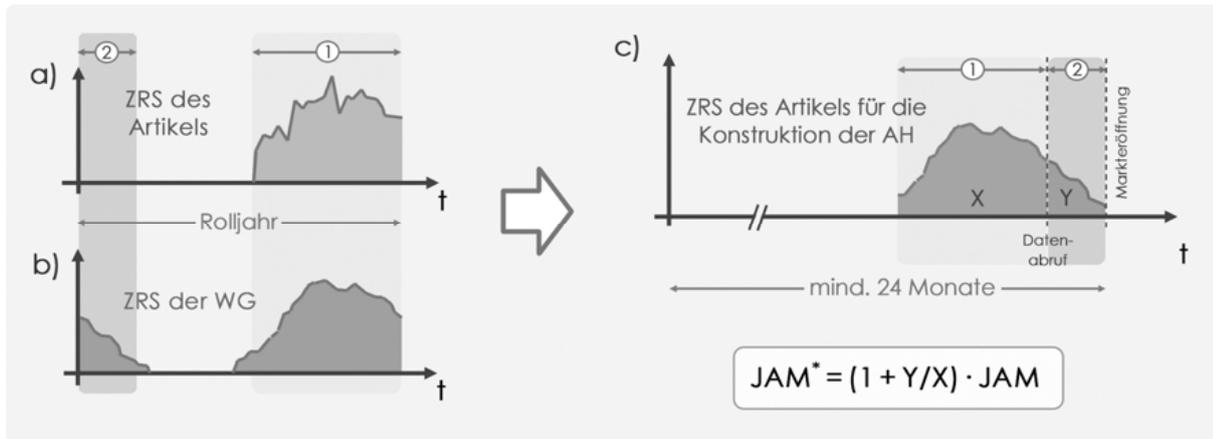


Abb. 10: Modifikation der JAM neu gelisteter Artikel

Eine Ausnahme bilden neu gelistete Artikel, für die noch keine Abverkaufsdaten über das gesamte letzte Rolljahr vorliegen (Abb. 10a). Ihnen wird daher, wie in Kapitel 4.2 dargelegt, die ZRS der zugehörigen WG zugeordnet (Abb. 10b), so dass bei der Konstruktion der AH auch das Abverkaufsverhalten im Zeitraum zwischen Datenabruf und Markteröffnung beschrieben werden kann (Abb. 10c). Die fiktiven JAM-Daten hingegen können für jeden Artikel nur über ihren tatsächlichen Verkaufszeitraum bestimmt werden und beziehen sich folglich auf eine kürzere Zeitspanne als die ZRS. Die JAM wird daher im Verhältnis des Flächenzuwachses erhöht, der sich durch die Erweiterung der ZRS ergeben hat.

5.3 Konstruktion der fiktiven Abverkaufshistorie

Nachdem nun die fiktiven ZRS und NP der Artikel des neuen Marktes für den Zeitraum des letzten Rolljahres vorliegen, sind die betreffenden Teilergebnisse gemäß den Anforderungen des eingesetzten Dispositionssystems (vgl. 2.2) zu kombinieren. Hierfür werden die bisher auf Kalenderwochenebene verdichteten ZRS des Rolljahres durch Spiegelung (vgl. 4.2) auf den gewünschten Zeitraum erweitert. Dies bezieht sich zum einen auf die Zeit zwischen dem Datenabruf und dem letzten fiktiven Verkaufstag vor Markteröffnung. Zum anderen muss der Zeitraum über das Rolljahr hinaus auf die vom Dispositionssystem geforderte maximale Länge der AH bzw. bis zum jeweiligen unternehmensweiten Listungsdatum der Artikel erweitert werden.

Im letzten Schritt müssen die bisher auf Kalenderwochen aggregierten fiktiven Abverkäufe für das Dispositionssystem auf Wochentage heruntergebrochen werden. Dazu wird über den Zeitraum eines Rolljahres für jede Warengruppe eine Gesamtverteilung der Abverkäufe auf

Wochentagbasis ermittelt. Gemäß dieser Struktur erfolgt die Aufteilung der fiktiven Abverkäufe von Kalenderwochen zu Kalendertagen.

6 Ergebnisse

Bei Vorliegen ausreichender Referenzdaten werden mit dem beschriebenen Verfahren fiktive AH für neue Märkte erstellt, die vom Dispositionssystem zur Bedarfsplanung genutzt werden können. Die Voraussetzung ausreichender Referenzdaten ist bei Einzelhandelsketten mit flächendeckender Marktpräsenz durchaus gegeben, so dass der Lösungsansatz z. B. auch auf den Elektronik-, Textil- oder Lebensmittelhandel übertragbar ist. Darüber hinaus beschränkt sich der Anwendungsbereich nicht allein auf die Neueröffnung von Märkten. Auch auf eine Standortverlagerung, eine umfassende Sortimentsveränderung oder eine Angebotserweiterung bestehender Märkte einer Handelskette ist das Verfahren übertragbar.

Das Verfahren ist systemtechnisch bereits umgesetzt und kann, bis auf Überwachungs- und Korrekturschritte, automatisiert durchgeführt werden, so dass die praktische Anwendbarkeit selbst bei Sortimentsbreiten von mehreren zehntausend Artikeln sichergestellt ist. Aufgrund der Neuheit kann über längerfristige Einsatzerfahrungen noch nicht berichtet werden. Bereits in der laufenden Erprobungsphase zeichnet sich aber eine positive Abschätzung für die Verbesserung der Bedarfsplanung in der Anfangsphase von Einzelhandelsmärkten ab.

Literaturverzeichnis

- [CoGr04] Corsten, D.; Gruen, T.: Stock-Out Cause Walkouts. Harvard Business Review, Mai 2004, S.26-28.
- [CoLe05] Cottrell, M.; Letrémy, P.: Missing Values: Processing with the Kohonen Algorithm. Proc. ASMDA, 2005, Brest 2005, S.489-497.
- [Fers79] Ferstl, O. K.: Konstruktion und Analyse von Simulationsmodellen. Verlag Hain, Königsstein 1979.
- [Götz99] Götz, N. D.: Automatische Dispositionssysteme für den Handel. Joseph Eul Verlag, Köln 1999.

- [Hert99] Hertel, J.: Warenwirtschaftssysteme – Grundlagen und Konzepte. 3. Auflage, Physica-Verlag, Heidelberg 1999.
- [HiSc01] Hippner, H.; Schmitz, B.: Data Mining in Kreditinstituten – Die Clusteranalyse zur zielgruppengerechten Kundenansprache. In: Hippner, H. u. a. (Hrsg.): Handbuch Data Mining im Marketing – Knowledge Discovery in Marketing Databases. Vieweg Verlag, Wiesbaden 2001, S.607-622.
- [KiK177] Kirsch, W.; Klein H. K.: Management-Informationssysteme. Band 1 – Wege zur Rationalisierung der Führung, Kohlhammer, Stuttgart 1977.
- [Koho01] Kohonen, T.: Self-Organizing Maps. 3. Auflage, Springer, Berlin 2001.
- [KüKa01] Küsters, U.; Kalinowski, C.: Traditionelle Verfahren der multivariaten Statistik. In: Hippner, H. u. a. (Hrsg.): Handbuch Data Mining im Marketing – Knowledge Discovery in Marketing Databases. Vieweg Verlag, Wiesbaden 2001, S.131-192.
- [NeKn05] Neckel, P.; Knobloch, B.: Customer Relationship Analytics – Praktische Anwendung des Data Mining im CRM. dpunkt.verlag, Heidelberg 2005.
- [Nowa05] Nowak, A.: Prognose bei unregelmäßigem Bedarf. In: Mertens, P.; Rässler S. (Hrsg.): Prognoserechnung. 5. Auflage, Physica-Verlag, Heidelberg 2005, S.61-72.
- [Pepe01] Pepels, W.: Kommunikations-Management – Marketing-Kommunikation vom Briefing bis zur Realisation. 4. Auflage, Schäffer-Poeschel Verlag, Stuttgart 2001.
- [Pete05] Petersohn, H.: Data Mining – Verfahren, Prozesse, Anwendungsarchitektur. Oldenbourg Verlag, München 2005.
- [Sach04] Sachs, L.: Angewandte Statistik – Anwendung statistischer Methoden. 11. Auflage, Springer, Berlin 2004.
- [ScSt01] Schlittgen, R.; Streitberg, B.: Zeitreihenanalyse. 9. Auflage, Oldenbourg Verlag, München 2001.

Konzeption des Unternehmensreportings

Ein modellgestütztes Vorgehensmodell zur fachkonzeptionellen Spezifikation

Jörg Becker, Stefan Seidel, Christian Janiesch

European Research Center for Information Systems (ERCIS)

Universität Münster

48149 Münster

{becker, seidel, janiesch}@ercis.de

Abstract

Die strukturierte Konzeption des Reportings als Grundlage für dessen Umsetzung und Optimierung stellt eine Herausforderung dar, die von der der Einführung operativer Softwaresysteme abweicht. Es werden nicht nur aussagekräftige Modellierungssprachen benötigt, die als Diskussionsgrundlage für Fachanwender und Systementwickler dienen, sondern auch detaillierte Vorgehensmodelle, die auf Basis theoretisch fundierter Methoden einen Lösungsweg bieten. Die Einführung von Business-Intelligence-Systemen mithilfe konzeptioneller Modelle verspricht Kosten- und Zeiteinsparungen, indem Iterationen des sonst üblichen prototypischen Vorgehens ausgelassen werden können. Das hier vorgestellte Vorgehensmodell umfasst die vier Phasen Initialisierung, Ist-Analyse, Soll-Konzeption und Wartung. Es wird anhand von Kurzbeispielen aus einem Evaluationsprojekt illustriert.

1 Anforderungen an ein modernes Reporting

Das Berichtswesen¹ ist die Schnittstelle zwischen strategischem Controlling, operativem Controlling und der Ausführungsebene. Folglich kommt dem Berichtswesen eine wichtige Rolle im Hinblick auf die Erreichung der Ziele des Controllings zu. Trotz dieser großen Bedeutung ist festzustellen, dass effiziente Berichtswesen heute eher die Ausnahme als die Regel sind [Wani02]. Ein wesentlicher Grund hierfür ist darin zu sehen, dass dem Management häufig nicht diejenigen Informationen zur Verfügung stehen, die für eine bestmögliche Entscheidungs-

¹ Die Begriffe Berichtswesen und Reporting werden in diesem Beitrag synonym verwendet.

findung nötig sind. Gleichzeitig kommt es oft zur so genannten Informationsüberflutung – d. h., den Entscheidungsträgern steht eine so große Menge an Informationen zur Verfügung, dass es kaum noch möglich ist, die relevanten herauszufiltern. Daraus ergibt sich die Forderung nach einer möglichst umfassenden Ermittlung und Spezifikation der Informationsbedarfe, um diese Defizite zu vermeiden.

Eine strukturierte Konzeption des Berichtswesens als Grundlage für die Umsetzung und Optimierung stellt somit eine Herausforderung für Unternehmen dar. Dabei werden nicht nur aussagekräftige Modellierungssprachen, die auch als Diskussionsgrundlage für Fachanwender und Systementwickler dienen, benötigt, sondern auch detaillierte Vorgehensmodelle, die auf Basis theoretisch fundierter Methoden einen Lösungsweg bieten.

Zu diesem Zwecke werden zunächst die notwendigen Grundlagen für das Vorgehensmodell erarbeitet: Nach Bestandsaufnahme moderner Berichtssysteme für das Unternehmensreporting und Grundlagen der Informationsbedarfsanalyse wird die dem Vorgehensmodell zugrundeliegende Modellierungssprache H2 kurz eingeführt. Das eigentliche Vorgehensmodell umfasst die Phasen Initialisierung, Ist-Analyse, Soll-Konzeption und Wartung. Es wird anhand von Kurzbeispielen aus einem Evaluationsprojekt illustriert.

2 Informationsbedarfsanalyse und Modellierung für Reporting-Systeme

2.1 Moderne Reporting-Systeme und Einordnung in das Controlling

Dem vorliegenden Beitrag wird die *koordinationsorientierte* Controlling-Konzeption von Küpper zugrunde gelegt [Küpp01]. Dieses Verständnis umfasst die Funktionen der *informationsorientierten* sowie der *planungs- und kontrollorientierten* Controlling-Konzeptionen [Meng95]. Die koordinationsorientierte Controlling-Konzeption geht darüber hinaus und betrachtet die Koordination aller *Führungsteilsysteme*. Führungsteilsysteme sind z. B. das *Planungssystem*, das *Kontrollsystem*, das *Personalführungssystem*, das *Organisationssystem* und das *Informationssystem*. Die Zerlegung der IT in Führungsteilsysteme ist aufgrund der wachsenden Komplexität der Führungsaufgaben notwendig [Koch94]. Gleichzeitig erwächst aus dieser Zerlegung die Notwendigkeit der Koordination, da zwischen den einzelnen Teilsystemen Interdependenzen bestehen. Diese erfordert die Ermittlung des Informationsbedarfs der einzelnen Systeme, die Gestaltung der Informationserzeugung sowie eine anwenderorientierte Informationsbereit-

stellung durch das *Berichtswesen* [Küpp01]. Dieses stellt somit das „Bindeglied des Informationssystems zu anderen Subsystemen der Führung“ [Glad03] dar. *Berichte* sind „unter einer übergeordneten Zielsetzung, einem Unterrichtszweck zusammengefasste Informationen“ [Blohm80].

Weder eine weitgefaste Definition des Reportings, welche alle Informationsübermittlungsprozesse der Unternehmung als Berichtswesen versteht (z. B. Blohm [Blohm70; Blohm80]), noch eine eng gefaste Definition, die lediglich die Übermittlung von Informationen in Form von Berichten als Gegenstand des Berichtswesens betrachtet (z. B. Horváth [Horv03]) ist treffend. Eine bedarfsgerechte Informationsversorgung der Entscheidungsträger kann in der Regel nur durch die Ergänzung herkömmlicher Berichte durch geeignete Analyse- und Auswertungssichten erfolgen. Folglich wird dieser Arbeit eine eigene Definition zugrunde gelegt: Das Berichtswesen ist der Teil des betrieblichen Informationssystems, welcher der Übermittlung und Bereitstellung von Informationen für innerbetriebliche Empfänger mittels Berichten im oben definierten Sinne sowie der Bereitstellung geeigneter Analyse- und Auswertungssichten auf einen verdichteten Datenbestand, der alle für den operativen und strategischen Geschäftsbetrieb relevanten Informationen enthält, dient.

Demzufolge werden heute neben herkömmlichen Standardberichten vor allem auch ein so genanntes Exception-Reporting sowie die Möglichkeit von Ad-hoc-Auswertungen und weiteren Analysen gefordert. Einfach zu bedienende, moderne Führungsinformationssysteme sowie ihnen zugrunde liegende Technologien wie Data-Warehousing und Online Analytical Processing (OLAP) bieten die entsprechende technische Grundlage für komplexe multidimensionale Analysen. Diese können jedoch nur dann erfolgreich sein, wenn die Datenbasis und die Kriterien, nach denen ausgewertet wird, richtig und widerspruchsfrei spezifiziert sind.

2.2 Informationsbedarfsanalyse

Um den oben genannten Anforderungen an ein Berichtswesen gerecht zu werden, ist die Kombination verschiedener Verfahren der Informationsbedarfsanalyse nötig. Zum einen müssen angebotsorientierte Verfahren eingesetzt werden, in deren Rahmen eine Analyse des Ist-Zustandes durchgeführt wird. Dies ist notwendig, um Schwachstellen zu identifizieren und eine Kommunikationsgrundlage für Entwickler und Fachanwender zu schaffen.² Daneben müssen

² Hier sind zwei Fälle zu unterscheiden: Golfarelli und Rizzi [GoRi98] verstehen unter angebotsorientierten Verfahren solche, bei denen ausgehend von einer Analyse der bestehenden operativen Systeme logische und semantische Schemata abgeleitet werden. Dabei handelt es sich um ein Vorgehen, das auf die Bestimmung des objektiven In-

nachfrage- oder aufgabenorientierte Verfahren verwendet werden, die der Ermittlung des Soll-Bedarfes dienen. List et. al. [LBMS02] begründen die Notwendigkeit der aufgabenorientierten Informationsbedarfsanalyse insbesondere damit, dass nur durch die Berücksichtigung zukünftiger Informationsbedarfe auch eine lange Lebensdauer für das zu entwickelnde System zu erwarten ist. Nachfrageorientierte Verfahren sind dabei in der Regel verhältnismäßig einfach durchzuführen und verbessern die Akzeptanz des zukünftigen Systems seitens der Systemnutzer.

Aufgrund der Verschiedenartigkeit von Projekten und der jeweils unterschiedlichen Rahmenbedingungen ist nicht davon auszugehen, dass ein Verfahren, das anhand bestimmter Kriterien als „das Beste“ identifiziert wurde, auch immer eingesetzt werden kann. Daher wird hier vorgeschlagen, grundsätzlich verschiedene Verfahren in Betracht zu ziehen und im konkreten Fall zu entscheiden, welches Verfahren bzw. welche Verfahren eingesetzt werden.

Die Informationsbedarfsanalyse muss darüber hinaus die fachkonzeptionelle Spezifikation der Informationsbedarfe unterstützen. Zur Explikation der Informationsbedarfe ist daher eine Methode nötig, die sowohl von den Berichtsempfängern als auch von Entwicklern verstanden wird, um eine reibungslose Kommunikation und somit eine zielführende Diskussion der entwickelten Modelle zu ermöglichen, da ein Hauptproblem darin zu sehen ist, dass die späteren Benutzer des Systems selten in der Lage sind, ihre Informationsbedarfe konkret auszudrücken und zu spezifizieren [Stra02; StWi02]. Im Folgenden wird zunächst eine entsprechende Modellierungsmethode, H2, vorgestellt, deren Anwendung im Rahmen eines Vorgehensmodells dann in Abschnitt 3 erläutert wird. Zur Auswahlentscheidung zugunsten von H2 und zum Unterschied zu vergleichbaren Modellierungsmethoden vgl. [BJPS06; KnSJ06].

2.3 Modellierung mit H2

Die H2-Methode basiert auf dem Prinzip, dass Kennzahlen (bspw. Gewinn, Lagerreichweite, Produktionskosten) und Dimensionen (bspw. Zeit, Produkte, Organisationsstruktur) unabhängig voneinander definiert werden können und aus der Kombination dieser Bausteine der Inhalt von Datenwürfeln (Cubes) sowie konkreten Berichten beschreibbar ist [BJKM07]. Dabei ergibt sich ein konkreter Fakt (bspw. Gewinn der Filiale Karlsruhe im Warenbereich Kleinuhren im aktuellen Geschäftsjahr) aus der berichtsspezifischen Zuordnung von Bezugsobjekten zu einer Kennzahl. Um festzulegen, welche Bezugsobjekte für einen bestimmten Analysezweck vonnöten

formationsbedarfes abzielt. Im vorliegenden Beitrag wird im Rahmen der Ist-Analyse das Informationsangebot des Ist-Zustandes des Berichtswesens betrachtet. Dabei handelt es sich im Wesentlichen um eine Dokumentenanalyse zur Bildung semantischer Schemata.

sind, werden Ausschnitte aus Dimensionen gebildet (Dimensionsausschnitte) und diese zu mehrdimensionalen Dimensionsausschnitt-Kombinationen (Bezugsgrößen) kombiniert. Die Methode erlaubt weiterhin, Fakten als Operanden in so genannten Faktberechnungen zu verwenden. Auf diese Weise können beispielsweise Abweichungsanalysen und Anteilsrechnungen abgebildet werden. Die Dokumentation erfolgt bei der H2-Methode in Form eines grafischen Modells basierend auf hierarchischen Strukturen. Vgl. Abb. 1 für ein kurzes Beispiel eines Datenwürfels sowie eine Legende der verschiedenen Elemente (Faktberechnungen sind nicht enthalten); Berichte werden auf ähnliche Weise modelliert.

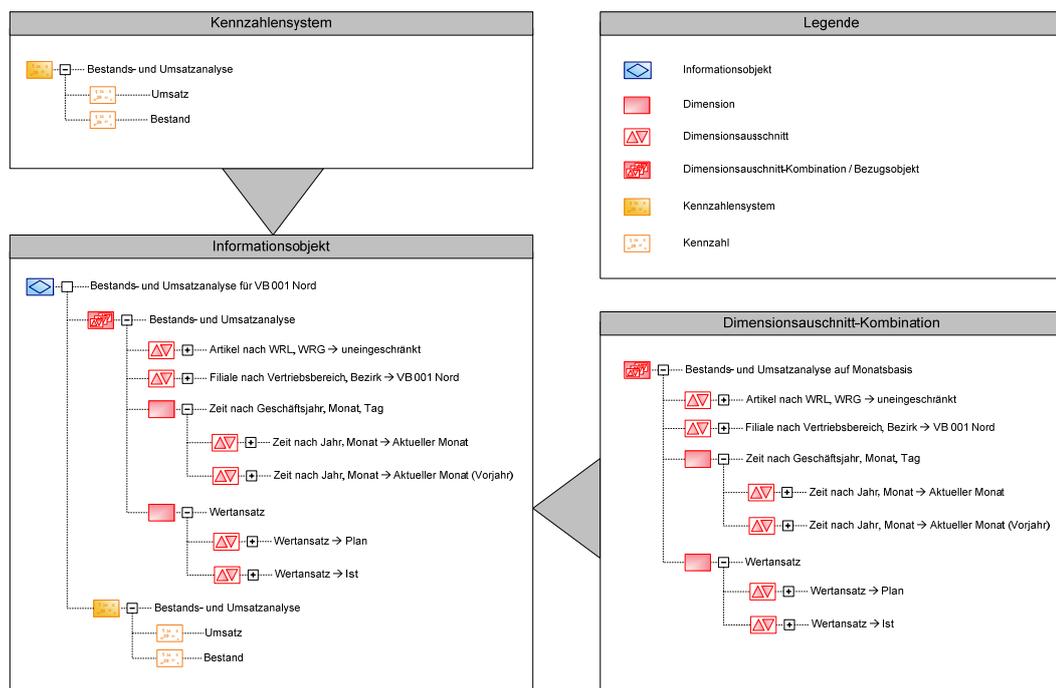


Abb. 1: H2-Modell

Die Anwendung dieser Methode im Rahmen eines entsprechenden Vorgehensmodells (vgl. Abb. 2) führt zu einer eindeutigen und transparenten Dokumentation der verwendeten Kennzahlen und Dimensionen im Bericht; Interpretationsspielräume werden minimiert. Die fachkonzeptionellen Modelle bilden dabei zum einen die Grundlage für die Diskussion zwischen Fachanwendern und IT-Entwicklern [Holt03a]. Zum anderen besitzen sie einen Formalitätsgrad, der die Überführung in logische Data-Warehouse-Schemata sowie die Implementierung eines konkreten Reporting-Systems unterstützt [Holt03b].

3 Vorgehensmodell zur Gestaltung des Unternehmensreportings

3.1 Übersicht

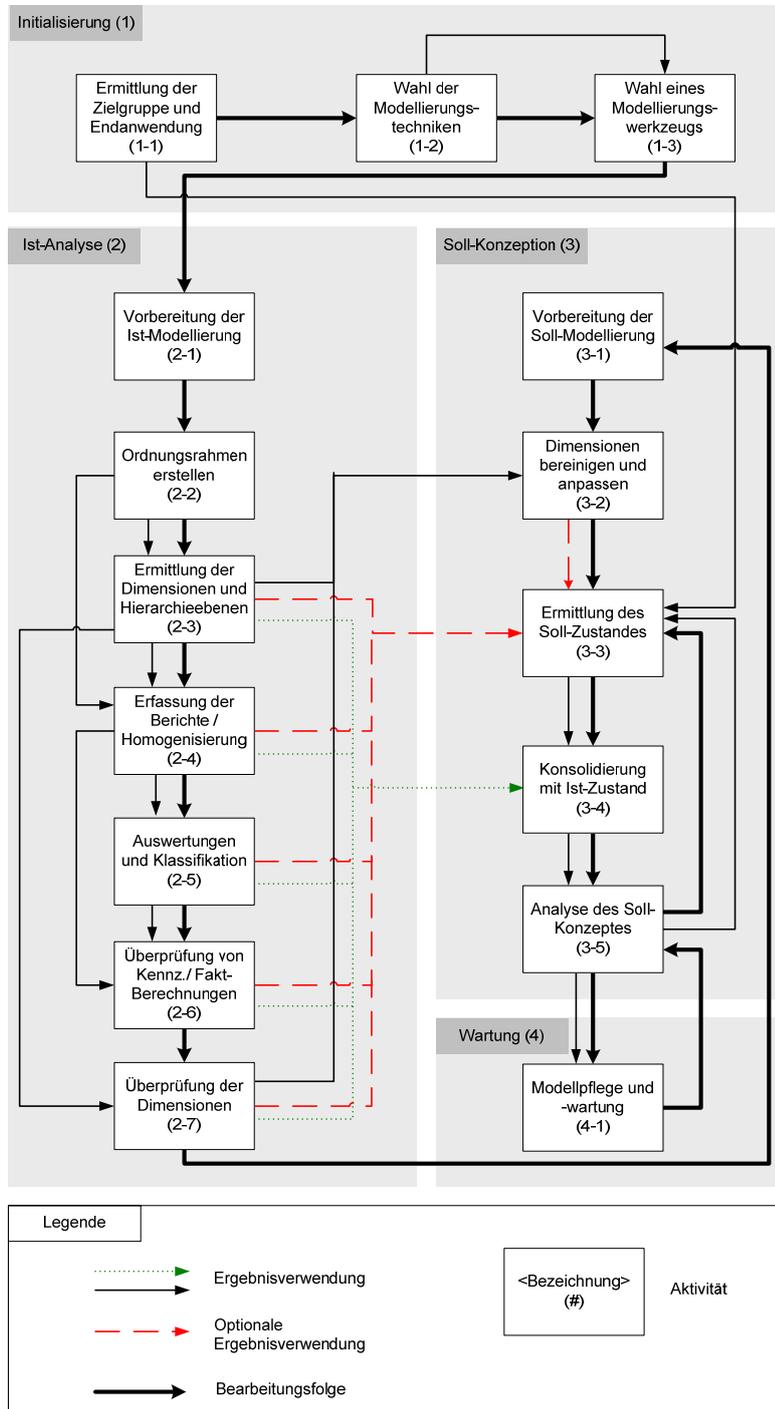


Abb. 2: Phasen und Aktivitäten des Vorgehensmodells³

³ Der Übersichtlichkeit halber wurden nicht alle Ergebnisverwendungen mit aufgenommen. So haben die Ergebnisse der Vorbereitung der Ist-Modellierung beispielsweise Auswirkungen auf nahezu alle Aktivitäten der Phase der Ist-Analyse. Gleiches gilt für die Vorbereitung der Soll-Modellierung. Aus Platzgründen werden im Folgenden nur die Phasen 2 und 3 detailliert dargestellt. Die Modellierungsentscheidung zugunsten H2 wird vorausgesetzt.

Das Vorgehensmodell umfasst die Phasen *Initialisierung*, *Ist-Analyse*, *Soll-Konzeption* und *Wartung*. Die einzelnen Phasen werden weiter unterteilt in *Aktivitäten*. In der Initialisierungsphase werden grundsätzliche Fragen bzgl. der bestehenden und neu zu entwickelnden Systeme behandelt. Im Rahmen der Ist-Analyse erfolgt eine – mehr oder weniger – detaillierte Erfassung des Ist-Zustandes. Dabei werden sowohl Stammdaten wie z. B. Produktstrukturen als auch Metadaten über die einzelnen Berichte erfasst. Das so entstandene fachkonzeptionelle Informationsmodell des Berichtswesens kann dann für Analysezwecke ausgewertet werden. Im Rahmen der Soll-Konzeption werden verschiedene Verfahren der Informationsbedarfsanalyse eingesetzt, um den Soll-Zustand des Berichtswesens zu bestimmen. Dabei können – abhängig vom Projekt – die Ergebnisse der Ist-Analyse als Grundlage für die Soll-Konzeption herangezogen werden. Darauf folgt eine Konsolidierung mit den Ergebnissen aus der Ist-Erhebung, in der das Soll-Fachkonzept verfeinert und ergänzt wird. Gegenstand der Phase der Wartung ist schließlich die kontinuierliche Pflege und Anpassung der Modelle. Diese sind im Zeitablauf immer wieder auf ihre Qualität hin zu überprüfen, was durch den Pfeil in Richtung der Analyse des Soll-Konzepts ausgedrückt wird. Die gestrichelten Verbindungslinien in Abb. 2 deuten an, dass die Verwendung der Ist-Modelle bzw. deren Auswertungen für die Soll-Konzeption optional sind. In jedem Fall aber ist eine Konsolidierung von Ist- und Soll-Zustand im Anschluss an die Soll-Konzeption vorzunehmen, um sicherzustellen, dass Fehler, Schwachstellen und Inkonsistenzen des Ist-Zustandes vermieden und gleichzeitig keine relevanten Informationen vergessen werden. Die Aktivitäten haben jeweils eine Nummer, diese bilden die Reihenfolge der Durchführung der einzelnen Aktivitäten ab und dienen als Orientierungshilfe.

Zur Illustration einzelner Phasen werden Beispiele aus einem Evaluationsprojekt mit der Christ Juweliere und Uhrmacher seit 1863 GmbH (im Folgenden Christ) verwendet. Im Rahmen dieses Reporting-Projekts wurde sowohl der Aufbau des Vorgehensmodells bestätigt als auch die Verwendbarkeit der fachkonzeptionellen Informationsmodelle überprüft. Kern des Projekts ist die Neugestaltung des Reportings und die Überführung des alten proprietären Berichtswesens (CHRIST A) in ein integriertes Data Warehouse, das in Zukunft laufend durch Daten des neuen Warenwirtschaftssystems Lawson (Intentia) Movex gefüllt werden soll. Zur Darstellung der Berichte werden Produkte der Firma Cognos eingesetzt werden: Cognos ReportNet für das Standard-Berichtswesen und Cognos Powerplay für die deskriptive, multidimensionale Datenanalyse per OLAP. Das Projekt befindet sich derzeit in der Implementierungsphase, die fachkonzeptionelle Spezifikation ist abgeschlossen.

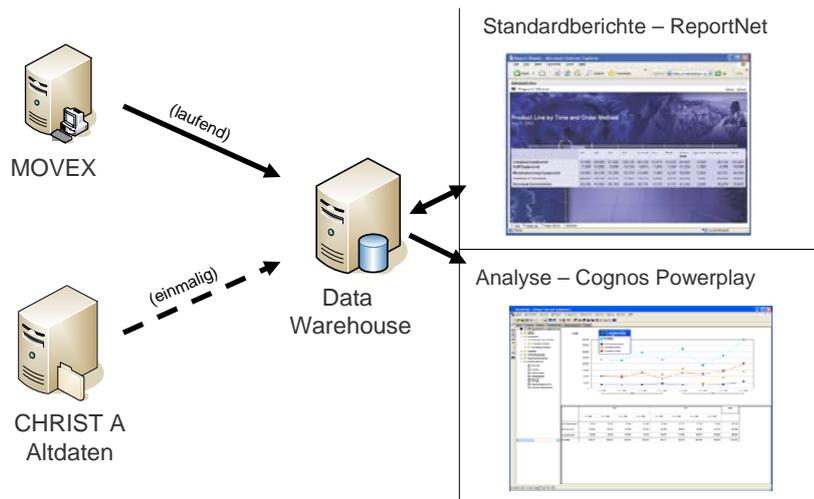


Abb. 3: Vorgehen bei der Christ Juweliers und Uhrmacher seit 1863 GmbH

3.2 Initialisierung

Ermittlung der Zielgruppe und Endanwendung (1-1), Wahl der einzusetzenden Methoden und Modellierungstechniken (1-2) und Wahl eines Modellierungswerkzeugs (1-3)

Es muss zunächst der Anwendungszweck bestimmt werden. In diesem Zusammenhang müssen vor allem die Zielgruppen des neuen Systems festgelegt werden. Dazu ist zu klären, wer die Berichtsempfänger sind und in welchen Positionen sie sich befinden. Unter Umständen ist es sinnvoll, sich aus Gründen der Komplexitätsverminderung zunächst auf einen Teilbereich des Unternehmens zu beschränken [Stra02]. Neben der Unterstützung der eigentlichen Modellierung und der Pflege und Wartung der Modelle sollten auch Werkzeuge zur Definition, Ausführung, Darstellung und Überwachung von Vorgehensmodellen zur Verfügung stehen.

3.3 Ist-Analyse

Vorbereitung der Ist-Modellierung (2-1) und Erstellung eines Ordnungsrahmens (2-2)

Die Ist-Modellierung kann mit einem hohen Aufwand verbunden sein. Bevor die Ist-Modellierung durchgeführt wird, ist entsprechend festzulegen, mit welcher Genauigkeit und Detaillierung dies geschehen soll. Insbesondere spielt hier die Einschätzung darüber eine Rolle, in welchem Maße Teile des Ist-Zustandes in die Soll-Konzeption übernommen werden können. Zu Vorteilen der Ist-Modellierung vgl. [BeKR05; Stra02], die Zweckmäßigkeit einer detaillierten Ist-Modellierung muss daher am Umfang des prognostizierten Handlungsbedarfes festgemacht werden [BeKR05; BeSc04; Rose00].

Daneben sind Modellierungskonventionen festzulegen. Durch die Reduktion von Freiheitsgraden bei der Modellierung wird eine Vergleichbarkeit der Modelle gewährleistet und Inkon-

sistenzen werden vermieden. Dies ist insbesondere auch in Bezug auf die Vergleichbarkeit von Soll- und Ist-Modellen zu fordern. Weiterhin muss der zu betrachtende Bereich des Berichtswesens systematisiert und zerlegt werden. Idealerweise wird an dieser Stelle ein *Ordnungsrahmen* erzeugt. Den am Projekt Beteiligten wird somit ermöglicht, den eigenen Arbeitsbereich in den Gesamtzusammenhang einzuordnen.

Ermittlung der Dimensionen und Hierarchieebenen (2-3)

Ziel dieser Aktivität ist es, die Dimensionen, die zum Zeitpunkt der Ist-Erhebung in den Berichten verwendet werden, zu erfassen. Existierende Dokumentationen sowie eine Betrachtung der Berichtsdefinitionen bzw. -outputs können eine Ausgangsbasis darstellen. Ebenso ist eine Betrachtung der entsprechenden Stammdaten in den operativen Systemen möglich. Existiert in der Unternehmung bereits ein Data Warehouse, so können auch die entsprechenden Data-Warehouse-Strukturen Aufschluss über den Aufbau der Hierarchien geben.

Bei der Erhebung der Informationsräume von Christ wurde die Struktur der über 100 Berichte untersucht und so die bislang übliche Dimensionierung der Berichte des Altsystems als auch der Excel-Berichte erfasst. Die übliche Dimensionierung der *Artikel* und *Filialen* des Altsystems sind auf Typ- und Instanzebene in Abbildung 4 dargestellt. Hierarchieebenen dienen der Strukturierung der Dimensionen. Die detaillierte Darstellung der Dimensionen kann häufig automatisiert aus den Altsystemen ausgelesen werden (bspw. mit SQL) und unterstützt die Kommunikation mit dem Fachpersonal besser als die abstraktere Darstellung der Hierarchieebenen.

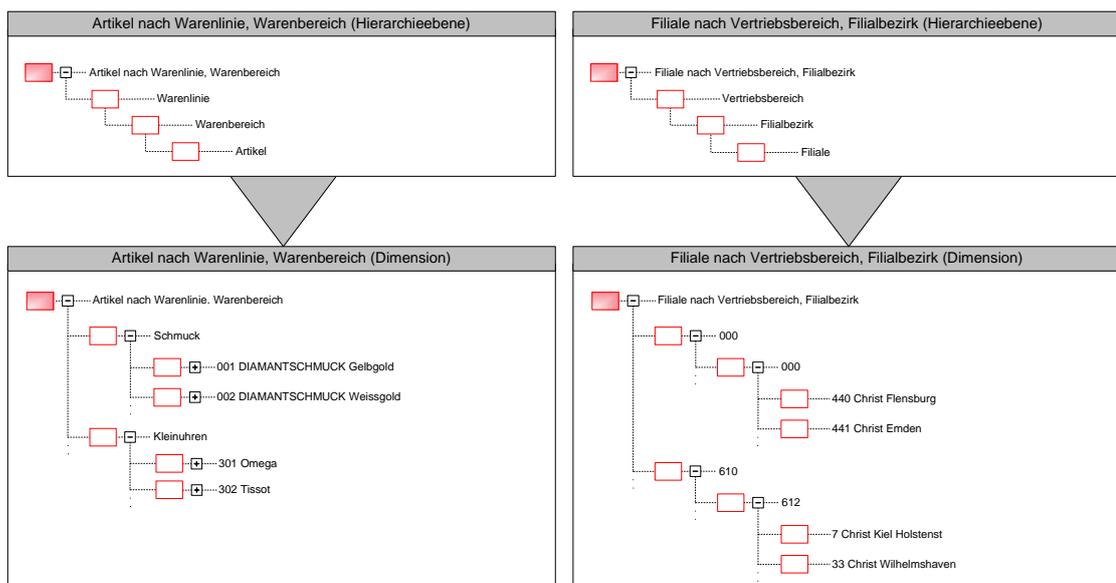


Abb. 4: Erfassung der Dimensionen und Hierarchieebenen

Erfassung der Berichte und Identifikation von Synonymen und Homonymen (2-4)

Berichte werden aufgenommen und strukturiert in einem Informationsmodell abgebildet. Von den Berichtsverantwortlichen ist eine Übersicht über die von ihnen erstellten Berichte bzw. Auswertungen zu erstellen [Stra02]. Die Metadaten über die einzelnen Berichte bilden die Grundlage für verschiedene Analysezwecke. Folgende Metadaten sind insbesondere relevant: Bezeichnung, Kurzbeschreibung, Kennzahlen sowie zugehörige Dimensionen, Berichtsempfänger, Berichtsverantwortlicher, Periodizität und die Anwendung, die den Bericht erzeugt.

Auch die Erfassung von synonymen und homonymen Bezeichnungen für Kennzahlen und Dimensionen wird im Rahmen der Berichtserfassung durchgeführt. Jeweils synonyme Bezeichnungen sind dabei durch eindeutige Begriffe für inhaltlich gleiche Objekte zu ersetzen. Homonyme liegen dann vor, wenn derselbe Begriff mit unterschiedlicher inhaltlicher Bedeutung verwendet wird. Hier müssen zwei verschiedene Bezeichnungen gefunden werden.

Es fanden sich bei Christ einige Beispiele, die jeweils von anderen Empfängern oder Berichtserstellern gepflegt wurden: *VK % erz.* und *Marge in %* standen beide für die erzielte Marge, auch *Lagerbestand* und *Bestand Stk.* wurden bspw. synonym verwendet. Homonyme fanden sich eher unabsichtlich durch Abweichungen im Berechnungsausdruck.

Auswertungen und Klassifikation (2-5)

Die verschiedenen Modellelemente des Informationsmodells des Ist-Zustandes können zu Auswertungen herangezogen werden. Diese führen zu einer so genannten *Informationslandkarte* [Stra02]. Sie kann im Rahmen der Soll-Konzeption als Diskussionsgrundlage für die Modellierung oder zur Aufdeckung von Redundanzen verwendet werden. Die Auswertungen bieten für sich genommen keine Möglichkeit, Aussagen über die inhaltliche Qualität des Ist-Berichtswesens zu treffen. Um festzustellen, inwieweit dieses bedarfsgerecht ist, muss eine Informationsbedarfsanalyse durchgeführt werden – dies ist Gegenstand der Soll-Konzeption.

Bereits eine einfache quantitative Analyse der Kennzahlenverteilung in den Berichten (vgl. Abbildung 5) zeigt deutlich, dass es eine kleine Anzahl von Kennzahlen gibt, die häufig verwendet wird und eine überwiegende Anzahl von Kennzahlen nur sehr selten, bspw. in Jahresabschlüssen, erhoben wird. Die Betrachtung dieser Kennzahlencluster gibt Aufschlüsse darüber, welche Kennzahlen zukünftig im Data Warehouse vorberechnet werden sollten und welche Kennzahlen zur Laufzeit berechnet werden können. Weiterhin kann die Auswertung Entscheidungsunterstützung für die zukünftige Darstellung von Synonymen und Homonymen leisten.

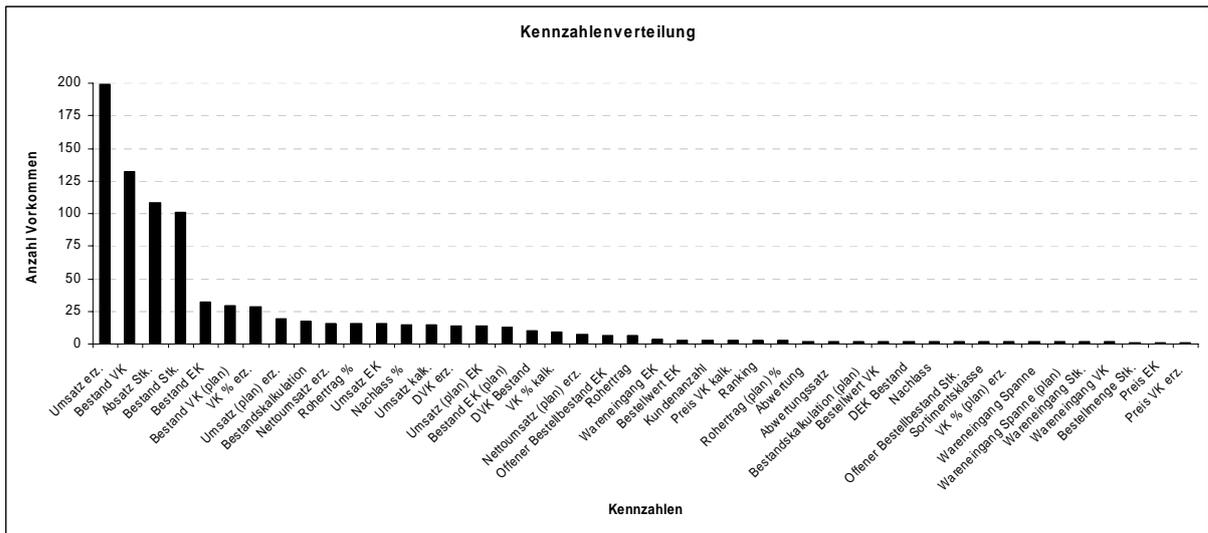


Abb. 5: Quantitative Analyse der Kennzahlenverteilung

Überprüfung von Kennzahlen (2-6)

Für berechnete Kennzahlen und sonstige Berechnungsausdrücke werden die Formeln auf inhaltliche Richtigkeit hin untersucht. Dies erfordert in der Regel die Zusammenarbeit mit den Fachbereichen – unter Umständen kann es erforderlich sein, bis auf Quellcodeebene zu gehen.

Bei Christ wurde neben Berechnungsmängeln im Detail, die im Rahmen eines Berichtswesens auf Excelbasis und einem gewachsenen, proprietären System kaum auszuschließen sind, insgesamt eine relativ hohe Datenqualität festgestellt. Zur genaueren Identifikation der Mängel wurde bspw. auch detailliert der Aufbau der einzelnen Kennzahlen analysiert und Zusammenhänge herausgearbeitet. Abbildung 6 zeigt einen Ausschnitt der Berechnung üblicher Kennzahlen unter Verwendung von Basiskennzahlen. Inhaltsgleiche Kennzahlen dieser Darstellung sind in grau schattiert und werden konsolidiert.

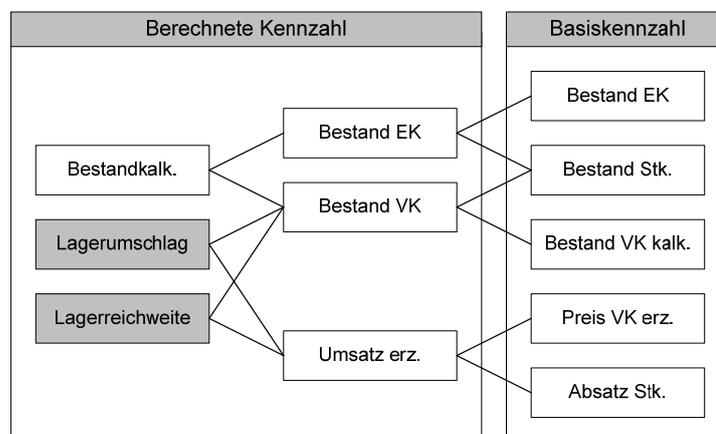


Abb. 6: Kennzahlenanalyse

Überprüfung der Dimensionen (2-7)

Die erfassten Dimensionen müssen die tatsächlichen Gegebenheiten im Unternehmen abbilden.⁴ Dies kann z. B. im Rahmen einer Diskussion des Berichtswesen-Analysten mit den Berichtsverantwortlichen und -empfängern überprüft werden, aber auch durch eine Untersuchung der Quellsysteme. In einer Mängelliste werden die unzureichend bzw. falsch abgebildeten Sachverhalte geführt.

Bei Christ entstanden durch die strukturierte Darstellung der Informationsräume bereits an dieser Stelle viele Ideen und Hinweise, was bei der Darstellung im neuen System ergänzt und verändert werden sollte.

3.4 Soll-Konzeption

Vorbereitung der Soll-Konzeption (3-1)

Analog zur Ist-Modellierung ist auch bei der Soll-Modellierung der Detaillierungsgrad zu wählen. Die Soll-Modelle bilden dabei die Grundlage für ein Berichtswesen-Inventar, das kontinuierlich zu pflegen ist. Je detaillierter die Soll-Modellierung erfolgt, umso teurer wird folglich auch die Pflege des Berichtswesen-Inventars. Andererseits ermöglicht ein hoher Detaillierungsgrad eine genaue Spezifikation der Informationsbedarfe und erleichtert die Umsetzung des Fachkonzepts. Wie im Rahmen der Ist-Modellierung werden für die Soll-Modellierung Modellierungs- und Namenskonventionen festgelegt, um die Lesbarkeit des Gesamtmodells zu erhöhen.

Dimensionen bereinigen und anpassen (3-2)

Ziel dieser Aktivität ist es, die im Rahmen der Ist-Analyse erfassten und analysierten Dimensionen zu bereinigen und anzupassen. Das bedeutet, dass z. B. Widersprüchlichkeiten beseitigt oder Bezugsobjekte, die nicht mehr existieren, entfernt werden müssen. Ziel dieser Aktivität ist es nicht, den Soll-Zustand zu erheben, sondern sachliche Fehler aus den Modellen des Ist-Zustandes zu entfernen.

Im Altsystem von Christ fanden sich zahlreiche Artikel, Filialen etc. die inzwischen ausgelaufen sind bzw. geschlossen wurden. Um nur aktuelle bzw. relevante Daten in das neue Data Warehouse zu übernehmen, wurden diese Dimensionen bereinigt. Abbildung 7 zeigt dies am Beispiel der Dimension *Filiale*. Nicht mehr existierende Filialen wurden aus Filialbezirken entfernt.

⁴ Dies bedeutet nicht, dass die Dimensionen tatsächliche Produkt- oder Organisationsstrukturen darstellen müssen. Es kann sich um Analyse- und Auswertungssichten handeln. Diese müssen jedoch inhaltlich richtig sein.

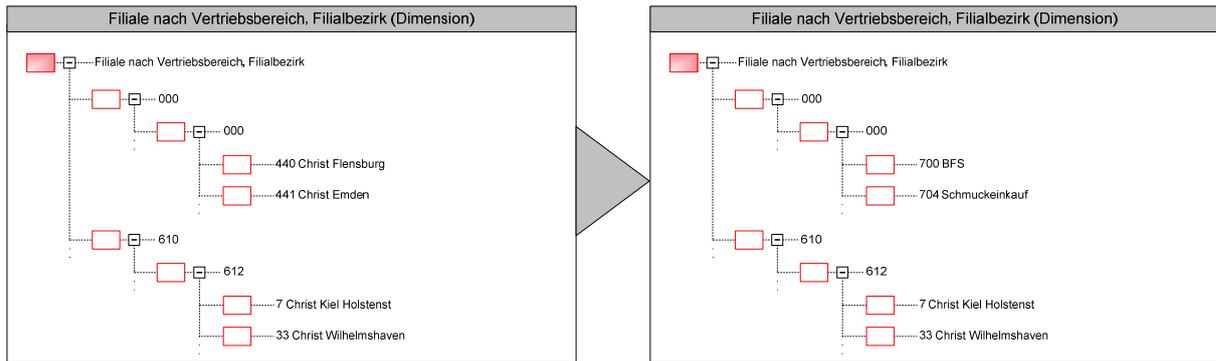


Abb. 7: Dimensionsbereinigung

Ermittlung des Soll-Zustandes (3-3)

Für die Ermittlung des Soll-Zustandes können verschiedene Verfahren der Informationsbedarfsanalyse zum Einsatz kommen. In welchem Maße die Modelle des Ist-Zustandes für die Ermittlung des Soll-Zustandes herangezogen werden, muss individuell entschieden werden. Sie können direkt als Ausgangsbasis für die Erstellung des Soll-Konzepts dienen.

Unabhängig davon, welche konkreten Verfahren angewendet werden, sind die folgenden Konstruktionsaufgaben im Rahmen der Ermittlung des Soll-Zustandes durchzuführen:

- *Konzeption der Informationsräume*

Im Rahmen der Konzeption der Informationsräume sind zunächst die für die Berichtsempfänger relevanten Basisobjekte zu identifizieren und diese (hierarchisch) in Dimensionen zu ordnen.

- *Identifikation von Kennzahlen und Konzeption der Kennzahlensysteme*

Es werden Basiskennzahlen und daraus berechneten Kennzahlen unterschieden. Auf ihrer Basis können Kennzahlensysteme gebildet werden. Kennzahlensysteme müssen keine algebraischen Zusammenhänge ausdrücken, sondern können auch Kennzahlen nach ihrer Wichtigkeit für die Analysezwecke ordnen [Holt03a].

- *Konzeption von bedarfsgerechten Sichten auf die Informationsräume*

In Abhängigkeit von den individuellen Informationsbedarfen der Berichtsempfänger werden Sichten auf Dimensionen definiert und zu (Teil-)Informationsräumen kombiniert. Aus den (Teil-)Informationsräumen und Kennzahlensystemen können Berichte (Informationsobjekte) zusammengesetzt werden (vgl. Abbildung 1).

Es ist bei der Konzeption zu vermeiden, dass neue Synonyme bzw. Homonyme entstehen. Dies wird durch die Pflege des Glossars gewährleistet. Wenn neue Elemente benötigt werden, ist

zunächst zu prüfen, ob diese bereits Bestandteil des existierenden Soll-Modells sind. Um das Ziel eines konsistenten Berichtswesens zu erreichen, werden hier die folgenden Vorgehensweisen zur Ermittlung des Soll-Zustandes vorgeschlagen:

- *Ermittlung der Informationsbedarfe auf Basis von Ist-Modellen*

Der nachfrageorientierte Ansatz basiert auf typischen benutzerorientierten Techniken wie dem Interview oder der Fragebogenmethode. Auf diese Weise können jedoch lediglich subjektive Informationsbedarfe ermittelt werden. Den Ausgangspunkt für die Diskussion mit den Entscheidungsträgern bilden Berichtsmengen, die aufgrund verschiedener Klassifikationsmerkmale gebildet wurden, und die Auswertungen der Ist-Analyse (z. B. um redundante Informationen zu identifizieren).

- *Ermittlung der Informationsbedarfe aus Unternehmenszielen*

Informationsbedarfe von Entscheidungsträgern lassen sich auch aus den Unternehmenszielen ableiten [BeDR03]. Die abstraktesten Ziele werden in Geschäftsstrategien definiert. Ein Hauptproblem von Geschäftsstrategien ist häufig in deren nicht-operationalem Charakter zu sehen. Für operationale Ziele müssen daher eine Zielgröße, ein Zielniveau, ein Bezug und ein Zeitrahmen definiert sein [Adam96]. Da Geschäftsstrategien in der Regel nicht messbar sind, müssen diese zunächst in operationale Ziele heruntergebrochen werden.

- *Verwendung von Referenzmodellen*

Der Einsatz von Referenzmodellen hat sich z. B. bei der Entwicklung operativer Informationssysteme für den Handel als erfolgreich erwiesen [BeSc04]. Auch für die Konzeption von Führungsinformationssystemen kann die Verwendung geeigneter Referenzmodelle förderlich sein.

- *Angebotsorientierter Ansatz*

Unter angebotsorientierten Ansätzen kann nicht nur der Einbezug des Informationsangebots in Form des bestehenden Berichtswesens verstanden werden, sondern auch eine Analyse operativer Informationssysteme [GoRi98].

Bei Christ wurde häufig ein Export der Standardberichte nach Excel vorgenommen, um die Berichte mit anderen, nicht im System enthaltenen Daten anzureichern und neu zu formatieren. Ein prägnantes Beispiel für eine wiederkehrende Anreicherung findet sich in der Dimension *Artikel* (vgl. Abbildung 7).

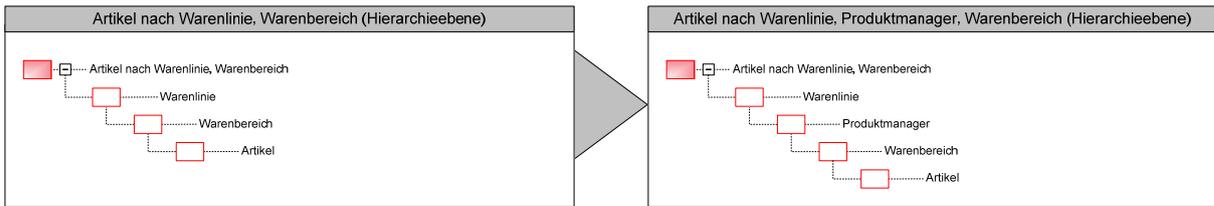


Abb. 8: Soll-Konzeption der Dimensionen

Artikel werden in Warenbereichen geführt, die wiederum Warenlinien zugeordnet sind. Für eine Auswahl mehrerer Warenbereiche einer Warenlinie ist ein Produktmanager zuständig. Dieser wurde im Altsystem nicht geführt. In fast allen Excelauswertungen wurde diese Hierarchieebene aber eingefügt. Daher wurde sie in das Soll-Konzept aufgenommen.

Konsolidierung mit Ist-Zustand (3-4)

Ziel der Konsolidierung von Ist- und Soll-Zustand ist vor allem, dass relevante Informationen, die im Ist-Zustand bereits berücksichtigt wurden, im Soll-Konzept nicht vergessen werden.

Die Schwachstellen des Ist-Berichtswesens, insbesondere Fehler, Inkonsistenzen und Redundanzen können dabei vermieden werden, da diese bereits im Rahmen der Ist-Analyse identifiziert und in entsprechenden Mängellisten vermerkt wurden. Es besteht allerdings die Gefahr, dass nicht-relevante Information im Rahmen der Konsolidierung in die Soll-Konzeption aufgenommen werden und so zu einer neuerlichen Informationsüberflutung führen.

Im Rahmen der Soll-Konzeption sind in der Regel neue Kennzahlen und Dimensionen definiert worden. Folglich sind die gewählten Bezeichnungen noch einmal mit den Entscheidungsträgern sowie den Berichtsverantwortlichen abzustimmen. Ziel dieser Zusammenführung ist es, ein Konzept zu erarbeiten, das die überflüssigen Standardbereiche eliminiert und Kapazitäten schafft, ein konsolidiertes, flexibleres Berichtswesen zu etablieren. Neben der Basis aus Standardberichten ist dann Raum für Ad-hoc Auswertungen und ein Exception Reporting

Analyse des Soll-Konzepts (3-5)

Bei der Analyse des Soll-Konzepts wird ähnlich vorgegangen wie bei der Analyse des Ist-Zustandes. Sie ist als eine Qualitätsprüfung zu verstehen, in der das erstellte Fachkonzept mit den im Vorfeld identifizierten Zielen verglichen wird. Wurden die Ziele nicht erreicht, muss es in einem iterativen Prozess entsprechend überarbeitet werden. Hier bietet sich unter Umständen ein Vorgehen basierend auf Prototypen an. Dabei ist das Fachkonzept in den jeweiligen Iterationsschritten anzupassen.

3.5 Wartung

Modellpflege und -wartung (4-1)

Die Entwicklung des Berichtswesens ist nicht mit der erfolgreichen Einführung abgeschlossen. Der sich ständig ändernde Informationsbedarf der Entscheidungsträger macht eine laufende Anpassung und Pflege des Berichtswesens erforderlich. Das Fachkonzept sollte daher nicht nur Grundlage für die einmalige Implementierung des Berichtswesens sein, sondern muss kontinuierlich gepflegt werden und ein ständig aktuelles Berichtswesen-Inventar sein. Viele der Ursachen, die zu einem nicht bedarfsgerechten Berichtswesen führen, können durch eine solche kontinuierliche Dokumentation vermieden werden. Bei sich ändernden Anforderungen muss überprüft werden, ob das Berichtswesen diese Informationen bereits an anderer Stelle bereitstellt. Nach größeren Änderungen sollte eine Analyse vorgenommen werden, um mögliche neu entstandene Schwachstellen in Form von Redundanzen und Inkonsistenzen zu identifizieren.

4 Zusammenfassung

Die strukturierte Konzeption des Berichtswesens stellt eine Herausforderung dar. Moderne Reportingsysteme können nur dann erfolgreich eingesetzt werden, wenn die Datenbasis und die Kriterien, nach denen ausgewertet werden soll, richtig und widerspruchsfrei spezifiziert sind. Die Nutzung aussagekräftiger Modellierungssprachen zur Diskussion mit den Fachanwendern unter Verwendung detaillierter Vorgehensmodelle, kann die Konzeption und anschließende Implementierung erleichtern.

H2 ist eine Methode, die die Spezifikation dieser Datenbasis und Kriterien erlaubt. In einem mehrstufigen Vorgehensmodell kann diese Methode effizient eingesetzt werden: Im Rahmen der Ist-Analyse wird das existierende Berichtswesen modelliert und analysiert. Ziel der Soll-Konzeption ist die fachkonzeptionelle Spezifikation des zukünftigen Berichtswesens. Diese ist mit dem im Rahmen der Ist-Analyse erhobenen Informationsangebot zu konsolidieren. Die Grundlage für die Konzeption des Soll-Zustandes bilden – soweit vorhanden – die Modelle und Auswertungsergebnisse der Ist-Analyse. Es schließt sich eine kontinuierliche durchzuführende Phase der Wartung und des Change Management an.

Das zentrale Artefakt des Vorgehensmodells ist dabei ein Informationsmodell, das den Ist- und Soll-Zustand des Unternehmens-Reportings abbildet. Das Informationsmodell bildet die Grundlage für verschiedene Auswertungen, die im Rahmen der Ist- und Soll-Analyse durchzuführen

sind. Neben den Auswertungen ist die Klassifikation von Berichten sowie Kennzahlen ein wichtiges Analyse-Werkzeug.

Es hat sich als essentiell herausgestellt, die Modellierung softwaretechnisch zu unterstützen. Nur bei konsistenter Erhebung und Pflege der Modelle ist es möglich, eine widerspruchsfreie Spezifikation zu erarbeiten. Zukünftige Entwicklungen des Vorgehensmodells und der Modellierungsmethode müssen sich auf die Weiterverwendung der Modelle zur semi-automatischen Konfiguration von Data Warehouses und Reporting-Software konzentrieren, um einen quantifizierbaren Mehrwert zu generieren, der bislang nur qualitativ zu messen ist.

Literaturverzeichnis

- [Adam96] Adam, D.: Planung und Entscheidung. Modelle - Ziele - Methoden. 4. Aufl. Gabler: Wiesbaden, 1996.
- [BeDR03] Becker, J.; Dreiling, A.; Ribbert, M.: Using Management Objectives to specify Management Information Systems - A Contribution to MIS Success. In: Proceedings of the Wirtschaftsinformatik 2003. 2003, S. 313-323.
- [BJKM07] Becker, J.; Janiesch, C.; Knackstedt, R.; Müller-Wienbergen, F.; Seidel, S. (2007). H2 for Reporting - Analyse, Konzeption und kontinuierliches Metadatenmanagement von Management-Informationssystemen. Arbeitsberichte des Instituts für Wirtschaftsinformatik No. 115 (J. Becker, H. L. Grob, S. Klein, H. Kuchen, U. Müller-Funk and G. Vossen Eds.), Münster.
- [BJPS06] Becker, J.; Janiesch, C.; Pfeiffer, D.; Seidel, S.: Evolutionary Method Engineering - Towards a Method for the Analysis and Conception of Management Information Systems. In: Proceedings of the 12th Americas Conference on Information Systems (AMCIS 2006). Acapulco, 2006, S. 3922-3933.
- [BeKR05] Becker, J.; Kugeler, M.; Rosemann, M.: Prozessmanagement. Ein Leitfaden zur prozessorientierten Organisationsgestaltung. 5. Aufl. Springer: Berlin, Heidelberg, New York, 2005.
- [BeSc04] Becker, J.; Schütte, R.: Handelsinformationssysteme. 2. Aufl. Moderne Industrie: Landsberg/Lech, 2004.
- [Bloh70] Blohm, H.: Die Gestaltung des betrieblichen Berichtswesens als Problem der Leitungsorganisation. Herne, Berlin, 1970.

- [Blo80] Blohm, H.: Berichtssysteme. In: Handwörterbuch der Organisation. Grochla, Erwin: Stuttgart, 1980, S. 315-320.
- [Glad03] Gladen, W.: Kennzahlen- und Berichtssysteme. 2. Aufl. Wiesbaden, 2003.
- [GoRi98] Golfarelli, M.; Rizzi, S.: A Methodological Framework for Data Warehouse Design. In: Proceedings of the ACM First International Workshop on Data Warehousing and OLAP (DOLAP 98). Washington D.C., 1998, S. 3-9.
- [Holt03a] Holten, R.: Integration von Informationssystemen - Theorie und Anwendung im Supply Chain Management. Habilitationsschrift, Universität Münster, 2003.
- [Holt03b] Holten, R.: Specification of Management Views in Information Warehouse Projects. *Information Systems* 28(7), 2003, S. 709-751.
- [Horv03] Horváth, P.: Controlling. 9. Aufl. München, 2003.
- [KnSJ06] Knackstedt, R.; Seidel, S.; Janiesch, C.: Konfigurative Referenzmodellierung zur Fachkonzeption von Data-Warehouse-Systemen mit dem H2-Toolset. In: Proceedings of the DW2006 - Integration, Informationslogistik und Architektur. Friedrichshafen, 2006, S. 61-81.
- [Koch94] Koch, R.: Betriebliches Berichtswesen als Informations- und Steuerungsinstrument. Dissertation, Technische Hochschule Darmstadt, 1994.
- [Küpp01] Küpper, H.-U.: Controlling. 3. Aufl. Stuttgart, 2001.
- [LBMS02] List, B.; Bruckner, R. M.; Machaczek, K.; Schiefer, J.: A Comparison of Data Warehouse Development Methodologies - Case Study of the Process Warehouse. In: Proceedings of the 13th International Conference on Database and Expert Systems Applications (DEXA 2002). Aix-en-Provence, 2002, S. 203-215.
- [Meng95] Mengele, A.: Controlling, Konzeptionen. In: Küpper, H.-U.; Weber, J. (Hrsg.) *Grundbegriffe des Controlling*. Stuttgart, 1995, S. 59-62.
- [Rose00] Rosemann, M.: Using Reference Models in the ERP Lifecycle. *Australian Accounting Review* 10(22), 2000, S. 19-30.
- [Stra02] Strauch, B.: Entwicklung einer Methode für die Informationsbedarfsanalyse im Data Warehousing. Dissertation, Universität St. Gallen, 2002.
- [StWi02] Strauch, B.; Winter, R.: Vorgehensmodell für die Informationsbedarfsanalyse im Data Warehousing. In: von Maur, E.; Winter, R. (Hrsg.) *Vom Data Warehouse zum Corporate Knowledge Center*. Heidelberg, 2002, S. 359-378.
- [Wani02] Waniczek, M.: Berichtswesen optimieren. Frankfurt, Wien, 2002.

Designing Open Source Business Intelligence for Public Administrations

Björn Niehaves, Felix Müller-Wienbergen

European Research Center for Information Systems (ERCIS)

Universität Münster

48149 Münster

{bjoern.niehaves | felix.mueller-wienbergen}@ercis.uni-muenster.de

Abstract

Business intelligence approaches are standard in private organisations. In public administrations, however, only first steps towards such intelligent managing support are undertaken. Deficits become apparent especially in the move of implementing new public management approaches, such as new accounting systems or an output-oriented management. While public administrations are often obliged to implement such approaches in order to collect management-relevant data, there is little support regarding how to employ these new structures in terms of an intelligent business management. Here, balanced scorecards (BSC) can provide a valuable business intelligence instrument. Core problems that such an approach faces in the domain of public administrations are domain specific requirements, data collection problems in a heterogeneous systems environment, and financial limitations due to budgetary restrictions. Our design science-oriented case study on BSC implementation in public administrations introduces a solution which bases on an open source approach and addresses the problem of data collection by BSC and project management software integration. A case study analysis identifies generalisable issues which can potentially be applied to other situational and organisational contexts.

Summary of Arguments

- Paradigmatic changes in public administration management, for instance, regarding the accounting system (Neues Kommunales Finanzmanagement) or output-oriented management (Neues Steuerungsmodell), reveal severe *deficits in strategic thinking in public agencies*.
- *Balanced Scorecards* (BSC), if adapted to the specific context of public administrations, are a valuable instrument for implementing strategic thinking in these organisations.

- An *integration of project management and BSC software* can be a feasible solution to the central problem of data collection in a heterogeneous information systems environment, such as in public administrations.
- *Open Source Software (OSS)* is a practicable, low-cost, and flexible alternative to proprietary solutions for business intelligence in public administrations.

1 Introduction

There is a high demand for business intelligence solutions (concepts and software) in public administrations. New Public Management (NPM) and Electronic Government (eGovernment) are widely applied approaches to modernise public administrations into IT-based, networked, and market- and citizen-oriented organisations [Falc02; ScPr03; ScEi03; TrWi01]. In recent years, German public administrations have been obliged to introduce concepts as product-oriented budgeting (*Produktorientierter Haushalt*) or double-entry accounting (*Neues Kommunales Finanzmanagement*) which basically aim at providing data and improving the environment for an output-oriented management. Implementing these concepts is for most German public administrations, by law, obligatory (e.g., in Northrhine-Westfalia by 2009, in Lower Saxony by 2010). On the other hand, however, the beneficial perspective, how to use this data and how to actually implement and benefit from output-oriented strategic management [ScPr03], is underemphasised in public administrations with a strategy deficit being the consequence.

BSC have been established as a valuable instrument addressing major management problems in private and public organisations [KaNo96a; KaNo00; OIRW99]. An empirical study conducted in major US-enterprises [KaNo96] has shown, for instance, that there exist significant deficits in actually aligning the business strategy and business operations, that classical financial measures often run too short when it comes to strategic management decisions, or that controlling and reporting systems are often perceived as too complex but insufficient when it comes to ad hoc requests. These and other significant problems in management practice have lead to developing BSC as a strategy management and controlling instrument [Horv01]. Hence, BSC aims at balancing performance measurement between strategy and operations, taking into account various types of measures, e.g. qualitative and quantitative, and including different stakeholder perspectives, e.g. customer or employee perspectives [KaNo96]. BSC is the most known and applied performance measurement concept [GüGr02] while its IT support is seen as a major success

factor [Buyt01; GüGr02]. Hence, a BSC establishes a possible conceptual basis for business intelligence and management information systems [Buyt01; OIRW99].

Implementing BSC in public administrations is confronted with severe problems. While the BSC approach provides a valuable conceptual basis for business intelligence solutions in public administrations, major problems arise when it comes to an operative implementation:

- The need for domain specific adaptations is often underestimated [Busc04]: While BSC is originally derived from the private sector, characteristics of public organisations are habitually not fully appreciated [Alt04].
- A heterogeneous information systems environment regularly leads to problems in systems interoperability [BADF05] and, thus, in (automatically) collecting BSC-relevant data.
- Budgetary restrictions, as omnipresent as in the majority of (German) public administrations, necessitate low-cost solutions.
- Technological know-how deficits and a latent change resistance in the public sector domain [Fisc02; ScPr03] suggest a technologically evolutionary, rather than a revolutionary approach.

Hence, the research question we seek to address within this paper is *how to design an integrated open source BSC system in public administrations?* The line of argumentation addresses the following sub-questions:

- What are BSC principles and their domain specific adaptations and what basic functionalities are to be covered within a BSC implementation process? (Sec 2)
- Regarding a design science case study, how can a web technology and open source approach resolve major problems of BSC implementation in public administrations? (Sec 3)
- What is the value added by integrating BSC and project management software? (Sec 3)
- Which aspects of the design science case study can potentially be generalised and support further related business intelligence projects? (Sec 4)

Addressing this research objective, the methods chosen are that of conceptual and empirical research. We will hence provide theoretical-logical arguments as well as empirical evidence by (briefly) conducting and analysing a BSC implementation case study in a public organisation. We consider the paper to contribute to and to be part of design science research in information systems [Bola89; HMPS04; MaSm95; Simo81; WaWE92]. We will therefore provide a brief

summarising assessment of this research, complying with the guidelines for evaluating design science in IS research [cf. HMPS04], within the concluding section (Sec 5).

2 Balanced Scorecards in Public Administrations

BSC is a performance measurement instrument which aims at balancing strategy and operations. It has been developed as a response to the discovery that, for instance, there exist significant deficits in actually aligning the business strategy and business operations, that classical financial measures often run too short when it comes to strategic management decisions, or that controlling and reporting systems are often perceived as too complex but insufficient when it comes to ad hoc requests [Horv01; KaNo96]. BSC aims, as the name of the concept reflects, at maintaining a balance “between short- and long-term objectives, between financial and non-financial measures, between lagging and leading indicators, and between internal and external performance perspectives” [KaNo96a, p. viii].

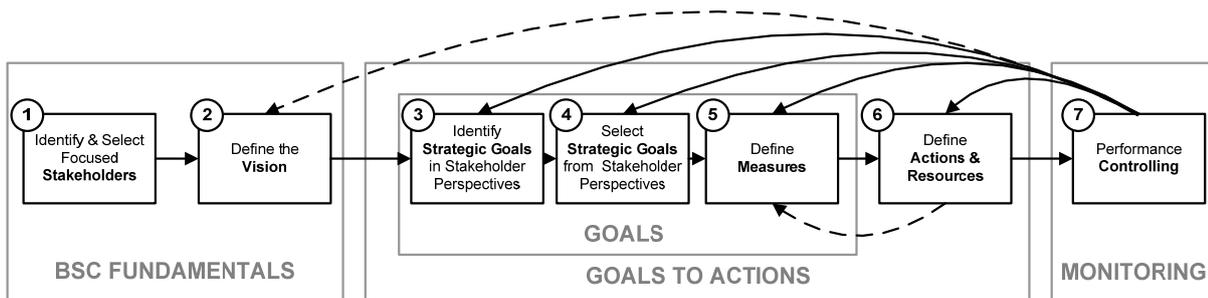


Figure 1: Schematic Balanced Scorecard Implementation Process

The BSC implementation process has several phases, each of them rendered by a specific task and concern. Literature provides us with a multitude of BSC implementation approaches [KaNo96], often featuring different granularity and practicality levels. Figure 1 outlines a schematic and archetypal BSC implementation procedure which takes into account the most common process features:

- 1) *Identify and Select Stakeholder Groups.* BSC seeks to balance between diverse stakeholder perspectives, including external and internal stakeholders. Starting the implementation process, it has to be decided upon which stakeholder perspectives ought to be taken into account. Financial perspective (‘How to deliver value to the shareholders?’), customer perspective (‘How to satisfy the customer needs?’), internal business perspective (‘Are we working effectively and efficiently?’), or innovation and learning perspective (‘What are emerging opportunities and challenges?’) are frequently taken into consideration. However, when it

comes to, for instance, BSC in public administrations, often the political or the citizens' perspective are considered as equally essential [Sche02]. Furthermore, it appears to be appropriate to differentiate between distinct 'customer' perspectives. As a consequence, five plus one perspectives are identified for a public organisation-wide BSC approach: Financial, employees, structures and processes (internal business perspective) in analogy to the original BSC concept. Additionally, the customer perspective has been differentiated into the citizens' and the businesses' perspectives. Furthermore, a political perspective was identified which varies from the other perspectives in the sense that the political perspective, as primacy of representative democracy, is 'superior' to the other perspectives (see also Point 3, Figure 2).

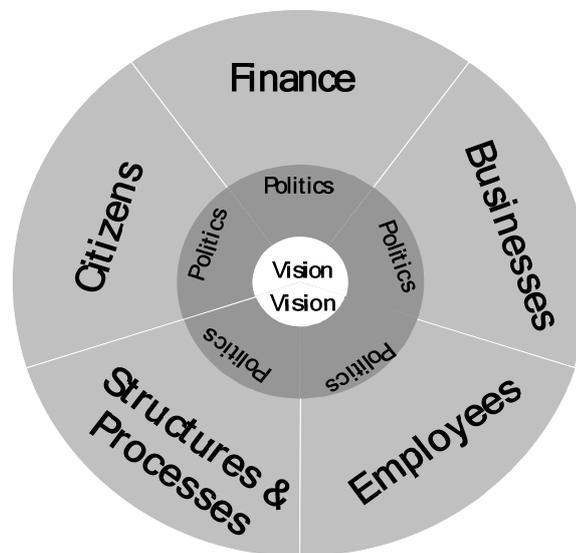


Figure 2: Balanced Scorecard Stakeholders

- 2) *Define the Vision.* The corporate vision is the first step towards policy within BSC: How does the organisation picture itself in a positive scenario in the long run? The vision functions as reference point for policy making, e.g. for defining strategic goals. Also with regard to the selected stakeholder groups, the goal here is to create a vision that is agreed on and widely accepted within the organisation. In many public administrations, the step of creating a common vision first opens the process of strategic thinking. At this point, often very divergent perspectives on the future development of a public administration become evident. Often two perspectives of the vision are differentiated from another: the internal perspective deals with the organisation, the public administration, itself while the external perspective is concerned with the public community (see again Figure 2).
- 3) *Identify Strategic Goals in Stakeholder Perspectives.* With regard to the corporate vision, policy making is taken one step further by identifying and discussing strategic goals that

ought to be pursued by the organisation. Here, the different stakeholder perspectives regularly frame the discussion of strategic goals, meaning: What are the strategic goals (abstract ‘demands’ at this point) stakeholder group X holds (implicitly and explicitly) with regard to the organisation? So-called goal landscapes for each stakeholder perspective can facilitate the discussion. Within public administrations, the defined perspectives (for instance, businesses) are still very heterogeneous. For instance, retail companies or agricultural companies might pursue very different goals regarding desirable infrastructures. Thus, it appears as very fruitful to conduct workshops with a selected group of representatives that defines such goal landscapes. As a second step, these goal landscape can be ‘filled’ with the help of specific workshops, interviews, questionnaires, document analyses etc. Here, identifying strategic goals (within stakeholder perspectives) is the first step to select certain goals.

- 4) *Select Strategic Goals from Stakeholder Perspective.* Based on the strategic goals identified within the stakeholder perspectives, certain goals have to be discussed, aligned, and selected to become part of the organisational policy. Often the goals identified are conflicting. However, some goal conflicts can be solved via discussion while other goals remain conflictory. In public administrations, this is a process that is regularly conducted by politics (democratic representative). In some cases, decisions can also be undertaken relying on direct democratic instruments.
- 5) *Define Measures.* BSC also features controlling characteristics. Thus, the definition of suitable measures for measuring if and how well particular goals have been achieved is the next step. As these measures are also sought to provide a motivational aspect, not only management representatives, but also employees responsible for achieving the goals are involved in defining them.
- 6) *Define Actions and Resources.* Defining measures and actions heavily depends on each another. Certain actions can efficiently be measured in only a certain way, some measures cannot be considered, if there is no data available. Therefore, a feedback loop is commonly recommendable. The group deciding upon actions and resources is often the same that defines certain goal measures. However, deciding upon what should be the ‘right’ things to do and what should be the ‘right’ resources for that is regularly an especially critical issue. In public administrations, one can regularly find evidence that these actions supporting BSC strategic goals often base on projects. As a consequence, project management and the definition of BSC-related actions are often strongly intertwined.

7) *Performance Controlling*. BSC aims at continuity. Therefore, a continuous performance controlling analyses if and how well the goals have been achieved according to the measures defined. Mostly in cases when problems occur, for instance, the measures have not been met, the resource limits have been exceeded, or the strategic goal has proven to be questionable, a problem analysis seeks to stimulate improvements. This often results in redefining goals, measures, actions, or resources. Besides these problem-driven improvements, also proactive steps can be taken, for instance, in terms of analysing and discussing on a regular basis if the strategic goals are still suitable or if environmental circumstances have changed. As a consequence of the strong interconnection between BSC and project management, performance controlling often heavily relies on project controlling.

The tasks to be performed within a particular phase of the BSC implementation are often interconnected with each other so that several feedback loops can be compulsory. Furthermore, the parties involved within the particular steps vary, however, stimulating and guiding a discussion among these parties is constantly the critical but often the least methodologically supported concern. An interconnection with project management can be a valuable way to save extra work, in the sense that project management is conducted anyways and thus provides relevant performance data to the BSC system.

3 Open Source BSC Software for Public Administration – A Design Science Case Study

Information Systems supporting the BSC implementation process have become an obligatory, in that valuable, part of almost every BSC project [Buyt01; GüGr02]. Systems of this kind allow for an efficient documentation and versioning of an enterprise BSC and foster its development process by providing graphical means illuminating a scorecard structure. Furthermore, they offer a technical infrastructure granting communication of a scorecard among all organisational entities of an enterprise. Moreover, providing means for real time monitoring of key performance indicators is one of the main advantages of a BSC software tool. Contemporary tools integrate data from operative systems or data warehouses for keeping objective compliance under surveillance and foster a real time exception handling and irregularity escalation.

Aside from the exigency to adapt the BSC concept to the specific context of public administration, the application of BSC software tools in this sector has to accomplish some domain specific requirements, as well. In the course of this section, we will pinpoint these context specific

needs and demonstrate how utilising web technologies, developing software on the basis and under the licence model of OSS, and integrating BSC and project management may help to meet these requisitions. Furthermore, the presentation of a case study is subject matter of this section. Here we demonstrate how former findings aided the successful conduction of a BSC implementation project for a medium-sized German public administration.

1) *Web technologies*

Applying web technologies for the development of a BSC software tool in the domain of public administration meets the prevailing situation of shrinking budgets in this sector.

First, most of the components necessary for setting up a web information system are free of charge. For instance, the Apache web server, being the most popular system of its kind, may be used without paying any licence fees. Second, as almost every public administration is hosting a web site the IT infrastructure necessary for operating a web information system is already in place. Therefore, the complexity of additional maintenance efforts is kept to a minimum. This aspect also holds good as the centralized architecture of web information systems limits administration and maintenance tasks to the web server side. As contemporary operation systems already provide web browsers, the installation of a web based software tool does not impose any further requirements on the client systems.

Extending the repository of every day software tools in public administrations on the basis of web technologies lowers user acceptance barriers. Especially in a surrounding characterised by latent change resistance as such in public administration [Fisc02; ScPr03], implementing a system which easily integrates with users' existing working environment fosters their acceptance. This holds good for web information systems. As users can handle a web based BSC software tool by using well-known concepts like navigation via back and forth navigation keys and setting bookmarks, convincing them of supporting the BSC idea and inducing commitment for the new strategic initiative becomes easier.

2) *Open Source Software*

Budgetary limitations in the public administration sector and the early status of the BSC adaptation process to this particular domain favours reverting to OSS and developing tools under Open Source licences. The term Open Source describes a software concept which comprises a particular licensing model, distribution model, and development model. In the context of OSS, several licensing models exist. Most of them accord to the Open Source Definition (OSD) which was developed by the Open Source Initiative (OSI)

[FeFi02]. Core issues are the right of unrestricted editing and adaptation, free distribution and use regarding the software as well as the source code [Reit04]. Every user (and developer) has the right to copy the software as often as needed. Furthermore, licenses allow to take the source code, to derive a new software product from the former, and to redistribute the new software using the same label [Schi02]. OSS development is a communicative process often characterised by a lack of central control [Koch03]. An OSS-community normally comprises several developers which regularly work independently from each other developing particular parts of the software. Moreover, an incremental and rapid publication of new releases is symptomatic for an OSS development process. Raymond stressed this fact by characterising open source development with the phrase “Release early, release often” [Raym99]. Thus, altered requirements or identified bugs are converted into new releases instantaneously. All in all the community-like OSS development process leads to the fact that several ideas as well as interests are considered during software development pursuing a best-of-breed approach. Open source code is generally more open to program testing and bug fixing. Another advantage of OSS is the fact that most developers are highly motivated due to their often intrinsic motivation as they often use their own software.

Against this background, the motivation for applying OSS in the area of public administration is manifold [BeRS04; Marq04]. The most obvious reason is related to the monetary dimension [DüWe04]. OSS may be employed free of charge and the internet offers an abundant source for gratuitous documentation and assistance for solving installation and maintenance issues. However, a sufficient degree of technical know-how within public administration is necessary to revert to OSS. Opposed to commercial software tools, OSS still requires kind of a do-it-yourself attitude for successfully harnessing its potential. In addition to the monetary advantages, the characteristics of its development process also militate in favour of applying OSS to meet the public administration particularities. Due to the recency of aligning BI concepts to the demands of the public administration sector, readjustment of developed IT solutions is in all probability. Aligning the BSC concept to the particular context of public administration will probably take course in a trail and error process of adaptations. Due to its incremental and rapid development process, the application of OSS supports flexible reaction to altered requirements [Fitz06]. As the development of BI tools looms large in the OS community [Klei06], it is in all likelihood that there will be a sufficient group of programmers promoting the flexible development of a BSC OSS tool.

3) *Integration of BSC and project management*

Integrating BSC and project management software tools makes for an efficient and effective way of deriving and automatically collecting BSC relevant measures. An effective provision of management relevant information necessitates a comprehensive business intelligence infrastructure comprising a data warehouse and ETL (extract, transform, load) interfaces to the data providing operative systems. However, the prevailing ‘backwardness’ of contemporary public administration’s IT infrastructure hinders the automatic surveillance of BSC goal compliance. Both public administration processes lacking a sufficient support by information systems and the application of proprietary stand-alone software solutions [BADF05] impede self-controlled provision of data for BSC measurement. In order to ease the measure maintenance in an environment like this the integration of ratios derived from project management provides a fruitful loophole out of this problematic situation. As actions defined to accomplish BSC objectives often show features characteristic for projects, ratios reflecting whether actions are conducted properly can be utilised for BSC measurement. In this regard supporting the execution of specified BSC actions via a project management software tool offers the ability to automatically obtain BSC relevant data and bypasses the lack of an extensive business intelligence infrastructure. However, information on whether a project is on time and budget do not entirely substitute the original BSC measures but provide a fruitful source for additional strategic monitoring.

The Integration of BSC and project management software tools fosters user acceptance and commitment to the BSC strategic management approach. Assigning projects to organisational objectives explains the contribution the participating employees make to the fulfilment of the overall vision. Every-day operative work becomes directly linked to strategic targets preventing the BSC concept from being misunderstood as an abstract concept out of touch with reality. On the whole, the integration approach serves for a tighter interlocking of operative and strategic management.

The former findings served as guidelines for a BSC software tool development project we accomplished for a German public administration in 2005. The environmental conditions for this project resemble the situation described in Section 1. The public administration faced the challenge of adopting the BSC concept to its own domain specifics to pave the way for output-oriented strategic management. A BSC software tool should support the entire implementation process as sketched in Section 2. However, due to budget limitations, utilising a commercial

tool for this purpose was not feasible and therefore applying OSS became the favoured solution. Furthermore, the existing information system environment consisted of a heterogeneous pool of autarkic applications hardly interconnected.



Figure 3: BASIS cockpit

A preceding evaluation revealed that there was no open source BSC solution so far. Therefore, it was decided to start a new open source development project named BASIS – BALanced Scorecard Based Information System. It was benefited from basing the project on existing open source solutions and reusing the framework of an available open source application. For the development of the web system the open source scripting language PHP¹ was chosen. Our development environment comprises of the open source database management system MySQL, the Apache web server, and the integrated development environment PHPeclipse.² Due to the employment of a database abstraction layer and sticking with the W3C HTML standard³ different database management systems and web servers are also supported by our application. In order to design business logic and user interface independently from each other, the PHP template engine smarty⁴ was utilised. The generic framework of the open source project management tool dotProject⁵ was adopted for realising a modular software architecture. Furthermore, dotProject provided some software artefacts serving as a starting point for the development of the project management module to be integrated with the BSC solution.

¹ See <http://www.php.net>.

² See <http://www.mysql.com>, <http://httpd.apache.org>, and <http://www.phpclipse.net>. PHPeclipse is based on the open source platform-independent software framework eclipse (<http://www.eclipse.org>).

³ See <http://www.w3.org/MarkUp>.

⁴ See <http://smarty.php.net>.

⁵ See <http://www.dotproject.net>.

BASIS provides a wide range of functions presented by a clearly structured user interface. As explicitly demanded by the public administration and consistent with the propositions stated above, the tool comprises a tight integration of BSC and project management functionality. The overall compliance with the defined BSC objectives is illustrated on a cockpit-screen serving as an entry point to the system (see Figure 3). Besides system and BSC administration functions, time recording, storage and maintenance of both contacts and events, the administration and monitoring of projects comprises a further corner stone of BASIS. The progress of tasks as well as the adherence to budget and time restrictions is supervised and evaluated by the system. On this basis performance indicators are derived pointing out possible negative effects on the attainment of BSC goals.

Both the specification of a BSC adhering the particularities of the public administration and the development of a conformable OSS BSC and project management tool took eight months in total. In this regard falling back on existing OS solutions and related assisting resources on the internet considerably saved time. At present the developed BSC and the supporting tool are successfully used in every-day business of the public administration. Although the developed web system does not act as an all-embracing BI solution it serves as a good starting point to gradually evolve IT infrastructure for backing public administration on its move to market- and output-oriented management.

4 Case Study Analysis and Results

According to Lee [Lee89] we will add a brief case study analysis in order to address the questions of repeatability and generalisability of the case study (results). Here, we will refer to four core questions in case study research:

1) What is the initial setting in the organisation (case study data) and in how far is it bound to specific situational and historical circumstances?

A core issue in the case study setting was the need for business intelligence systems resulting from severe societal, legal, and information-technological changes. Constraints for designing the desired system were mainly interoperability problems in a heterogeneous system environment (about 40 different information systems), budgetary restrictions, and a latent resistance to (technological) change. The described situation is not expected to alter significantly within a five year horizon; merely certain efforts to reduce the heterogeneity in the system landscape have been undertaken.

2) *Which (design) conclusions are drawn from the case study data and in how far are these conclusions bound to specific situational and historical circumstances?*

The main design decisions consist of pursuing the BSC concept, an open source approach, an integration of project management functionalities, and a web technology implementation. Here, especially the BSC concept can be understood as an initial approach to strategic management and business intelligence for public administrations [Alt04; Busc04; Sche02]. It is expected that, over time, other management functionalities can provide valuable extensions to this single-concept approach. Anyhow, still BSC is expected to be the major conceptual basis for business intelligence efforts in the organisation within a five year horizon. The open source approach, even if being considered as better-quality solution, will still have to practically prove its positive network effects in this particular setting.

3) *In how far do other settings (here: public administrations) show similar features and, thus, in how far is the case study setting generalisable?*

Especially medium-sized German public administrations share major features with the described case study setting. Not only that the services and processes that these organisations have to provide are existentially similar by law, they also have major problems in common, for instance, budgetary restrictions, heterogeneous systems environments [BADF05], or latent change resistance [Fisc02; ScPr03]. This being the overall trend, individual differences do evidently exist. However, especially their business intelligence maturity can be expected similar as, for instance, legal changes in accounting systems (*Neues Kommunales Finanzmanagement*) will affect most public administration within the same time horizon (depending on the state/Bundesland in most cases 2009 or 2010).

4) *Are the (design) conclusions made in the case study setting transferable to other organisational settings?*

The major design decisions in the case study setting consist of pursuing the BSC concept, an open source approach, an integration of project management functionalities, and a web technology implementation. Especially the integration of BSC and project management had major advantages in the particular case study setting (see again Sec 3). Firstly, project management software support was explicitly demanded by several employee(s)/(groups). Secondly, the given public administration had restricted resources available for the project so that a feasible and practicable alternative to a huge information systems integration (comprising, for instance, the population register, land register, (diverse) accounting systems, external data-

bases, finance monitoring system, council information system, etc.) was found. However, other BSC implementation projects chose dissimilar paths [Sche02], mainly against the background of different resource alimentations.

As a consequence, the case study setting can be – regarding its business intelligence maturity – considered as typical medium-sized German public administration. Major features affecting the design decisions will be apparent in most organisations of this kind within a five year time horizon. However, major differences in resource availability and financial support may affect design decisions, for instance, the open source approach or an integration of project management functionalities.

5 Conclusions and Future Research

There exists a high demand for strategic management and business intelligence solutions in public administrations. Here, BSC – adapted to the specific domain of public administrations – has proven to be a valuable approach enacting strategy and vision in daily operations. At this juncture, several key success factors were identified for the domain of public administrations:

- A strong involvement of employees and stakeholder representatives is crucial to the BSC project success. Defining the corporate vision (phase 2) is often the first step towards manifesting strategic thinking in public administrations. Very heterogeneous perspectives often become evident at this point of time which requires rich conceptual and motivational support. A conceptual combination of BSC-based business intelligence systems and project management can contribute to solving the data problem, avoid additional work, increase employee acceptance.
- OSS is in most cases low-cost and a valuable alternative to proprietary software products, also for business intelligence systems in public administrations.
- Applying web technologies for public administration information systems has proven to be a low-cost, but effective solution which is accompanied by a relatively high user acceptance also due to its recognition effect.

At this juncture, a brief paper self-assessment seeks to bring further clarity to presenting our research findings and follows Hevner et al.'s (2004) guidelines for design science research evaluation (see Table 1).

Guideline	Addressed in the paper
Epistemological Positioning	<ul style="list-style-type: none"> ▪ The epistemological position taken is that of linguistic interpretivism. Assuming that a real world' exists, the perceptions of it are influenced by the subject. The reason for such subjective perceptions of reality is assumed to be language differences, as languages not only provide representative means, but also form perceptions and constitute a differentiation instrument [in IS, see also Ortn04]. As a consequence, an aim is to create a language community relating to the issue of interest [BeNi07; KaLo73].
Addressing a relevant problem	<ul style="list-style-type: none"> ▪ The need for domain specific adaptations of BSC is often underestimated: While BSC is originally derived from the private sector, characteristics of public organisations are habitually not fully appreciated. ▪ A heterogeneous information systems environment regularly leads to problems in systems interoperability and, thus, in (automatically) collecting BSC-relevant data. ▪ Budgetary restrictions, as omnipresent as in the majority of (German) public administrations, necessitate low-cost solutions. ▪ Technological know-how deficits and a latent change resistance suggest a technologically evolutionary, rather than a revolutionary approach.
Making a research contribution	<ul style="list-style-type: none"> ▪ Design, case study implementation, and analysis of an integrated BSC and project management software which is based on an open source approach, employs web technologies and enacts a domain-specific BSC concept for public administrations
Choosing an adequate research method	<ul style="list-style-type: none"> ▪ Conceptual and theoretical-argumentative method as well as empirical case study implementation
Addressing the question of research rigour	<ul style="list-style-type: none"> ▪ It was sought to rigorously apply the conceptual and theoretical-argumentative method by remaining a high clarity of argumentation. Furthermore, empirical data collection within the case study was conducted rigorously against the background of the underlying linguistic interpretivist epistemology.
Designing an artefact	<ul style="list-style-type: none"> ▪ BASIS software as design artefact; an implementation of a BSC-based business intelligence system
Research evaluation	<ul style="list-style-type: none"> ▪ The first steps of research evaluation, implementing and evaluating the BASIS software in a medium sized German public administration in terms of a design science case study, yet indicated its feasibility. However, further evaluation is necessary.
Adequate communication of research	<ul style="list-style-type: none"> ▪ Clarity of argumentation was primary goal for research communication (see above) ▪ The business intelligence community on WI2007 is considered as adequate audience, potentially providing further feedback on special business intelligence issue, also from other domains than public administrations.

Table 1: Design Science Research Assessment

In order to provide evidence supporting the concept feasibility, a case study implementation was conducted. Following the case study implementation, further evaluation is necessary. For further future research, an analysis of OSS capability for other application areas in the field of eGovernment is to be conducted. Furthermore, OSS applications, such as the presented BASIS, should be made open to the public in order to facilitate a discussion of advantages, disadvantages, and application experiences. As the application of OSS looms large in the area of business intelligence at present [heis06; Klei06], further effort in integrating the BASIS application with existing OSS solutions is projected. Here, especially the connection with OSS data warehouses is in the focus of our endeavour to also foster the automatic provision of measures unre-

lated to project monitoring. Additionally, further management-oriented perspectives on OSS in PA are needed, for instance addressing Total-Cost-of-Ownership analyses.

References

[Alt04] *Alt, J.M.*: Balanced Government - Die Eignung der Balanced Scorecard als Organisationsentwicklungsprozess in der Öffentlichen Verwaltung. In: *Scherer, A.G.; Alt, J.M. (Eds.)*: Balanced Scorecard in Verwaltung und Non-Profit-Organisationen. Schäffer-Poeschel, Stuttgart 2004.

[BeRS04] *Becker-Pechau, P.; Roock, S.; Sauer, J.*: Open Source Software für die Softwareentwicklung. In: HMD - Praxis der Wirtschaftsinformatik 238 (2004), pp. 19-25.

[BADF05] *Becker, J.; Algermissen, L.; Delfmann, P.; Falk, T.; Niehaves, B.*: Identifikation von Best Practices durch Geschäftsprozessmodellierung in öffentlichen Verwaltungen. In: HMD - Praxis der Wirtschaftsinformatik 41 (2005), pp. 86-96.

[BeNi07] *Becker, J.; Niehaves, B.*: Epistemological Perspectives on IS Research – A Framework for Analysing and Systematising Epistemological Assumptions. In: Information Systems Journal 17 (2007).

[Bola89] *Boland, R.D.*: The Experience of System Design: A Hermeneutic of Organizational Action. In: Scandinavian Journal of Management 5 (1989) 2, pp. 87-104.

[Busc04] *Busch, V.*: Wettbewerbsbezogene Controllinginstrumente im Rahmen des New Public Management. Vahlen, München 2004.

[Buyt01] *Buytendijk, F.*: Balanced Scorecard Tools: Comparing Apples and Oranges. Gartner Group Research Note DF-12-8143 2001.

[DüWe04] *Dürr, C.; Weske, D.*: Einfluss von Open-Source-Software in kommerziellen Softwareprojekten. In: HMD - Praxis der Wirtschaftsinformatik 238 (2004), pp. 72-82.

[Falc02] *Falck, M.*: Business Process Management - As a Method of Governance. In: *Lenk, K.; Traunmüller, R. (Eds.)*: Proceedings of the First International Conference on Electronic Government (EGOV'02). Aix-en-Provence, France 2002, pp. 137-141.

[FeFi02] *Feller, J.; Fitzgerald, B.*: Understanding Open Source Software Development. London, U.K. et al. 2002.

- [Fisc02] *Fisch, R.*: Widerstände gegen Veränderungen in Behörden - Sozialpsychologische Perspektiven. In: *König, K. (Ed.): Deutsche Verwaltung an der Wende zum 21. Jahrhundert. Nomos, Baden-Baden 2002.*
- [Fitz06] *Fitzgerald, B.*: The Transformation of Open Source Software. In: *MIS Quarterly* 30 (2006) 3, pp. 587-898.
- [GüGr02] *Günther, T.; Grüning, M.*: Performance Measurement-System im praktischen Einsatz. In: *Controlling* 14 (2002) 1, pp. 5-13.
- [heis06] *heise online*: Acht Millionen Dollar für Open Source Business Intelligence. <http://www.heise.de/newsticker/meldung/75528>, Date of Access: 2006-10-08.
- [HMPS04] *Hevner, A.R.; March, T.S.; Park, J.; Sudha, R.*: Design Science in Information Systems Research. In: *MIS Quarterly* 28 (2004) 1, pp. 75-105.
- [Horv01] *Horváth, P.*: Balanced Scorecard umsetzen. Schäffer-Poeschel, Stuttgart 2001.
- [KaLo73] *Kamlah, W.; Lorenzen, P.*: Logical Propaedeutic. Pre-School of Reasonable Discourse. 1st ed., University Press of America, Lanham/MD 1973.
- [KaNo96a] *Kaplan, R.S.; Norton, D.P.*: The balanced scorecard: translating strategy into action. Harvard Business School Press, Boston/MA 1996.
- [KaNo96] *Kaplan, R.S.; Norton, D.P.*: Using the BSC as a Strategic Management System. In: *Harvard Business Review* 74 (1996) 1, pp. 75-85.
- [KaNo00] *Kaplan, R.S.; Norton, D.P.*: The balanced scorecard: measures that drive performance. Harvard Business School Press, Boston/MA 2000.
- [Klei06] *Kleijn, A.*: Business Intelligence mit Open Source. <http://www.heise.de/open/artikel/73725>, Date of Access: 2006-10-08.
- [Koch03] *Koch, S.*: Das Open-Source-Entwicklungsmodell: Grundprinzipien, Fragen und Erfahrungen. In: *HMD - Praxis der Wirtschaftsinformatik* 231 (2003), pp. 55-62.
- [Lee89] *Lee, A.S.*: A Scientific Methodology for MIS Case Studies. In: *MIS Quarterly* 13 (1989) 1, pp. 33-52.
- [MaSm95] *March, T.S.; Smith, G.*: Design and Natural Science Research on Information Technology. In: *Decision Support Systems* 15 (1995) 4, pp. 251-266.

- [Marq04] *Marquardt, C.*: Open-Source-Software-Strategie der deutschen Verwaltung. In: HMD - Praxis der Wirtschaftsinformatik 238 (2004), pp. 58-61.
- [OIRW99] *Olve, N.G.; Roy, J.; Wetter, M.*: Performance drivers: a practical guide to using the balanced scorecard. Wiley, Chichester 1999.
- [Ortn04] *Ortner, E.*: Anthropozentrik und Sprachbasierung in der (Wirtschafts-) Informatik. In: *Hammwöhner, R.; Rittberger, M.; Semar, W. (Eds.): Wissen in Aktion - Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft. Festschrift für Rainer Kuhlen.* Hochschulverband für Informationswissenschaft 2004, pp. 141-152.
- [Raym99] *Raymond, E.S.*: The Cathedral and the Bazaar. O'Reilly, Sebastopol, CA 1999.
- [Reit04] *Reiter, B.E.*: Wandel der IT: Mehr als 20 Jahre freie Software. In: HMD - Praxis der Wirtschaftsinformatik 238 (2004), pp. 83-91.
- [ScPr03] *Schedler, K.; Proeller, I.*: New Public Management. UTB, Bern et al. 2003.
- [Sche02] *Scherer, A.G.*: Besonderheiten der strategischen Steuerung in Öffentlichen Institutionen und der Beitrag der Balanced Scorecard. In: *Scherer, A.G.; Alt, J.M. (Eds.): Balanced Scorecard in Verwaltung und Non-Profit-Organisationen.* Schäffer-Poeschel, Stuttgart 2002, pp. 3-26.
- [ScEi03] *Scherlis, W.L.; Eisenberg, J.*: IT research, innovation, and e-government. In: *Communications of the ACM* 46 (2003) 1, pp. 67-68.
- [Schi02] *Schiffner, T.*: Open Source Software. Freie Software im deutschen Urheber- und Vertragsrecht. München, Germany et al. 2002.
- [Simo81] *Simon, H.A.*: The Sciences of the Artificial. MIT Press, Cambridge/MA 1981.
- [TrWi01] *Traunmüller, R.; Wimmer, M.*: Directions in E-Government: Processes, Portals, Knowledge. In: *Proceedings of the DEXA International Workshop "On the Way to Electronic Government"*. Los Alamitos, CA 2001, pp. 313-317.
- [WaWE92] *Walls, J.; Widmeyer, G.; El Sawy, O.*: Building an Information System Design Theory for Vigilant EIS. In: *Information Systems Research* 3 (1992) 1, pp. 36-59.

Move-to-the-User? Eine Analyse der verlagernden Wirkung von Business Intelligence im Controlling

Michael Samtleben und Thomas Hess

Institut für Wirtschaftsinformatik und Neue Medien
Ludwig-Maximilians-Universität München
D-80539 München
{samtleben, thess}@bwl.uni-muenchen.de

Abstract

Mittels Fallstudien wird in diesem Beitrag die verlagernde Wirkung von Business Intelligence auf eine unternehmensinterne Aufgabe, das Controlling, untersucht. Im Vordergrund der Betrachtung steht dabei die Allokationsveränderung controllingspezifischer Aufgaben zwischen verschiedenen Aufgabenträgern im Unternehmen. Für ein theorieorientiertes Vorgehen bieten Kompatibilitätskalküle dazu ein passendes Analysewerkzeug. Im Ergebnis zeigt sich, dass der Einsatz von Business Intelligence zu einer Delegation controllingspezifischer Aufgaben von der Controlling- und der IT-Abteilung in die Fachabteilung führt und insbesondere das interne Berichtswesen verlagert wird.

Keywords:

Business Intelligence, technological imperative, Controlling, Delegation, Fallstudien

1 Einführung

Die betriebliche Nutzung von Business Intelligence Technologien (BI-Technologien) wurde in den letzten Jahren intensiv untersucht. Dabei standen die Identifikation neuer Anwendungsfelder und Leistungsverbesserungen der Technologie im Mittelpunkt. In der Praxis zeigt sich nun, dass diese Technologien auch die Organisation eines Unternehmens verändern können. Ein paar Unternehmen haben diese Idee schon aufgegriffen und zum Beispiel eigene BI-Abteilungen gegründet, andere stehen vor diesbezüglichen Entscheidungen. Allerdings ist noch unklar, welche Wirkung BI auf Aufbau- und Ablauforganisation haben wird. In der nachfolgenden Analyse greifen wir diese Fragestellung auf. Wir konzentrieren uns dabei auf das Controlling als eines der traditionell wichtigsten BI-Anwendungsfelder.

Analysen zur Wirkung neuer Technologien auf das Unternehmen haben mittlerweile schon Tradition. Im Sinne von Markus und Robey [MaRo88] folgen wir mit der Analyse der Wirkung der BI-Nutzung auf die betriebliche Organisation dem „technological imperative“. In diesem Feld sind wiederum zwei Gruppen von Forschungsarbeiten zu unterscheiden: einerseits zu den Auswirkungen der Informationstechnologie auf Unternehmensgrenzen und zwischenbetriebliche Beziehungen [PiRW03], andererseits zum Einfluß der Informationstechnologien auf die interne Organisationsstruktur [WyMa96]. Arbeiten zu den Auswirkungen der IT auf die Verteilung von Controlling-Aufgaben wären im zweiten Feld zu suchen, finden sich aber nur ansatzweise. So postulieren Gehra et al. BI ein hohes Potenzial für den Einsatz in verschiedenen Fachbereichen sowie zahlreiche nicht in der Controllingabteilung angesiedelte User [GeGH05]. Die Wirkung von Business Intelligence auf die Allokation controllingspezifischer Aufgaben wird daher Gegenstand dieses Beitrags sein. Im zweiten Abschnitt geben wir einen Überblick über die wichtigsten Grundlagen. Unter Rückgriff auf die Organisationstheorie entwickeln wir in Kapitel drei Thesen zur Wirkung von BI auf die Aufgabenverteilung im Controlling. Im vierten Kapitel folgen dann die Fallstudien, deren Ergebnisse im fünften Kapitel verglichen und interpretiert werden. Das letzte Kapitel rundet diesen Beitrag mit einer Zusammenfassung und einem Ausblick auf den weiteren Forschungsbedarf ab.

2 Grundlagen

2.1 Business Intelligence und seine Anwendung im Controlling

Business Intelligence umfasst eine Vielzahl unterschiedlicher Werkzeuge bzw. Systeme, die das Management bei Planung, Steuerung und Kontrolle unterstützen. Chamoni und Gluchowski zählen hierzu alle Werkzeuge und Anwendungen mit entscheidungsunterstützendem Charakter, die zum besseren Verständnis der Mechanismen relevanter betriebswirtschaftlicher Wirkungsketten beitragen [ChGl04, S. 120]. Wird dieser Zuordnung gefolgt, so können mittels eines zweidimensionalen Ordnungsrahmens drei gängige Typen zur Einordnung von BI-Technologien vorgenommen werden (vgl. Abb. 1). Während die horizontale Achse den Fokus auf die Sichtweise des derzeitigen Diskussionsschwerpunkts richtet, werden auf der vertikalen Achse die jeweiligen Phasen der analytischen Datenverarbeitung, von der Datenbereitstellung bis zu deren Auswertung, aufgetragen.

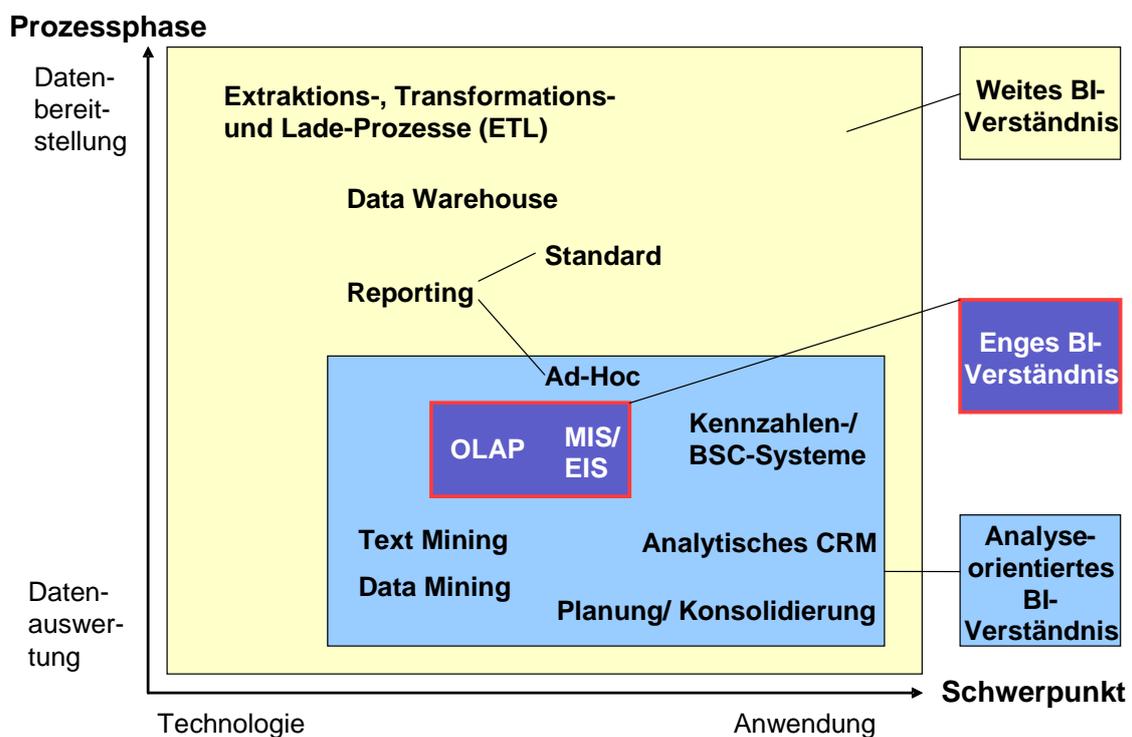


Abb. 1: Übersicht über die Einordnung von BI-Technologien [Gluc01, S. 7]

Nach dem engen BI-Verständnis werden lediglich einige wenige Kernapplikationen als BI verstanden, die die Entscheidungsfindung unmittelbar unterstützen. Diese sind zum einen das Online Analytical Processing (OLAP), zum anderen das Managementinformationssystem (MIS) bzw. das Executive Information System (EIS). Das analyseorientierte BI-Verständnis umfasst dagegen bereits mehrere Technologien. Eingeschlossen werden sämtliche Anwendungen bei denen der Entscheider unmittelbar am System arbeitet und somit einen direkten Zugriff auf die

Benutzeroberfläche mit interaktiven Funktionen besitzt. Es geht dabei um die zielgerichtete Analyse von vorhandenem Datenmaterial [ChGl04, S. 120]. Neben OLAP und MIS/EIS fallen darunter auch Text- und Data Mining, Ad-hoc-Reporting, IT-gestützte Balanced Scorecards (BSC), analytisches Customer Relationship Management (CRM), sowie Systeme zur Unterstützung von Planung und Konsolidierung [Hess03]. Unter das weite BI-Verständnis fallen alle direkten und indirekten zur Entscheidungsunterstützung eingesetzten Anwendungssysteme [KeMU04, S. 3-4]. Damit sind alle Systemkomponenten gemeint, „die operatives Datenmaterial zur Informations- und letztlich zur Wissensgenerierung aufbereiten und speichern sowie Auswertungs- und Präsentationsfunktionalitäten anbieten“ [ChGl04, S. 120]. Innerhalb dieses Beitrags wird das BI-Verständnis im engeren Sinn zu Grunde gelegt.

Ein wichtiges Anwendungsfeld von BI ist traditionell das Controlling. Vereinfachend verstehen wir unter Controlling nachfolgend alle Aufgaben, die den Management-Prozess der Zielfindung, Planung und Steuerung gestalten und planen. Dabei ist die Rolle des Einzelnen, für Strategie-, Ergebnis-, Finanz- und Prozesstransparenz zu sorgen und die Wirtschaftlichkeit zu steigern [Webe04, S. 20]. Die wichtigsten Aufgaben des Controllings sind das interne Berichtswesen, Kontrolle/Abweichungsanalysen, Budgetierung, operative und strategische Planung, internes Rechnungswesen und Investitionsrechnung [SaHe06]. Business Intelligence dient bei verschiedenen Aufgaben der Informationsbeschaffung und Aufbereitung sowie der Bewertung und Präsentation der Information. Konkrete Einsatzfelder der Technologie sind einerseits analytische Auswertungen wie im Marketingcontrolling oder Investor Relations Management, aber auch klassische Anwendungsfelder wie das interne Berichtswesen oder die Abweichungsanalyse [Hess03]. Der Einsatz von Business Intelligence wird dabei insbesondere durch die wachsende Sammlung im Unternehmen anfallender Daten, wie Markt-, Rechnungswesen- oder Kundendaten, begünstigt [GeGH05].

2.2 Ökonomische Kalküle zur Aufgabenallokation

Zur Analyse der technologieinduzierten Veränderungen bietet sich das Kompatibilitätskalkül nach Laux und Liermann an [LaLi05, S. 238ff], wobei wir uns auf die Anforderungskompatibilität konzentrieren und die von der IT weniger tangierte Frage der Anreizkompatibilität ausklammern. Die Anforderungskompatibilität lässt sich wiederum in Informations- und die Kalkülkompatibilität zerlegen, auf die wir nachfolgend eingehen. Als informationskompatibel gelten Aufgabenträger, wenn sie die Fähigkeit besitzen, die Menge der Handlungsalternativen ausreichend zu berücksichtigen und die Informationen entsprechend wahrzunehmen. Kalkülkompatibilität baut auf der Informationskompatibilität auf und stellt Anforderungen an die Fähigkei-

ten eines Individuums bezüglich Prognosebildung, Ergebniseinschätzung und Entscheidungsmodell.

Sowohl bei der Informations- als auch bei der Kalkülkompatibilität sind für die Eignung eines Akteurs im Wesentlichen die Aufgabenmerkmale Strukturiertheit, Variabilität und Umfang einer Aufgabe ausschlaggebend – und diese können sich wiederum durch IT ändern [LaLi05, S. 242-251; PiDF05, S. 226-227]. Strukturiertheit, auch Komplexität oder Analysierbarkeit einer Aufgabe genannt, bezeichnet ganz allgemein, inwieweit die Umsetzung genauestens präzisiert bzw. die Lösungsschritte weitestgehend vorgegeben sind. Mit Variabilität oder Dynamik wird das Ausmaß beschrieben, inwiefern während der Umsetzung einer Aufgabe einzelne Entscheidungsgrößen verändert werden müssen. Als Umfang einer Aufgabe wird das erwartete Volumen bzw. die Menge der zu treffenden Entscheidungen bezeichnet. Neben diesen Kriterien können noch Aufgabeninterdependenzen im Unternehmen auftreten, die sich insbesondere dann ergeben, wenn Controllingaufgaben auf mehrere Aufgabenträger verteilt werden. Interdependenzen mit anderen Aufgaben, die durch die Abhängigkeiten zwischen Aufgaben verschiedener Mitarbeiter entstehen, wirken sich erhöhend auf die Gesamtkomplexität der Koordination aus. Je höher diese wird, desto genauer muss eine Abstimmung zwischen den Mitarbeitern erfolgen, was Transaktionskosten in Form von Informations- und Entscheidungskosten verursacht [Jost01, S.330].

Ein konkretes Anwendungsbeispiel dieser Kriterien auf eine controllingspezifische Aufgabe, wenn auch ohne IT-Bezug, stellt die Diskussion um den Wandel der Budgetierung von der klassischen bis hin zu Beyond Budgeting dar [Webe04, S. 374-380]. Dabei eignen sich abhängig von der Komplexität und Dynamik der Aufgabenanforderung unterschiedliche Formen der Budgetierung. Die klassische Budgetierung vermag gering dynamische und hoch komplexe Kontexte zu bewerkstelligen, während dies bei Beyond Budgeting genau umgekehrt der Fall ist.

3 Theoretische Analyse

Zu untersuchen ist nun, inwiefern Business Intelligence Einfluss auf die einzelnen Aufgabenmerkmale nimmt und inwiefern es zu einer Aufgabendelegation vor dem Hintergrund veränderter Transaktionskosten kommen kann. Aufgrund höherer Anforderungskompatibilität und reduzierter Motivations- oder Koordinationskosten ist eine Delegation oder Redelelegation einer controllingspezifischen Aufgabe womöglich effizienter. Für die fortführenden Betrachtungen muss an dieser Stelle festgehalten werden, dass Controlling primär eine entscheidungsunterstützende Funktion für das Management darstellt. Für die Untersuchung beziehen wir daher in die arbeits-

teilige Erbringung der Controllingleistung den Controller, die IT-Abteilung und Fachabteilungen ein. Der Controller sorgt dabei im Unternehmen für die nötige Managementunterstützung mittels transparenter und aufbereiteter Informationen und besitzt diesbezüglich notwendiges Methodenwissen. Die IT-Abteilung ist im Rahmen des Controlling für die Informationstransparenz verantwortlich und Fachabteilungen besitzen die fachliche Verantwortung über ihre operativen Tätigkeiten. Die Unternehmensführung wurde dabei nicht näher betrachtet, da der Fokus auf die Managementunterstützung gelegt wird. Abbildung zwei stellt schematisch die Beziehung zwischen den befragten Aufgabenträgern und die Wirkung von Business Intelligence auf die Allokation controllingspezifischer Aufgaben dar.

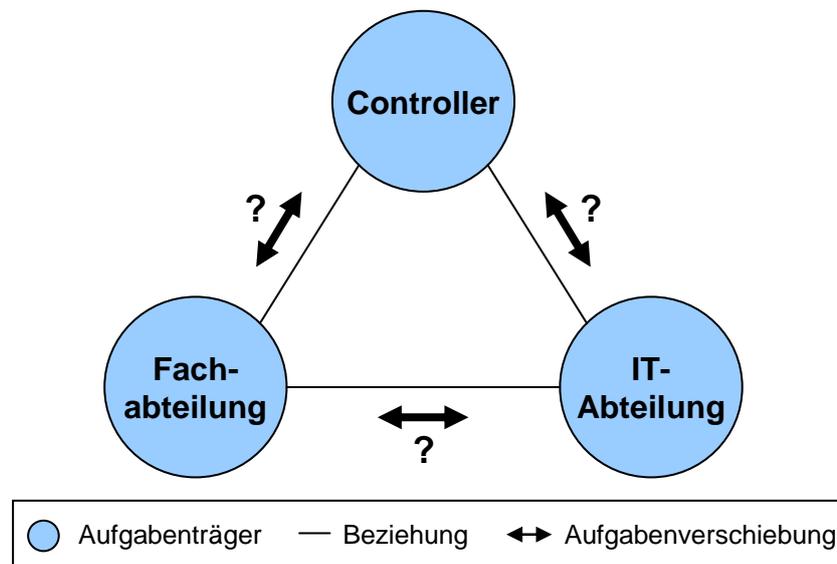


Abb. 2: Aufgabenträgermodell

Business Intelligence bietet die Möglichkeit, Daten zu analysieren und nach vorgegebenen Kriterien zu strukturieren. Des Weiteren erlauben BI-Systeme eine bessere Informationsversorgung höherer Hierarchieebenen und zeichnen sich durch Visualisierung und rollengerechte Aufbereitung sowie Verwaltung der Informationen aus [Geor00]. Aus diesem Grund wird sich die Anforderungskompatibilität verändern, indem durch eine Verringerung des Information Overload beim Aufgabenträger die Informationskompatibilität steigt oder einfach zu bedienende Analysewerkzeuge zu einer höheren Kalkülkompatibilität führen.

Werden nun die Aufgabenmerkmale des theoretischen Kalküls für die Formulierung der Thesen herangezogen, können interessante Aussagen zur Wirkung von Business Intelligence getroffen werden. Zur Strukturiertheit lässt sich beispielsweise die Informationsbeschaffung anführen, die von ihrem Bedarf und ihrer Beschaffungsquelle abhängig ist. Wenn diese Variable verändert wird, indem sich durch BI die Beschaffung der Information erleichtert, dann wird die Informationsbeschaffung und damit das Gesamtausmaß der Aufgabe strukturierter. Dies ist beispiels-

kurzfristig ermöglicht, individuelle Reports zu gestalten und in diesen zu slicen und dicen. BI erhöht demzufolge die Strukturiertheit einer Aufgabe. Bei der Variabilität gilt es, die Menge der Handlungsalternativen zu betrachten. Wenn ständig neue Handlungsalternativen aufgrund von andauernden Ausnahmefällen gesucht werden müssen, damit eine Aufgabe gut erfüllt wird, dann ist die Variabilität der Handlungsalternativen hoch. BI dämmt die Ausnahmefälle aufgrund der besseren Informationsbasis und der damit verbundenen schnelleren Reaktionsfähigkeit ein. Somit ist die Variabilität der Menge der Handlungsalternativen als auch die Variabilität der Gesamtaufgabe geringer. Kurzum: BI verringert die Variabilität einer Aufgabe. Wie viele und welche Entscheidungsvariablen der Entscheidungsträger bei seiner Entscheidung berücksichtigt, ist nicht objektiv vorgegeben und bleibt ihm überlassen. Um eine gute Entscheidung bei einer umfangreichen dispositiven Aufgabe zu treffen, ist tendenziell viel Zeit zur Beschaffung und Verarbeitung der Information nötig. Komplexe, umfangreiche Aufgaben sind deshalb häufig unstrukturiert. Je größer die Variabilität der einzelnen Primärdeterminanten, umso größer ist tendenziell auch der Umfang der Aufgabe. Eine subjektive Entscheidungsvariable des Umfangs könnte bspw. die Zeit sein. Wenn diese Variable verändert wird, indem BI die Beschaffung und Verarbeitung der Information beschleunigt, dann wird auch der Umfang einer gesamten Aufgabe verringert. Daraus folgt: BI verringert den Umfang einer Aufgabe.

Insbesondere die systematische Analyse und Aufbereitung großer Datenbestände durch BI führt dazu, dass sich durch die Erhöhung der Strukturiertheit und die Verringerung der Variabilität das Ausmaß der Kompatibilität zwischen Aufgabenträger und Aufgabe verändert. So kann unter Umständen auf Methodenwissen verzichtet werden und Aufgabenträger, die sich nicht im Kern mit Controlling beschäftigen, sind in der Lage, die Aufgabe umzusetzen. In diesem Fall ändern sich entscheidende Aufgabenmerkmale, so dass es je nach Ausprägung zu einer Umverteilung der Aufgaben bzw. zur Bildung funktionaler Stabsstellen kommen kann. Verbesserungen bei Visualisierung und Personalisierung sowie einfacher bedienbare Anwendungssysteme ermöglichen den Fachabteilungen, controllingspezifische Informationen selber aufzubereiten. Aufgrund steigender Akzeptanz für Informationssysteme und einfachere Handhabung lautet daher unsere erste These: Business Intelligence verschiebt Controllingaufgaben von der IT-Abteilung in Richtung Fachabteilung. Dabei stehen insbesondere systemseitige Auswertungen für das Reporting im Vordergrund, die früher noch auf Endlospapier bei der IT-Abteilung in Auftrag gegeben wurden.

Neben der Verbesserung auf technischer Seite sind in aktuellen Business Intelligence Systemen methodische Fähigkeiten hinterlegt. So konnte früher eine Delegation von Analysetätigkeiten in die Fachbereiche nur erfolgen, wenn Mitarbeiter statistische Verfahren beherrschten. Heutzuta-

ge reduzieren die Systeme die methodische Komplexität, so dass eine Delegation eher infrage kommt. Darunter fällt neben der Sammlung notwendiger Informationen auch die Analyse und Bereitstellung. Die zweite These lautet demnach: Business Intelligence verschiebt Aufgaben von der Controllingabteilung in Richtung Fachabteilung und IT-Abteilung.

Wenn BI die Abstimmung zwischen den Aufgabenträgern durch eine gemeinsame Informationsbasis verbessert, so werden Datenbrüche vermieden. Der geringere Abstimmungsbedarf zeigt sich einerseits in verringerten Informations- und Entscheidungskosten, und andererseits in einer reduzierten Gesamtkomplexität der Koordination. Auch hier besteht ein Anreiz, die Arbeit durch bessere Abstimmungsmöglichkeiten zu erleichtern. Folglich verringert BI Interdependenzen mit anderen Aufgaben. Darin spiegeln sich auch schon erste Auswirkungen von BI auf die Transaktionskosten wider. So kann der Einsatz integrierter Standardsoftware zu einer Reduzierung der Koordinationskosten führen, da beispielsweise die Kosten für die Informationsbeschaffung durch den Einsatz eines Data Warehouses oder die Abstimmungskosten durch Groupware sowie verkürzte Kommunikationswege sinken [SaSH06].

Beide Thesen formulieren einen starken Trend der Delegation in die Fachabteilungen, was auf die Studie von Samtleben und Hess zurückzuführen ist. Darin zeigt sich, dass die Betrachtung der Fachabteilung als Aufgabenträger im Controlling nicht vernachlässigt werden darf [Sa-He06]. Während in der gängigen Controllingliteratur hauptsächlich das Zusammenspiel von Controller und Manager betrachtet wird, wurde hier bewusst die Aufgabenverschiebung in Richtung Fachabteilung fokussiert.

4 Darstellung der Fallstudien

Zur Überprüfung der zwei Thesen haben wir die Einführung von BI-Technologien in zwei Fallstudien analysiert. Für die Erhebung der Fallstudien wurden insgesamt zehn Interviews mit Personen aus dem Controlling, der IT-Abteilung sowie aus Fachabteilungen geführt und um Informationen aus Dokumenten ergänzt. Da die Aufgabenträger aus unterschiedlichen Organisationsbereichen innerhalb eines Unternehmens stammen, wurden rollenspezifische Interviewleitfäden entwickelt. Eine Vergleichbarkeit der beiden Fallstudien wird erreicht, indem Aufgabenträger mit gleicher bzw. ähnlicher Funktion in beiden Unternehmen befragt wurden.

4.1 Fallstudie: Einführung eines OLAP-Systems bei Knorr Bremse

4.1.1 Hintergrund

Die Knorr Bremse AG (KB AG), mit Hauptsitz in München, ist Hersteller von Bremssystemen für Schienen- und Nutzfahrzeuge und kam im Jahr 2005 auf einen Umsatz von 2,74 Mrd. Euro. Weitere Produktfelder, in denen sich die Knorr Bremse etabliert hat, sind On-Board-Systeme und Bahnsteigtürsysteme für Schienenfahrzeuge, sowie Drehschwingungsdämpfer für Nutzfahrzeuge. Die Steuerung des Konzerns erfolgt regional. Dadurch sollen die jeweiligen Markt- und Kundenanforderungen in den drei großen Regionen Europa, Amerika sowie Asien/Australien bestmöglich berücksichtigt werden. Derzeit sind bei der KB AG ca. 12.000 Mitarbeiter beschäftigt. Durch die fortschreitende Internationalisierung des Konzerns war Ende der 90er Jahre sowohl eine neue Organisationsstruktur als auch deren technologische Unterstützung notwendig. Zur Unterstützung der regionalen Steuerung führte die KB AG im Jahr 1999 das OLAP-System Hyperion Essbase ein, was inzwischen aus konzernstrategischen Gründen durch SAP BW ersetzt wurde. Die nachfolgende Darstellung bezieht sich auf die Einführung von Hyperion.

4.1.2 Veränderung von Aufgabenmerkmalen

Die konkreten Aufgabenmerkmale haben sich durch das OLAP-System in unterschiedlicher Art und Weise verändert. Durch die bessere und vor allem einzige Informationsbasis des OLAP-Systems, als auch durch die kontinuierliche Verbesserung der technologischen Prozesse innerhalb des OLAP-Systems, lassen sich Aufgaben Schritt für Schritt besser bearbeiten. Grund hierfür ist die schnellere und detailliertere Betrachtung der Zusammenhänge, als auch die Vereinfachung und Beschleunigung der Aufgabenschritte durch ihre technologische Unterstützung. Durch die transparentere Zerlegung der Problemstellung und Überführung in vorgegebene Lösungsschritte hat sich bei allen befragten Aufgabenträgern die Strukturiertheit durch das OLAP-System erhöht. Was den Grad der Variabilität bei der Aufgabenbewältigung betrifft, so müssen die Ausnahmefälle genauer betrachtet werden. Jeder einzelne Aufgabenträger greift auf eine detailliertere Informationsbasis zu und hat damit eine schnellere Reaktionsfähigkeit, wodurch die Anzahl der Ausnahmefälle sinkt. Der Menge an unvorhersehbaren Ereignissen bei der Aufgabenbearbeitung kann demnach durch die Datenbasis entgegengewirkt werden. Allerdings ist dies bei der KB AG nur eingeschränkt der Fall, da gerade bei Ausnahmefällen stets eine Verifizierung der Daten, Zusammenhänge und Ergebnisse nötig ist. Ein ausschließliches „sich auf das System verlassen“ gibt es nicht. Die Variabilität einer Aufgabe sinkt demnach bei der KB AG, aber nicht in dem Ausmaß wie eingangs vermutet. Das zeitliche Ausmaß der Aufgabenbewältigung

gung wird sowohl in der Controlling- als auch in der Fach- und IT-Abteilung reduziert. Daraus folgt eindeutig das Ergebnis, dass der Umfang der Aufgaben bei allen Befragten durch das OLAP-System gesunken ist. Ob das System neben diesen ersten Merkmalsänderungen auch transaktionskostensenkend wirkt, wird anhand der folgenden Ausführungen geklärt. Informationen sind seit der OLAP-Einführung leichter zugänglich, was sich hinsichtlich der Informationskosten bemerkbar macht. Durch OLAP werden zumindest diejenigen Informationskosten reduziert, die durch die automatische Verarbeitung und Aufbereitung der Daten eingespart werden. Bezüglich Entscheidungen bzw. Entscheidungskosten werden Veränderungen differenziert betrachtet. Es wird davon ausgegangen, dass theoretisch selbständiges Entscheiden durch die bessere Informationsbasis des Systems verstärkt möglich ist, praktisch aber dennoch Hierarchien eingehalten werden müssen. Auch wenn in der KB AG eine flache Hierarchiestruktur vorherrscht, welche die Delegation von Verantwortung auf kleine organisatorische Einheiten mit großer Sachkompetenz zulässt, so werden dennoch Rangfolgen befolgt. Entscheidungskosten treten auf, wenn Mitarbeiter zusammen Entscheidungen treffen. Ob diese gesenkt werden, lässt sich aufgrund der Ergebnisse nicht klar sagen und ist eher zu verneinen. Was sich aber deutlich zeigte, ist die Verbesserung der Entscheidungsqualität. Schlechte oder gar falsche Entscheidungen kommen sowohl in geringerem Ausmaß, als auch in geringerer Anzahl vor. Kosten durch nicht konforme Entscheidungen lassen sich demzufolge eindeutig reduzieren. Aufgrund des Ergebnisses, dass OLAP weder eine Kontrolle noch eine Leistungsbewertung über den jeweiligen Aufgabenbeitrag der einzelnen Aufgabenträger zulässt, und so eine Messbarkeit der Leistung nicht gegeben ist, lassen sich weder Kontroll- und Überwachungskosten, noch Kosten der Leistungsbewertung reduzieren.

4.1.3 Organisatorische Veränderung

Durch die OLAP-Einführung werden in den Fachabteilungen vermehrt ganze Aufgabenprozesse selbständig bearbeitet, da mehr Transparenz für einen größeren Teil von Mitarbeitern möglich ist. Des Weiteren hat sich durch die OLAP-Einführung eine neue Abteilung im Bereich Finance und Controlling gebildet, welche rein technologisch für OLAP zuständig war. Aufgrund des ständigen Wachstums und der inzwischen weit über die ursprünglichen Controllingaufgaben hinauswachsenden Systemlandschaften wurde diese Abteilung aber vor ca. einhalb Jahren in die zentrale IT-Abteilung als Shared Services für Business Intelligence eingegliedert. Aufgaben dieser Abteilung liegen in der konzernweiten Verantwortung für BI, in der Umsetzung von globalen betriebswirtschaftlichen Berichtsanforderungen, in der Entwicklung,

dem Betrieb und der Weiterentwicklung des Konzernberichtswesens, sowie in der Sicherstellung der Datenintegrität.

4.2 Fallstudie: Einführung eines Managementinformationssystems bei der Deutsche Bahn

4.2.1 Hintergrund

Die Deutsche Bahn AG (DB AG), mit Hauptsitz in Berlin, ist mit einem Umsatz von 25 Mrd. Euro im Jahr 2005 der führende Mobilitäts- und Logistikdienstleister Deutschlands. Die ca. 216.000 Mitarbeiter sind an den jeweiligen Standorten ihrer Büros, auf dem 35.000 km langen Schienennetz sowie auf den 5.700 Bahnhöfen Deutschlands tätig und wirken so am Transport von Personen, Gütern und Sendungen mit. Die Systemlandschaft der DB AG ist überwiegend inhomogen und die Anzahl der Schnittstellen ist damit sehr hoch. Innerhalb des Konzerns gibt es unterschiedlichste MIS, welche größtenteils nicht konzernweit, sondern nur in einzelnen Bereichen zum Einsatz kommen. Diese Fallstudie konzentriert sich auf das MIS von Hyperion mit der Bezeichnung „MIS FaSt“ (Managementinformationssystem Fachdienst Steuerung). Das MIS FaSt wurde im Jahr 1997 ursprünglich in den Bereichen DB Konzern, DB Cargo, DB Reise und Touristik sowie DB Traktion, welcher noch zum Jahresende 1997 zu DB Regio eingegliedert wurde, eingeführt. Aufgrund mehrerer Umorganisationen ist MIS FaSt derzeit bei DB Regio, DB Railion und DB Fernverkehr im Einsatz.

4.2.2 Veränderungen von Aufgabenmerkmalen

Durch MIS FaSt teilweise verändert sind Daten in gewünschtem und benötigtem Umfang schneller und besser abrufbar, wodurch die Controller eine höhere Informationsversorgung besitzen, mittels dessen sie ihre Aufgaben schrittweise effizienter erledigen können. Durch die verbesserte Abgrenzung der Problemstellung hat sich durch MIS FaSt die Strukturiertheit der Aufgaben erhöht. Unvorhersehbare Ereignisse lassen sich mittels MIS FaSt dennoch nicht einfacher bewältigen. Aufgrund der begrenzten Aktualität der Daten wird keine schnelle Reaktion auf die Veränderlichkeit der Aufgaben festgestellt. Damit ergibt sich keine Senkung der Variabilität, da die Unsicherheit nur mit zeitlicher Verzögerung reduziert wird. Was den Arbeitsaufwand der Aufgaben betrifft, so muss dieser differenziert betrachtet werden. Bei Controllern und Fachabteilungen ist aufgrund der MIS FaSt-Einführung ein verminderter Umfang an Controllingaufgaben festzustellen. Aufgrund der schnelleren Informationsabfrage durch MIS FaSt reduziert sich der notwendige Zeitaufwand zur Bearbeitung der Aufgaben. Die Anzahl der zu berücksichtigenden Entscheidungsvariablen wird dahingehend reduziert, dass die entstehenden

zeitlichen Kapazitäten auf das eigentliche Problem gelenkt werden. Diese Konzentration führt zu einem verminderten Arbeitsumfang. Bei der IT-Abteilung war allerdings das Gegenteil zu beobachten. Aufgrund der Systemvielfalt in der DB AG muss DB Systems verstärkt auf die Datenintegration und Kompatibilität zu den Schnittstellen der einzelnen operativen Systeme achten. Dennoch soll hier der Fokus auf die Benutzung und nicht die Bereitstellung von MIS FaSt gelegt werden, so dass aufgabenträgerübergreifend in begrenztem Ausmaß von einem verminderten Umfang gesprochen werden kann. Nach der Betrachtung der konkreten Aufgabenmerkmale wird nun auf die Transaktionskosten eingegangen. Dabei ist festzustellen, dass die Informationskosten eindeutig gesenkt werden, da vor allem eine Reduktion der Suchkosten erfolgte. Der Informationsabruf „per Knopfdruck“ erleichtert und beschleunigt die Bearbeitung der Controllingaufgaben enorm. Auch die Übermittlung, Verarbeitung oder Anweisung für Teilaufgaben fällt geringer aus, da sich jeder Aufgabenträger aufgrund der gemeinsamen Informationsbasis selbständig informieren kann. Dagegen werden Entscheidungskosten durch das MIS FaSt nicht direkt tangiert. Generell werden durch MIS Entscheidungsprozesse mittels der verbesserten Informationsbereitstellung beschleunigt, was wiederum individuelle Entscheidungen fördert und erleichtert. Allerdings hat dies bei der DB AG keine Kompetenzerweiterung zur Folge. Auch wenn eine Entscheidungsdezentralisation in der DB AG unterstützt wird, werden durch die Komplexität und Fachlichkeit des Konzerns nur vermindert Entscheidungen allein gefällt. Kosten, die durch die Abstimmung in Gremien oder Gruppen auftreten, werden durch das MIS FaSt damit nicht reduziert. Ob sich zumindest die Qualität der Entscheidungen verbessern konnte, geht aus den Ergebnissen nicht klar hervor. Generell wird von einem qualifizierteren Entscheidungsniveau durch eine gemeinsame Informationsbasis ausgegangen. Dennoch führen die Befragten dies nicht auf das MIS FaSt zurück. Kosten durch nicht konforme Entscheidungen konnten anhand der Ergebnisse damit nicht signifikant reduziert werden. Das MIS FaSt unterstützt technologisch gesehen weder eine Kontrolle noch eine Leistungsbewertung, womit die Aufgabenträger des Controllings kontrolliert oder bewertet werden. Aussagen bezüglich Kontroll- und Überwachungskosten bzw. Kosten der Leistungsbewertung werden damit nicht getätigt. Alles in allem war durch die Einführung des MIS FaSt nur eine Senkung der Koordinationskosten und somit letztendlich eine geringe Reduzierung der Transaktionskosten festzustellen.

4.2.3 Organisatorische Veränderung

Die MIS FaSt-Einführung hat in gewisser Weise auch eine organisatorische Änderung mit sich gebracht. Aus Sicht von DB Systems werden nun Controlling-Aufgaben zusätzlich von Fachab-

teilungen bearbeitet. Die Zentralen der Geschäftsbereiche werden so hinsichtlich Controlling und Monitoring gestärkt. Die Folge ist eine zunehmende Dezentralisierung von Entscheidungsbefugnissen mit dem Ziel, die operative Schlagkraft und wirtschaftliche Flexibilität der Regionen durch zusätzliche Reporting-Aktivitäten auf den einzelnen Ebenen zu erhöhen. Die einfache Bedienung erlaubt es jedem unkompliziert Auswertungen durchzuführen. Darüber hinaus wurde bei DB Systems eine Organisationseinheit eingerichtet, welche unmittelbar für Business Intelligence zuständig ist. Diese Organisationseinheit ist allerdings nicht neu und hat sich eher durch eine interne Umstrukturierung gefestigt. Business Intelligence wirkte dabei aber zumindest unterstützend, da Methodenwissen gebündelt wurde.

5 Interpretation und Einordnung der Fallstudien

Organisatorisch betrachtet ist in beiden Fällen eine Aufgabendelegation zu beobachten. Zur Managementunterstützung wird festgehalten, dass diese in beiden Konzernen nach dem gleichen Schema abläuft, auch wenn die Aufgaben unterschiedlich sind. Bei den Aufgabenmerkmalen lässt sich festhalten, dass Business Intelligence messbaren Einfluss auf diese besitzt. Die KB AG kann aufgrund der Aktualität der Daten besser auf unvorhersehbare Ereignisse reagieren. Dennoch bleibt bei Ausnahmefällen eine zusätzliche Verifizierung unerlässlich. Aufgrund der begrenzten Aktualität der Daten bestätigen sich diese Rückschlüsse bei der DB AG nicht, was den Unterschied bei der Variabilität von Aufgaben erklärt. Die Reduktion des Arbeitsumfanges ist bei der KB AG eindeutiger ausgefallen. Dies liegt an der Systemvielfalt in der DB AG, wodurch ein höherer Aufwand seitens der IT-Abteilung vorzuweisen ist. Bezüglich der Transaktionskosten sind in beiden Konzernen nur bestimmte Arten gesunken. Die Informationskosten reduzieren sich sowohl bei der KB AG als auch bei der DB AG, da Business Intelligence die Informationssuche und die Abstimmung zwischen den Aufgaben aufgrund der besseren Informationsbasis erleichtert. Kosten durch nicht konforme Entscheidungen werden aufgrund der besseren Entscheidungsgrundlage nur im Fall der KB AG reduziert. Die DB AG geht von einem qualifizierteren Entscheidungsniveau aus, führt allerdings bessere Entscheidungen nicht direkt auf Business Intelligence zurück. Grund hierfür ist mangelnde Messung und Quantifizierung [Walt99, S. 449]. Bezüglich der Motivationskosten kann nur insofern eine Aussage getroffen werden, dass in beiden Unternehmen diese als unabhängig von der Business Intelligence Einführung gesehen werden und daher kein Einfluss beobachtet werden kann.

Nach diesem synoptischen Vergleich stellen wir nun die beiden Fälle unseren in Abschnitt 3 aufgestellten Thesen gegenüber. Sieht man von kleineren Unterschieden ab ergibt sich ein glei-

ches Bild. In beiden Fällen konnte sowohl die Anforderungs- als auch die Anreizkompatibilität verbessert werden. Dies ist einerseits auf die gestiegene Qualifikation der Aufgabenträger und andererseits auf das gesunkene Arbeitsleid durch Business Intelligence zurückzuführen. Die ersten beiden Thesen, die Aufgabenverschiebung von der IT-Abteilung in die Fachabteilung, sowie von der Controllingabteilung in die Fachabteilung lassen sich damit bestätigen. Lediglich eine Verschiebung von der Controllingabteilung zur IT-Abteilung konnte nicht eindeutig nachgewiesen werden. Bezüglich der Transaktionskosten kann nur begrenzt eine Aussage getroffen werden. Generell lässt sich eine Reduzierung einzelner Transaktionskosten feststellen, deren Anzahl bei der KB AG stärker ausfällt. Ein klarer Trend zeichnete sich allerdings nur bei der KB AG ab. Abbildung 4 zeigt die beobachteten Veränderungen der Allokation durch Business Intelligence.

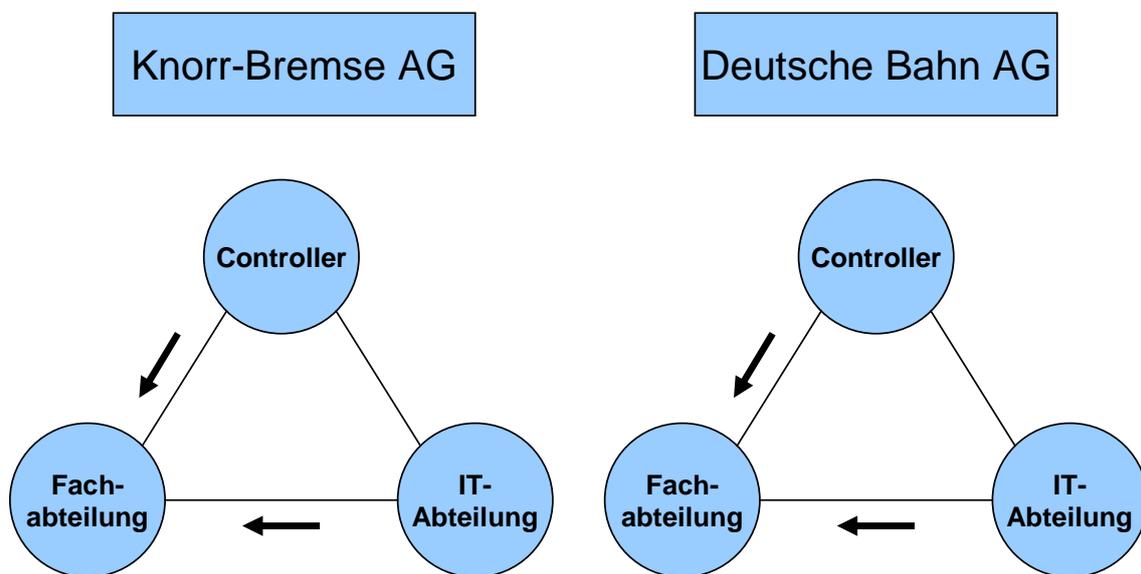


Abb. 3: Wirkung von Business Intelligence auf die Aufgabenverteilung

Zur Bewertung der Güte der hier dargelegten empirischen qualitativen Befunde sind drei Aspekte anzusprechen. Generell werden qualitative Verfahren daran gemessen, inwieweit sie die Kriterien Objektivität, Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit) erfüllen [ScHE05]. Offene Fragestellungen sowie die Interpretation der Antworten hängen zwar von der jeweiligen Untersuchungsperson ab. Dennoch ist davon auszugehen, dass bei der Durchführung eine Objektivität besteht, da keine Beeinflussung durch Interaktion stattgefunden hat und die in dieser Analyse dargestellten Aussagen von anderen überprüfbar sind. Reliabilität ist gesichert, da die Resultate der beiden Fallstudien gut übereinstimmen. Qualitative Fragebögen stellen ein gutes Mittel dar, um Thesen zu überprüfen, wenn sie in ausreichender Menge durchgeführt werden. Inwiefern die Daten in den vorliegenden Fallstudien valide sind, ist aufgrund der gerin-

gen Anzahl der Befragten in den einzelnen Abteilungen fraglich, aber anzunehmen [ScBr05, S. 45; Yin03, S. 33-39].

6 Zusammenfassung und Ausblick

Alles in allem haben die Untersuchungen gezeigt, dass Business Intelligence den Aufwand für die durchzuführenden Controllingaufgaben reduziert und den Ablauf verkürzt. Dies führt bei den Mitarbeitern in den Fachabteilungen dazu, vermehrt Controllingaktivitäten selber wahrzunehmen („Selbstcontrolling“). Dem Einfluss von Business Intelligence kann insgesamt keine Reduzierung des Controllingbedarfs beigemessen werden, wohl aber eine Veränderung ihres Aufgabenspektrums, was sich unmittelbar auch auf das der übrigen Aufgabenträger des Controllings auswirkt. Eine Verschiebung der Controllingaufgaben von der Controlling- und der IT-Abteilung in Richtung Fachabteilung ist in den Einzelfallstudien festzustellen. Dabei hat sich auch gezeigt, dass insbesondere das interne Berichtswesen in einem höheren Ausmaß von Fachabteilungen erbracht wird. Des Weiteren finden Analysen nun vermehrt in den Fachbereichen statt, da diese einerseits durch die einfache Bedienbarkeit der Systeme dazu befähigt werden und andererseits direkt vor Ort Controllingaufgaben ausüben können. Die zunehmende Delegation hat demnach auch den Controllingbereich erfasst. Unsere Ergebnisse passen damit ins Bild des Wandels des Controllers hin zum Berater [WeDP01].

Neben dem in der controllingrelevanten Literatur anzufindenden Wandel des Controllers stützen die beiden Fallstudien, wenn auch mit Einschränkungen, die Auffassung eines „technological imperative“ im Controlling. Um diese Aussage aber noch zu spezifizieren, bedarf es weiterer Forschungsarbeiten. Dafür erscheint die Betrachtung des Aufgabenträgermodells bestehend aus den drei Akteuren zu restriktiv, so dass eine um die Unternehmensführung erweiterte Untersuchung sinnvoll erscheint. Dies sollte auch in einer höheren Anzahl von Fallstudien überprüft werden, damit eine ausreichende Validierung gewährleistet werden kann.

Literaturverzeichnis

- [ChGI04] Chamoni, Peter; Gluchowski, Peter: Integrationstrends bei Business-Intelligence-Systemen - Empirische Untersuchung auf Basis des BI Maturity Model. In: Wirtschaftsinformatik 46 (2004) 2, S. 119-128.

- [GeGH05] Gehra, Bernhard; Gentsch, Peter; Hess, Thomas: Business Intelligence for the Masses - Datenaufbereitung und Datenanalyse für den Controller im Wandel. In: Zeitschrift für Controlling und Management 49 (2005) 3, S. 236-242.
- [Geor00] Georges, Patrick M.: The Management Cockpit - the human interface for management software: reviewing 50 user sites over 10 years of experience. In: Wirtschaftsinformatik 42 (2000) 2, S. 131-136.
- [Gluc01] Gluchowski, Peter: Business Intelligence - Konzepte, Technologien und Einsatzbereiche. In: Hildebrand, Knut (Hrsg.): HMD - Praxis der Wirtschaftsinformatik. Heidelberg 2001, S. 5-15.
- [Hess03] Hess, Thomas (Hrsg.): Anwendungssysteme im Controlling. In: Zeitschrift für Controlling und Management, 47 (2003) Sonderheft 2.
- [Jost01] Jost, Peter-Jürgen: Innerbetriebliche Koordination. In: Jost, Peter-Jürgen (Hrsg.): Der Transaktionskostenansatz in der Betriebswirtschaftslehre. Schäffer-Poeschel, Stuttgart 2001, S. 301-326.
- [KeMU04] Kemper, Hans-Georg; Mehanna, Walid; Unger, Carsten: Business Intelligence - Grundlagen und praktische Anwendungen : eine Einführung in die IT-basierte Managementunterstützung. Vieweg, Wiesbaden 2004.
- [LaLi05] Laux, Helmut; Liermann, Felix: Grundlagen der Organisation - die Steuerung von Entscheidungen als Grundproblem der Betriebswirtschaftslehre. 6. Auflage, Springer, Berlin 2005.
- [MaRo88] Markus, M. Lynne; Robey, Daniel: Information Technology and Organizational Change: Causal Structure in Theory and Research. In: Management Science 34 (1988) 5, S. 583-598.
- [PiDF05] Picot, Arnold; Dietl, Helmut; Franck, Egon: Organisation: eine ökonomische Perspektive. 4. Auflage, Schäffer-Poeschel, Stuttgart 2005.
- [PiRW03] Picot, Arnold; Reichwald, Ralf; Wigand, Rolf: Die grenzenlose Unternehmung: Information, Organisation und Management. 5. Auflage, Gabler, Wiesbaden 2003.

- [SaHe06] Samtleben, Michael; Hess, Thomas: Aufgabenverteilung und Nutzung der Informationstechnologie im Controlling - Ergebnisse einer empirischen Studie. In: Controlling 18 (2006) (im Druck).
- [SaSH06] Samtleben, Michael; Stadlbauer, Florian; Hess, Thomas: Anwendungssystemintegration im Controlling - aktueller Stand und wichtige Trends. In: Zeitschrift für Controlling und Management 50 (2006) 2, S. 86-93.
- [ScBr05] Schäffer, Utz; Brettel, Tanja: Ein Plädoyer für Fallstudien. In: Zeitschrift für Controlling und Management 49 (2005) 1, S. 43-46.
- [ScHE05] Schnell, Rainer; Hill, Paul B.; Esser, Elke: Methoden der empirischen Sozialforschung. 7. Auflage, Oldenbourg, München 2005.
- [Walt99] Walterscheid, Heinz: Systembewertung und Projektmanagement bei analytischen Informationssystemen. In: Chamoni, Peter; Gluchowski, Peter (Hrsg.): Analytische Informationssysteme: Data Warehouse, On-Line Analytical Processing, Data Mining. Berlin 1999, S. 427-451.
- [Webe04] Weber, Jürgen: Einführung in das Controlling. 10. Auflage, Schäffer Poeschel, Stuttgart 2004.
- [WeDP01] Weber, Jürgen; David, Ulrich; Prenzler, Carsten: Controller Excellence - strategische Neuausrichtung der Controller. WHU-Otto-Beisheim-Hochschule Lehrstuhl für Betriebswirtschaftslehre, Vallendar 2001.
- [WyMa96] Wyner, George M.; Malone, Thomas W.: Cowboys or Commanders: Does Information Technology Lead to Decentralization? International Conference on Information Systems (1996). Cleveland, S. 63-79.
- [Yin03] Yin, Robert K.: Case Study Research: Design and Methods. 3. Auflage, Sage, Thousand Oaks, California 2003.

Einführung in den Track

Business Process Engineering

Prof. Dr. Dimitris Karagiannis

Universität Wien

Prof. Dr. Markus Nüttgens

Universität Hamburg

Dr. Peter Küng

Credit Suisse, Zürich

Die Gestaltungspotenziale von Wertschöpfungsnetzwerken sind ein kritischer Erfolgsfaktor zur Steigerung der Wettbewerbsfähigkeit von Organisationen. Mit dem Einsatz prozessorientierter Methoden und Werkzeugen können die resultierenden Szenarien und Zielkonflikte frühzeitig abgebildet und weiterentwickelt werden. Geschäftsprozessmodelle fungieren hierbei als Bindeglied zwischen Strategie und IT-gestützter Implementierung.

Ziel des Tracks ist es, wissenschaftlich abgesicherte Konzepte und Fallstudien aus praxisnahen und anwendungsbezogenen Forschungsarbeiten unter ökonomischen, organisatorischen und IT-technischen Aspekten zu thematisieren.

Programmkomitee:

Prof. Dr. Rainer Alt, Universität Leipzig

Dr. Jan vom Brocke, Universität Münster

Prof. Dr. Jörg Desel, KU Eichstätt

Prof. Dr. Otto Krickl, Universität Graz

Prof. Dr. Frank Rump, FH Oldenburg/Ostfriesland/Wilhelmshaven

Oliver Thomas, Universität Saarland/DFKI

Prof. Dr. Mathias Weske, HPI Potsdam

Bernd Oestereich, oose Innovative Informatik GmbH, Hamburg

Supporting Inter-Business Collaboration via Contract

Negotiation and Enactment

Peter Rittgen

University College of Borås
501 90 Borås, Sweden
peter.rittgen@hb.se

Abstract

The increasing complexity of products and services encourages more and more companies to form collaborative networks. As these companies are independent organizations there is often an issue of governance. We suggest a possible architecture for such a business network that proposes a contract as the principal means of coordination and describes how such a contract can be designed and enacted.

1 Introduction

Today we can witness two seemingly opposed trends in the cooperation between businesses: On the one hand companies are forced to concentrate on their core competencies and to outsource all activities that lie outside the core. On the other hand customers demand that a supplier covers an increasing range of products and services. They want to buy a complete solution from only one supplier instead of buying bits and pieces from many. This latter point seems to suggest an increased amount of “insourcing”. The solution to both is that companies have to engage in closer cooperations, each concentrating on its area of expertise, but jointly offering a complete suite of related products and services that are well matched (one face to the customer). But this scenario represents an enormous challenge both in terms of organization and regarding the information system support.

Companies that want to engage in a closer cooperation, e.g. a value network, a virtual enterprise or the like, bring into this cooperation not only their different organizational cultures but also

different, often incompatible, information systems. A successful cooperation therefore requires the alignment or integration of both the business processes and the information systems to a certain degree. In some industries, such as the automotive industry, this can go as far as the customer forcing the suppliers to introduce the ERP system of the customer's choice (e.g. SAP). But on the whole it is more common that the organizations involved will strive for some kind of mutual adaptation of their business processes and information systems. In a very simple case this could be the introduction of a file transfer accompanied by suitable import and export functionalities and some organizational measures for providing and handling the new data. In more advanced cases it will imply substantial reorganization of business processes and changes to existing information systems and/or introduction of new ones.

Our goal is to support the set-up and operation of a business network. The first phase consists mainly of the design of a contract that can be used to coordinate the behavior of network actors. The design process is cooperative, i.e. the actors negotiate this contract among themselves. Such negotiations can be either bilateral or multilateral but both types will contribute to creating one common contract that is binding for all parties involved. We also call this negotiation process a co-design process. Negotiation is a social process that can be supported by a negotiation support system. This eliminates the need for partners to meet face to face and contributes to a flexible set-up of the business network. It implies that the lead-times for setting up the network are relatively short and replacing members that have left and adding new ones can be done with a minimum of effort. These are crucial issues for a business network.

The second phase, operation, consists of enacting the behavior specified in the contract. Here the business logic concerning the coordination of actors is incorporated into the communication network. In this this we phase "translates" from the business network to the communication network by managing the respective message exchange via the technical network and a coordination server. We have used this approach to improve the governance of an existent network that consisted of three partners: the headquarters of a retail chain in the home textile and home decoration industry, the shops of this chain and a third-party logistics provider. Although this is a minimal case of a business network it nevertheless provides fundamental insights into the workings of such networks.

The contribution of our work consists in outlining a method for governing business networks by negotiating and enacting contracts that take the form of business process models. It is based on existing methods for contract negotiation, business process modeling and workflow manage-

ment. Apart from combining these methods we also enrich them, e.g. by extending the concept of a trade contract to cover process aspects.

The remaining chapters are structured as follows. We first address the issue of coordination in organizational networks in general and business networks in particular, which leads us to the identification of a suitable class of contracts, namely behavior-based contracts. The next sections study the negotiation process and a language for formulating behavioral contracts. We then proceed by describing the enactment of the formalized contract based on a communication network and a coordination server. We conclude this paper by summarizing the major findings and presenting an outlook on future research.

2 Coordination in a Business Network

In a business network organizations strive for the provision of complex products and services by coordinating their activities in an “intelligent” way. This implies that the coordination effort is much higher than in a conventional supply chain. In the latter an individual company can focus on managing the relation to a few immediate major suppliers for creating a product or service. In a business network this is not enough but coordination is also required among the suppliers. Theoretically we move from a tree structure to a graph topology which implies that we have to hit a new balance between market and hierarchical coordination. The general problem behind this is quite old and several theories have been advanced to explain the use of a particular form of coordination, most notably Agency Theory [AlchianDemsetz'72; JensenMeckling'76; Ross'73; Wilson'68] and Transaction Cost Economics [Coase'37; KleinCrawfordAlchian'78; Williamson'75, '81, '85]. Based on these theories the internal and external coordination costs can be determined [GurbaxaniWhang'91]. High external costs favour centralization, high internal costs promote decentralization. It is typically assumed that organizations in a supply chain choose their organizational structure and network of trading partners in such a way that the sum of both costs is minimized. There has also been some debate on the impact of information technology (IT). Early work by Malone et al. [MaloneYatesBenjamin'87] suggested that IT will lower transaction costs and therefore, *ceteris paribus*, lead to an increase in market coordination. Later work posited that organizations will “move to the middle”, i.e. to “more outsourcing, but from a reduced set of stable partnerships” [ClemonsReddiRow'93] if non-contractible issues (e.g. quality and trust) play an important role. Empirical evidence [HollandLockett'97] shows

that companies often operate in a “mixed mode” blending aspects from both markets and hierarchies.

But the majority of these studies was performed in the context of conventional supply chains. In the face of a network topology the balance between hierarchical and market coordination needs to be readjusted: In the absence of a central coordination unit we typically use the contract as an instrument for coordination. Agency Theory suggests two principal forms of contracts, behavior-based contracts and outcome-based contracts. Between an employer and an employee, for example, a contract with a fixed annual salary would be behavior-based as such a contract demands that the agent performs to the best of his capabilities. An outcome-based contract would specify a remuneration that depends on the results that the agent has achieved (e.g. a commission). If the costs for monitoring agent behavior are high, an outcome-based contract is often superior. This is because an unobserved agent is assumed to shirk (i.e. underperform) knowing that he has no consequences to fear. This problem is called moral hazard. An outcome-based contract can be seen as a special case of a behavior-based contract where delivering the outcome is considered to be the only observable behavior of the agent. In addition to that, the costs for monitoring agent behavior have become marginal in many cases due to the omnipresence of information technology. These arguments apply also to the context of business networks. We will therefore focus our investigation on behavior-based contracts.

3 Negotiation as a Social Process

We define negotiation as the process whereby a group of two or more individuals tries to reach an agreement on the performance of future actions. The individuals are human beings that might act on behalf of organizations or on their own behalf. For the purpose of this paper we focus on electronic negotiations, i.e. negotiations that are supported by information and communication systems. They can be divided into three different types: bargaining, auction and agent negotiation [KöhneSchoopStaskiewicz'05]. Auctions are very common, especially in electronic commerce. They assume that the traded products or services can be described in detail and are hence comparable. The auction proceeds in the form of a bidding process where potential buyers can make (money) offers for a certain product or service. There are different models to organize the bidding process [Bichler'00]. A comprehensive classification of negotiations with respect to auctions is provided by the Montreal taxonomy [StröbelWeinhardt'03]. Agent

negotiation means that an inanimate agent, i.e. a software artefact, carries out the process of negotiation on behalf of a principal, typically a human being. The principal delegates the task of negotiating to the agent by providing it with his or her preferences regarding the product or service to be procured. The agent has a certain autonomy to act within the limits of these preferences. Some models for agent negotiation are given in [DignumCortés'01]. The specification of preferences requires that the product or service in question can be described in detail. Hence both auctions and agent negotiation only work with standardized products / services.

The models we have discussed so far assume that most parameters of the contract are already predetermined and very few can actually be negotiated. Most often the only free parameter is the price. In many cases this restriction is not acceptable, i.e. we need more freedom in negotiating. This can, for example, happen if the product or service to procure is not standardized so that we have to negotiate many of its parameters. In such a case we need the third model, bargaining. In bargaining we assume that in principal all parts of a contract are negotiable, i.e. we start with an empty contract (although existing reference contracts or contract templates can be used as a starting point if desired). A number of bargaining models has been suggested such as the Three-Layer Architecture [ChiuCheungTill'03], SilkRoad [Ströbel'01], DOC.COM [SchoopQuix'01], MeMo Business Negotiation Support Metamodel [de MoorWeigand'04], Protocols for Electronic Negotiation Systems [KerstenStreckerLaw'04], and the Generic Model [MathieuVerrons'02]. To find a suitable negotiation model for business networks we must first identify the criteria that it should fulfill. Based on the studies mentioned above we have derived the following criteria: Communication, documents, deontics and time. The next sections argue for the necessity of these criteria.

3.1 Communication

The term communication is ambiguous. Communication takes place both on the business network level and on the communication network level but the meanings of the term in these contexts differ fundamentally. In the case of a communication network, communication is the central issue (hence the name). It consists primarily of an exchange of messages between inanimate agents, i.e. computers, IT systems or the like. On the other hand, communication in a business network consists of interaction between human beings (actors). Inanimate agents do not exhibit many of the qualities of human beings, such as conscience, responsibility, creativity and so on. This affects their ability to act as they cannot engage in social action, which requires

these capabilities. Negotiating a business contract is an example for a complex social process that involves social actions, e.g. making commitments.

Communication is the primary instrument for social interaction in general and for negotiation in particular. Negotiation consists basically of an exchange of messages between the negotiators. With these messages the negotiators create, modify and extend the contract, e.g. by making requests or commitments that ultimately lead to contractual obligations. It is therefore evident that a negotiation model for business network contracts must incorporate communication on a fundamental level. The importance of language for social action has been recognized early which led to the development of several theories, most notably Speech-Act Theory [Austin'62; Searle'69] and the Theory of Communicative Action [Habermas'84]. Many of the negotiation models that address the issue of communication are based on these theories.

3.2 Documents

The result of negotiation is a contract, which is obviously a document. Contract and negotiation are duals of each other in the same way that document and communication are. They are so tightly interwoven that it is impossible to separate the one from the other. The contract is a negotiation cast into a document. A negotiation model must therefore provide some mechanism to derive the contract from the negotiation messages in a transparent and traceable way. But documents play an important role already during negotiation. Preliminary contracts (contract versions) are a record of the negotiations that have been made so far. In this sense documents are an embodiment of past communications. We need the contract versions to mark important achievements in contract development, to understand why the contract has developed in that particular way and to go back to an earlier version if something has gone wrong. As this holds for all types of contracts we can conclude that documents must form an integral part of the foundation of a negotiation model for business networks.

3.3 Deontics

Deontic logic is concerned with reasoning about obligations and permissions. It has a direct bearing on negotiation as contracts are about determining obligations in exchange for granting permissions. For example, if Henry signs a contract about the sale of a car he enters into an obligation to pay a certain amount of money but in return he is granted the permission to take the car into his possession and to dispose of it in any way he wishes. Deontics trace the status of commitments during the course of a negotiation. As a rule an obligation arises only if all parties

agree on it. If Sally commits herself to do the shopping she is not yet under any obligation. Mike might, for example, make a counter-offer to do it for her. Only if Mike accepts Sally's commitment is she actually obliged to keep it. The same holds if Sally requests Mike to do the shopping, which he might simply deny. Only his agreement makes it an obligation. Keeping track of the deontic state is therefore important for any negotiation model, especially in the case of the complex negotiations involved in designing the business network.

3.4 Time

Time restrictions are an issue for many business actions. Some actions are not allowed to start before a certain point in time, others must be finished before a deadline has expired. A particular action might be required to be performed precisely at a specific time or repeatedly in certain intervals. It is therefore necessary that time restrictions for future actions can be negotiated as they are an important characteristic of the actions. But time-related issues are not only relevant at the level of the business process but also concerning the negotiation process itself. The time order of messages is relevant for the negotiation and the establishments of obligations and there are time limits for the completion of the process. Our negotiation model should therefore offer a language that provides a concept of time.

4 Architecture of a Business Network

[KöhneSchoopStaskiewicz'05] has performed an evaluation and comparison of 11 negotiation models with respect to 11 criteria among which the above mentioned criteria can also be found. The closest match to the requirements for a negotiation model is represented by DOC.COM [SchoopQuix'01] which fulfills three of the four criteria fully and one, deontics, at least partially. We have therefore chosen to adopt this model for the purpose of our study. As deontics is an important issue we have decided to add respective functionality to the negotiation system. But there is yet another problem that needs to be solved. The objective of DOC.COM is to represent a negotiation about the execution of a process instance, e.g. the delivery of a particular item on a particular date. But negotiations regarding the set-up and maintenance of a business network concern process types, e.g. the general business logic of order processing. The resulting contract is called a frame contract as it regulates the interaction among network members regarding a significant number of orders over time. To enable such negotiations we

have introduced a meta-layer into the negotiation language. Figure 1 shows the architecture of a system to set up and operate a business network.

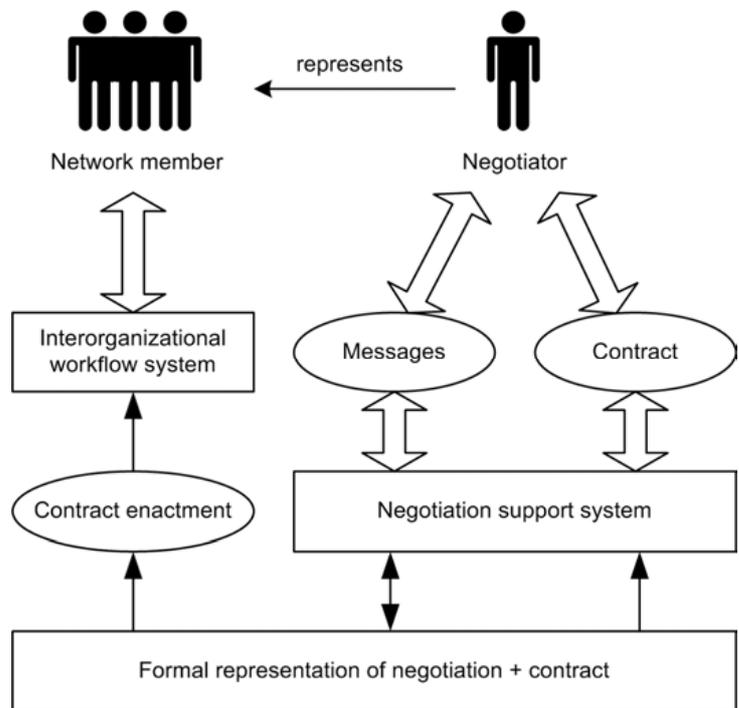


Figure 1: Architecture of a business network

A business network consists of a number of members. Each such member is typically an organization (i.e. a business) but could also be an individual who acts as an economic agent. Each member organization is represented by a negotiator who is entitled to carry out such negotiations and to sign a binding contract on behalf of the organization. This negotiator will interact with negotiators from the other members via a negotiation support system (NSS). The NSS consists of a message component and a contract component. The former handles both the translation of “human” negotiation messages into the formal representation in DOC.COM and the presentation of recorded formal negotiations in a human-readable form. The contract component stores the binding negotiations, which together make up the contract and which are also stored in DOC.COM, and represents this contract in a way that is similar to conventional, written contracts. The specific NSS for DOC.COM is called Negoisst [SchoopJertilaList'03]. The next section describes how negotiation and contract formation proceed.

The left part of Figure 1 shows how the operation of a business network is supported. We assume that the process of negotiation has led to a contract that deals with all relevant issues of the cooperation. This could be the negotiation of a completely new frame contract, i.e. the set-

up of a new business network. On the other hand the negotiation can also be about business network maintenance which involves adapting to the loss of members, incorporating new members, replacing parting members, reacting to changed requirements or the like. The contract under consideration will in any case be subject to enactment which yields a description of the interactions between the members in some kind of workflow language. The choice of this language depends on the workflow system that we choose to coordinate the workflow between members. In principle any workflow system can be used that allows for the implementation of the workflow patterns identified in [AalstHofstedeKiepuszewskiBarros'03]. Most commercial systems qualify if we allow for workarounds and coding but there is little native support for many of the advanced patterns. Only FLOWer supports directly or indirectly 16 of the 20 patterns. In a prototypical environment it can be useful to employ YAWL [AalstHofstede'05] that provides all patterns but one, together with the YAWL Engine. YAWL uses the same serialization language as the negotiation and contract language DOC.COM, i.e. XML. This facilitates enactment of the contract. YAWL makes also use of XQuery and XPath to extract data from XML input files and for generating XML output. This supports the integration with the enterprise application systems of the business network members, most of which can import and export in XML format. The resulting workflow system is run on a coordination server which can be seen as part of a communication network. An example of this is given in the section 4.3 based on a YAWL implementation. An overview and comparison of other languages for interorganizational workflows is given in [BernauerKappelKramlerRetschitzegger'03].

4.1 From Negotiation to Enactment

The previous section has described the general architecture of setting up and operating a business network. In this section we describe how the procedures in that architecture are performed and what the results look like. For this purpose we consider a simple negotiation, the corresponding part of the contract and the resulting workflow net (enactment) in some detail. This example represents only a very small part of the case and just serves to illustrate the way our approach works. The complete example is shown in the next section on a more general level.

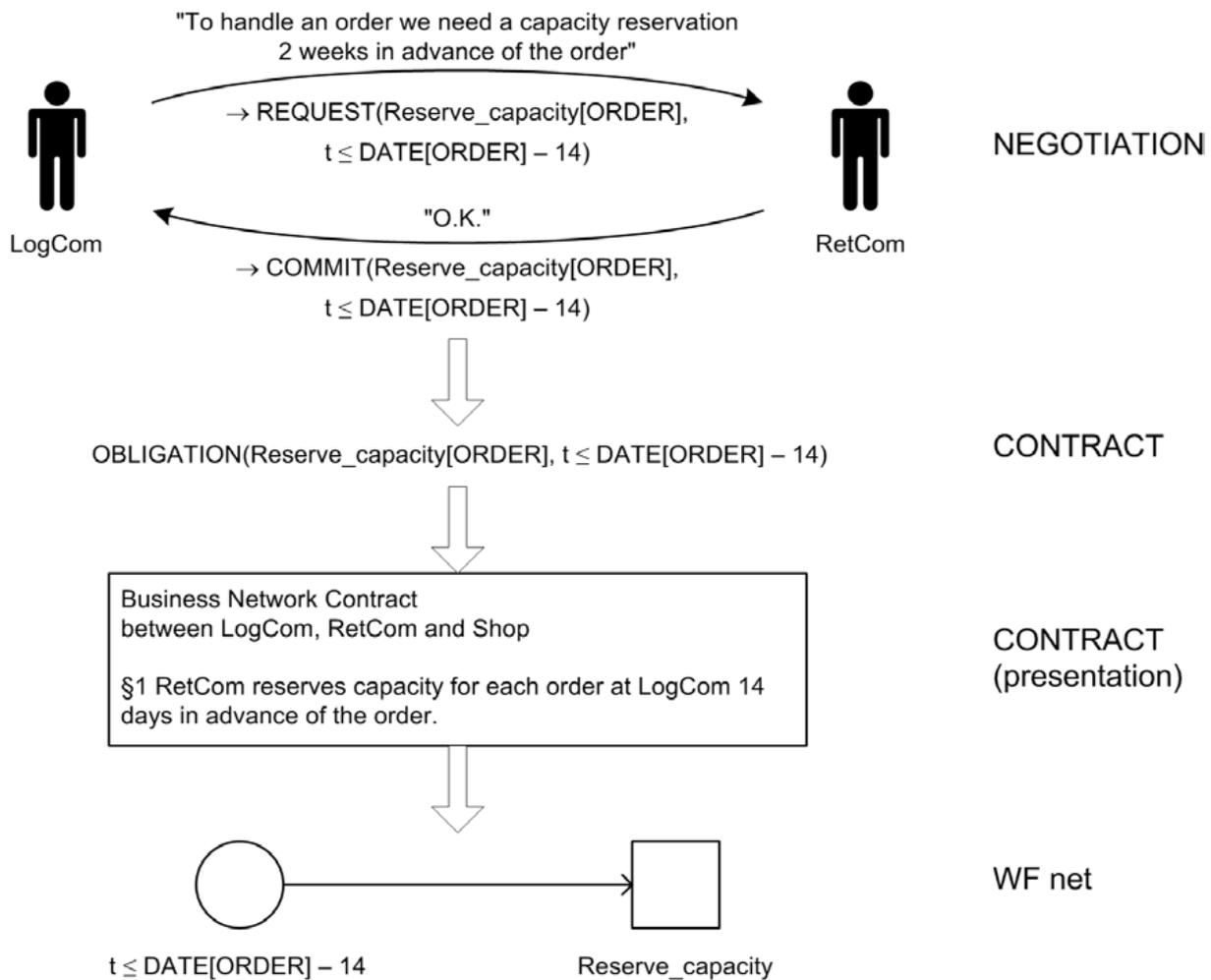


Figure 2: From negotiation to workflow net (example)

Our case involves three business partners: A retail chain in the home decoration industry (RetCom), the shops of this chain and a logistics company (LogCom). RetCom want that LogCom take over the delivery of orders for them. Figure 2 shows two steps in the respective negotiation between them. The representative from LogCom writes an email saying that they need a capacity reservation 2 weeks in advance of the order to be able to handle it. The negotiation support system helps with translating this request from the natural language to the internal, formal representation in DOC.COM:

$\text{REQUEST}(\text{Reserve_capacity}[\text{ORDER}], t \leq \text{DATE}[\text{ORDER}] - 14)$

The keyword REQUEST indicates that LogCom would like to introduce a new action into their cooperation. The propositional content of this message tells us what that action is, namely the reservation of capacity for each order. The request also specifies a time restriction for this action, i.e. 14 days in advance of the order date. This message is stored in the message memory of the negotiation system so that it can be matched with RetCom's reaction to it. In this case

RetCom fully agree with the action that was suggested by LogCom by answering with “O.K.”. Again the NSS will help with translating this to the formal representation:

COMMIT (Reserve_capacity[ORDER], $t \leq \text{DATE}[\text{ORDER}] - 14$)

The speech act COMMIT signals that RetCom agree to fulfil the request. A request that is followed by a commit with the same propositional content and restrictions leads to a binding obligation of the committing party towards the requesting party with respect to the content. An alternative reaction of RetCom could be:

COMMIT (Reserve_capacity[ORDER], $t \leq \text{DATE}[\text{ORDER}] - 7$)

which would be interpreted as: “We agree to reserve capacity but we cannot do it earlier than one week in advance.” Such a speech act does not create an obligation but constitutes a counter-offer. An acceptance of this counter-offer by LogCom would then create an obligation concerning the modified terms. In our example the original request is granted and a respective obligation is inserted into the contract:

OBLIGATION (Reserve_capacity[ORDER], $t \leq \text{DATE}[\text{ORDER}] - 14$)

The presentation component of the NSS can at any point in time display the contract that has been negotiated so far in a human-readable form (see Figure 2). In the final step the obligation is translated to a corresponding workflow.

In the case we have used an orchestration language (YAWL) for that purpose which allowed us to actually implement the coordination processes on a coordination server. The ultimate aim was for LogCom to operate this server and to offer this as an additional service. Another option, from a theoretical perspective, would be to use a choreography language (e.g. WS-CDL) and to decentralize the coordination so that each partner would be responsible for their respective part. In larger networks without any “central” node this is sometimes preferable, especially if the network is unstable. But often companies want to outsource this kind of information logistics in the same way they would outsource physical logistics.

4.2 Example

The example in the previous section was on a detailed level but covered only a small part of the case. Here we give a complete account of the case without the details concerning negotiation. We primarily focus on the “old” architecture of the retail network and the result of applying the procedure described in the previous sections with the aim of supporting network governance. We started our project with performing an analysis of the business processes between the companies we have already mentioned. These companies had already an established business

relationship that was based on a conventional frame contract. In the analysis we discovered the structure of the cooperation (see Figure 3) and a number of problems such as: broken interaction patterns, missing business rules, unclear communication structures, different contract interpretations and excessive interpersonal communication. As a consequence the parties were unsatisfied with the current situation.

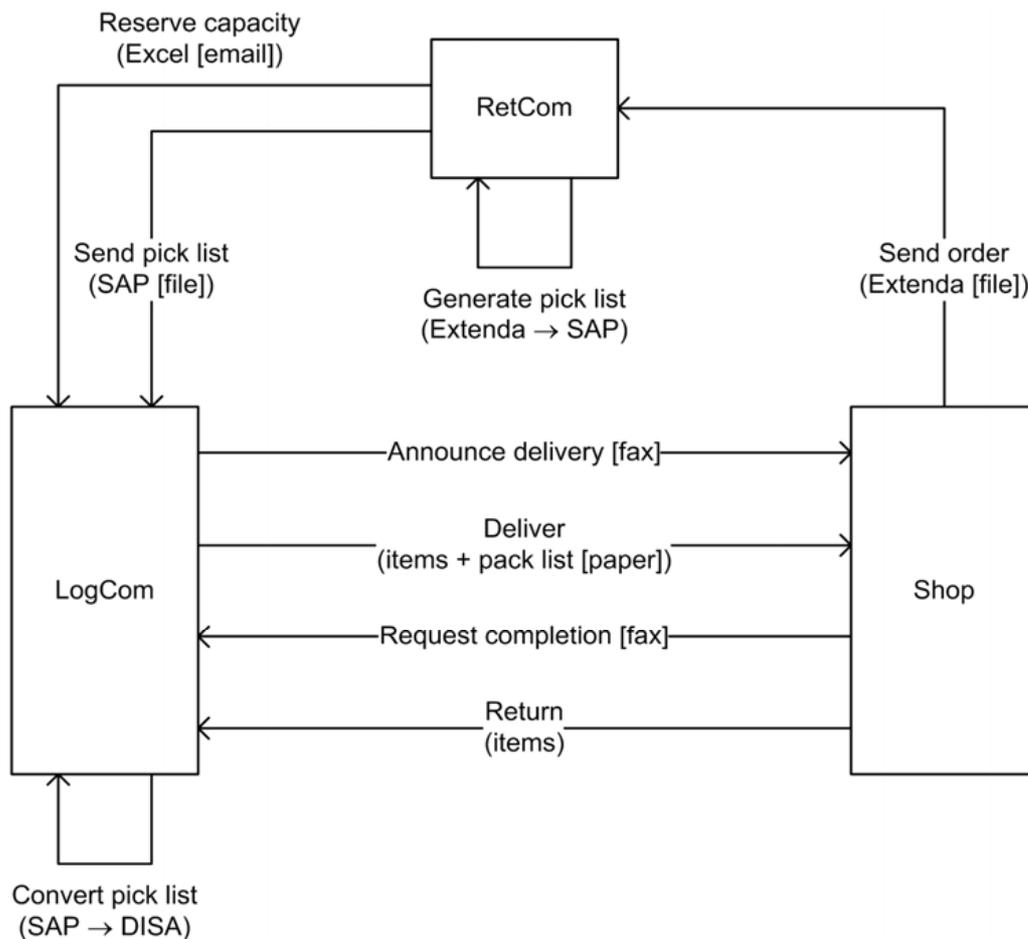


Figure 3: Original architecture of the network

To solve these problems we decided to support the coordination between the network members with the architecture introduced in section 4. We started with negotiating the formal contract. This was done in a seminar where the representatives of each organization were present and the seminar leader manually translated their requests and commitments into a formal representation according to the procedure described above. The reason for this is that the NSS does so far only support bilateral negotiations. We consider this as a technical restriction rather than a conceptual one and it should be possible to extend the NSS to multi-part negotiation. Enactment of the contract was done with the help of YAWL and the YAWL engine which was run on a

coordination server that connects all parties. The conversions between the involved formats (SAP, DISA, Extenda and Excel) have been performed with the help of XML Script and the X-Tract XML Script processor. This led to the architecture depicted in Figure 4.

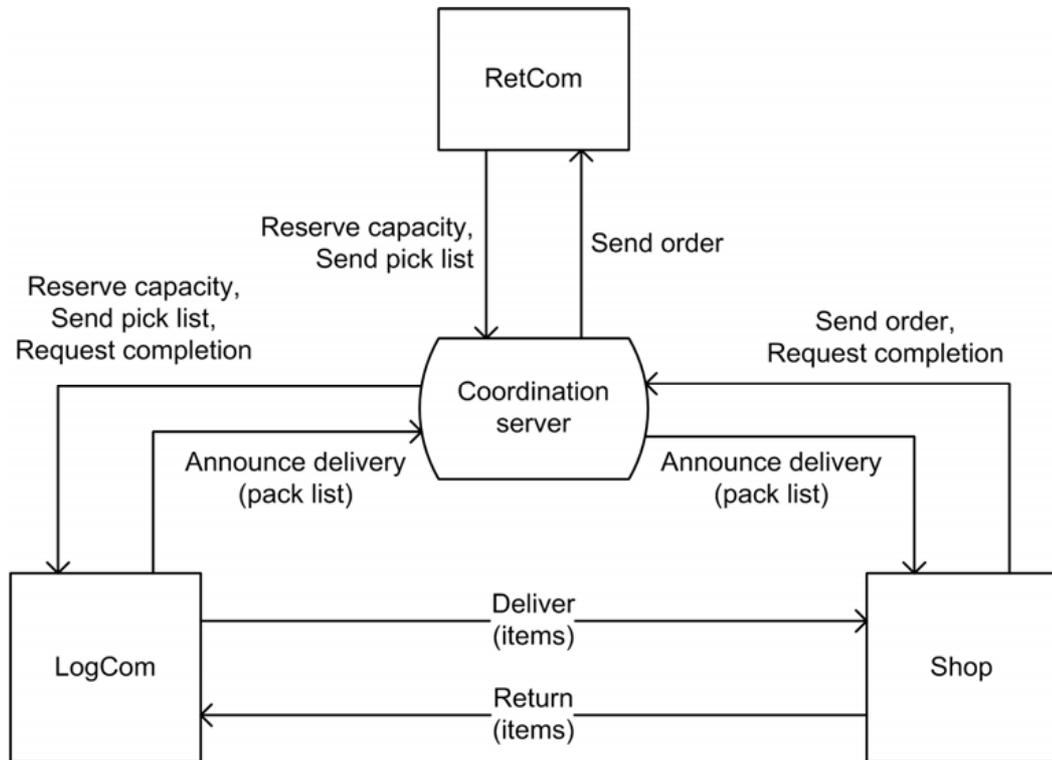


Figure 4: Revised architecture of the network

In the revised architecture each business network partner only exchanges messages with the coordination server. This reduces the complexity of the coordination considerably. The server takes care of forwarding messages to the right recipients, converting between formats, triggering time-controlled messages and so on. The business network architecture also offers ways to improve the efficiency of the communication. In our case, for example, the paper-based communication can be replaced by electronic messages, e.g. concerning the fax containing the pick list. The physical exchange between LogCom and the Shop can in this way be restricted to the exchange of the items themselves.

5 Conclusion

We started from the assumption that a business network consists of a group of businesses that collaborate on an equal footing and coordinate their interaction via a multi-part agreement. Based on relevant theories we identified a suitable type of contract, i.e. behavior-based, and suggested an architecture to negotiate and enact such a contract. Negotiation is a social process based on interaction between human actors, each of them possibly representing an organization. The nature of social systems and their processes requires that we select an approach that supports human communication as well as documents (as records of human or artefact activity), deontics (as states in the social world) and time. Based on these requirements we arrived at a language for expressing both the process of negotiation and its result, i.e. the contract. This language is DOC.COM. For the enactment of the contract we need an interorganizational workflow management system and a corresponding language. For this step we can so far not make a suitable suggestion as the commercial systems do not provide sufficient support for all required workflow patterns, and the research prototypes do not (yet) possess the maturity and stability required in real-life business applications. To show the feasibility of our approach, we have used it to set up and operate a prototypical business network.

6 References

[AalstHofstede'05]

Aalst, W.M.P. van der; Hofstede, A.H.M. ter: YAWL: Yet Another Workflow Language. In: Information Systems 30 (2005) 4, pp. 245-275.

[AalstHofstedeKiepuszewskiBarros'03]

Aalst, W.M.P. van der; Hofstede, A.H.M. ter; Kiepuszewski, B.; Barros, A.P.: Workflow Patterns. In: Distributed and Parallel Databases 14 (2003) 1, pp. 5-51.

[AlchianDemsetz'72]

Alchian, A.A.; Demsetz, H.: Production, information costs and economic organization. In: American Economic Review 62 (1972) 5, pp. 777-795.

[Austin'62]

Austin, J. L.: How to Do Things with Words. Oxford University Press, Oxford 1962.

[BernauerKappelKramlerRetschitzegger'03]

Bernauer, M.; Kappel, G.; Kramler, G.; Retschitzegger, W.: Specification of Interorganizational Workflows - A Comparison of Approaches. Paper presented at the 7th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2003), Orlando, FL, July 27-30, 2003.

[Bichler'00]

- Bichler, M.* : A Roadmap to Auction-based Negotiation Protocols for Electronic Commerce. Paper presented at the 33rd Hawaii International Conference on Systems Sciences, Hawaii, 2000.
- [ChiuCheungTill'03]
- Chiu, D. K. W.; Cheung, S. C. ; Till, S.*: A Three-Layer Architecture for E-Contract Enforcement in an E-Service Environment. Paper presented at the 36th Hawaii International Conference on System Sciences, Hawaii, 2003.
- [ClemonsReddiRow'93]
- Clemons, E.K.; Reddi, S.P.; Row, M.C.* : The Impact of Information Technology on the Organization of Economic Activity: The “Move to the Middle” Hypothesis. In: *Journal of Management Information Systems* 10 (1993) 2, pp. 9-35.
- [Coase'37]
- Coase, R.H.* : The nature of the firm. In: *Economica* 4 (1937), pp. 386-405.
- [de MoorWeigand'04]
- de Moor, A.; Weigand, H.*: Business Negotiation Support: Theory and Practice. In: *International Negotiation* 9 (2004) 1, pp. 31-57.
- [DignumCortés'01]
- Dignum, F.; Cortés, U.* (Eds.). (2001). *Agent-Mediated Electronic Commerce III*. Berlin: Springer.
- [GurbaxaniWhang'91]
- Gurbaxani, V.; Whang, S.* : The Impact of Information Systems on Organizations and Markets. In: *Communications of the ACM* 34 (1991) 1, pp. 59-73.
- [Habermas'84]
- Habermas, J.*: *The Theory of Communicative Action 1 - Reason and the Rationalization of Society*. Beacon Press, Boston 1984.
- [HollandLockett'97]
- Holland, C.P.; Lockett, A.G.* : Mixed Mode Network Structures: The Strategic Use of Electronic Communication by Organizations. In: *Organization Science* 8 (1997) 5, pp. 475-488.
- [JensenMeckling'76]
- Jensen, M.C.; Meckling, W.H.* : Theory of the firm: Managerial behavior, agency costs and ownership structure. In: *Journal of Financial Economics* 3 (1976), pp. 305-360.
- [KerstenStreckerLaw'04]
- Kersten, Gregory E.; Strecker, Stefan E.; Law, Ka Pong* Protocols for Electronic Negotiation Systems: Theoretical Foundations and Design Issues. In: *Bauknecht, K.; Bichler, M.; Pröll, B. (Eds.): E-Commerce and Web Technologies: Proceedings of the 5th International Conference, EC-Web 2004, Zaragoza, Spain, August 31-September 3, 2004*. Springer, Berlin 2004,
- [KleinCrawfordAlchian'78]
- Klein, B.; Crawford, R.; Alchian, A.* : Vertical integration, appropriable rents, and the competitive contracting process. In: *Journal of Law and Economics* 21 (1978), pp. 297-326.
- [KöhneSchoopStaskiewicz'05]
- Köhne, F.; Schoop, M.; Staskiewicz, D.*: A Meta Model for Electronic Negotiations - Comparison of existing approaches. In: *Tan, Y. H. (Ed.), Proceedings of the 12th Research Symposium on Emerging Electronic Markets (RSEEM 2005) “Governance of Electronic Markets”*. Vrije Universiteit Amsterdam, Amsterdam 2005,
- [MaloneYatesBenjamin'87]

- Malone, T.W.; Yates, J.; Benjamin, R.I.* : Electronic Markets and Electronic Hierarchies. In: Communications of the ACM 30 (1987) 6, pp. 484-497.
- [MathieuVerrons'02]
- Mathieu, P.; Verrons, M.-H.*: A generic model for contract negotiation. Paper presented at the AISB'02 Symposium on Intelligent Agents in Virtual Markets, Imperial College of Science, Technology and Medicine, University of London, 2002.
- [Ross'73]
- Ross, S.*: The economic theory of agency: The principal's problem. In: American Economic Review 63 (1973) 2, pp. 134-139.
- [SchoopJertilaList'03]
- Schoop, M.; Jertila, A.; List, T.*: Negoisst: a negotiation support system for electronic business-to-business negotiations in e-commerce. In: Data & Knowledge Engineering 47 (2003) 3, pp. 371-401.
- [SchoopQuix'01]
- Schoop, M.; Quix, C.*: DOC.COM: a framework for effective negotiation support in electronic marketplaces. In: Computer Networks 37 (2001) 2, pp. 153-170.
- [Searle'69]
- Searle, J. R.*: Speech Acts - An Essay in the Philosophy of Language. Cambridge University Press, London 1969.
- [Ströbel'01]
- Ströbel, M.*: Design of Roles and Protocols for Electronic Negotiations. In: Electronic Commerce Research 1 (2001) 3, pp. 335 - 353.
- [StröbelWeinhardt'03]
- Ströbel, Michael ; Weinhardt, Christof* The Montreal Taxonomy for Electronic Negotiations. In: Group Decision and Negotiation 12 (2003) 2, pp. 143-164.
- [Williamson'75]
- Williamson, O.E.*: Markets and Hierarchies. Free Press, New York 1975.
- [Williamson'81]
- Williamson, O.E.*: The modern corporation: Origins, evolution, attributes. In: Journal of Economic Literature 19 (1981), pp. 1537-1568.
- [Williamson'85]
- Williamson, O.E.*: The Economic Institutions of Capitalism. Free Press, New York 1985.
- [Wilson'68]
- Wilson, R.*: The theory of syndicates. In: Econometrica 36 (1968), pp. 119-132.

Integration of Conceptual Process Models

by the Example of Event-driven Process Chains

Carlo Simon

Institut für Management
Universität Koblenz-Landau
56016 Koblenz, Deutschland
simon@uni-koblenz.de

Jan Mendling

Institut für Wirtschaftsinformatik und Neue Medien
Wirtschaftsuniversität Wien
1090 Wien, Österreich
jan.mendling@wu-wien.ac.at

Abstract

It has become common place in business life that companies with related operations engage in a so-called merger in order to benefit from synergies or from combined products and services. In order to handle the complexity of such an endeavour, it is important to utilise a structured approach for finding similarities and contradictions in business process models of both partners. In this paper, we present a suitable procedure for this task. Furthermore, we demonstrate how to identify those specific activities within the overall business processes which must be adapted. In particular, we discuss how such integration can be conducted if the processes of both parties are modelled with Event-driven Process Chains, one of the most popular conceptual business process modelling languages. By the help of a running example we illustrate the join operator for the integration of these models and the interpretation of the result.

1 Introduction

It has become common place in business life that companies with related operations engage in a so-called merger in order to benefit from synergies or from combined products and services. The rationale behind such a merger is that the combination of both companies' operations is expected to result in lower total cost as for the sum of both and a wider range of capabilities. It is a prerequisite for the leveraging of these benefits that the operational infrastructure of the business processes of the merging partners is integrated. Since the complexity of such an endeavour is a considerable challenge, the enterprise model repositories of both companies play an important role to guide and structure this integration procedure.

Several perspectives have been proposed to document information systems of enterprises. The control flow perspective of the business process is the most important one mentioned e.g. by [ÖBH91, p. 173, Sch00, p. 41]. Other views are organisation, function and data in accordance with the St. Galler information system's architecture of Österle et al. [ÖBH91, p. 173]. Scheer extends this model by output and control within the architecture of integrated information systems (ARIS) [Sch00, p. 41]. Axenath et al. [AKR05, p. 6] add authorisation and authentication as well as assignment as further perspectives. These rather technical views on businesses may be extended by strategies as proposed by Frank [Fra94, p. 170] or Krcmar [Krc05, p. 43].

With respect to the integration of information models in case of a merger, there is extensive work reported in the database community on *view* and *schema integration of data models*. In the 1980s, Batini et al. [BLN86] provide a comparative analysis of schema integration methodologies. They distinguish *preintegration*, *comparing*, *conforming*, *merging*, and *restructuring* as schema integration activities. Several contributions focus on specific activities of this integration process. Rahm and Bernstein provide a survey on how matches across different schemas can be identified automatically [RB01]. Rizopoulos and McBrien discuss a hypergraph data model (HDM) with a set of semantic relationships to support the merge operation [RM05]. A comprehensive integration method is provided by [SS05].

While the integration of data models is a rather mature research discipline, surprisingly little work has been conducted on the integration of process models in theory and practice. Most of these contributions offer integration procedures on the level of Petri nets with only a few like [GRSS05a, GRSS05b] covering generic aspects. For this paper, we adopt the integration process for a conceptual process modelling language as presented in [MS06] since it provides a straight forward support for Event-driven Process Chains (EPCs). For further related work on

behaviour integration, we also refer to [MS06]. The approach which is presented in the following builds on general insights from database schema integration and integration operators that borrow ideas from the Semantic Process Language [Sim06]. Moreover, it focuses on the control flow aspect of business process models. The remainder of the paper is structured as follows. Section 2 gives a theoretical foundation of the process integration methodology applied in this paper. Section 3 provides a formal syntax definition of EPCs and introduces a running example of two procurement EPCs. Section 4 applies the integration process as defined in Section 2 to the two example EPCs. The paper closes with a conclusion in Section 5.

2 The Process Model Integration Process

The integration of two business process models needs to consider those parts of the processes that coincide or contradict each other. This bears the following challenges:

- If the business processes are described using different modelling languages, then a translation of these models into a unique representation must be made first. For this reason a *meta-model* or *concept-ontology* must be defined for each language, i.e. the source languages and the target language which, however, do not necessarily have to be different. Beside the syntactical concepts also semantics must be described and mapped to each other.
- A *domain ontology* is needed to identify those actions which can be used for the synchronisation of the business process models. Therefore, the domain ontology must describe the language of the various domain experts in order to match their views of the modeled field of application.
- A formal *integration operator* must be defined to join two process models into a single model which can then be used to advice mergers concerning their mutual adoption of business processes. For conceptual process models, only heuristics can be formulated for this task, since without a state semantic a formal equivalence relationship cannot be defined.

The purpose of an ontology for the development of information systems is to describe human (domain specific) language in such a way that it can be represented and processed by software [EGHS05, p. 204]. For the particular modelling problem discussed here we need such a formali-

sation of the domain which justifies each merge operation on event and function level. A possible representation of ontologies ranges from *unstructured* over *semi-structured* to *fully structured* [GFC04, p. 169]. An unstructured ontology is described in natural language, a semi-structured ontology in natural language which is restricted to a specific form, and a structured one is defined with the aid of a formal language which supports proofs concerning soundness and completeness.

In this paper, we use a semi-structured approach to define the domain ontologies of the examples focusing on two aspects: first, the purpose of each process function is described by natural language. Second, the relation to other functions within the process is taken down. While formal integration of (business) processes has been reported by Simon [Sim06] for a Semantic Process Language (SPL), for Petri nets in general by Best et al. [BDK01], and for Process Algebra [BPS01, Fok00], we define and apply a process model integration process that is especially tailored for conceptual business process modelling languages such as EPCs. We specify the merge operator on the level of the formal syntax of EPCs. In this way, we extend previous work on the integration of Petri nets (see e.g. [PCS01, Sim06]).

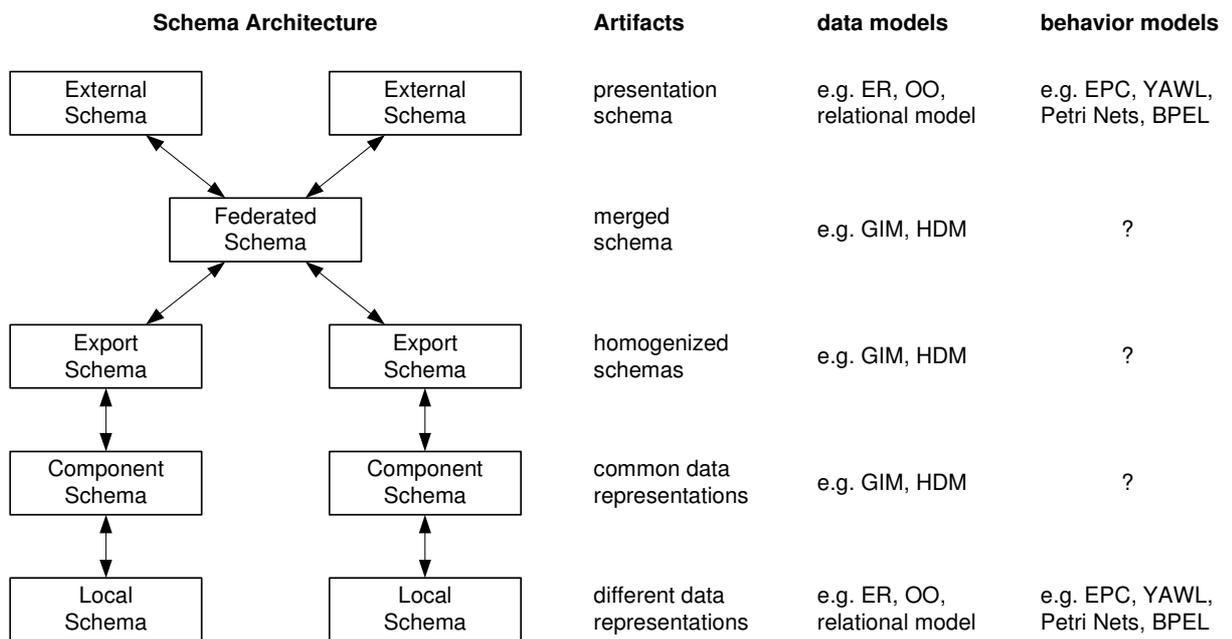


Fig. 1: Database Schema Integration and Process Model Integration [MPZ05]

In essence, the process model integration process can be specified analogously to the database schema integration process as defined in [SL90, SS05], compare Figure 1. While explaining the integration steps of the process model integration process, we give the terms from database integration according to [SL90] in brackets to establish the link. The process takes two process

models as input (“local schema”). These models might comply with two different business process modelling languages, e.g. one EPC and one BPEL model. In a first step, the models have to be mapped to a common business process modelling language that is utilised throughout the further integration. As a result, this yields the two models in the same language (“component schema”). In a second step, the elements of the process models have to be matched. In particular, potential homonyms and synonyms have to be analysed with special attention. The resulting models are called “export schemas” according to [SL90]. The third step represents the application of the merge operator and we achieve an integrated process model (called “federated schema” in [SL90]). Depending on user requirements, the integrated process model could be mapped to another process modelling language for presentation purposes (“external schema”). In this paper, we consider two business process models that are both modelled as EPCs. Therefore, we do not have to apply the mapping to a common process modelling language. Furthermore, we have to select the right variant of the merge operator: if the process models capture different views on the same business process, the integration is a kind of conjunction of the models (“integration by specialization” in [PCS01]). If the models represent process variants, the integration is a kind of disjunction (“integration by generalization” in [PCS01]). For a conjunction based integration of EPC business process models and further related work, refer to [MS06].

3 Event-driven Process Chains (EPCs)

Event-driven Process Chains (EPCs) are a conceptual business process modelling language. EPCs capture the temporal and logical dependencies of activities of a business process [KNS92]. So called function type elements represent activities of a business process. Event type elements describe pre- and post-conditions of functions, and three kinds of connector types including AND, OR, and XOR. Control flow arcs are used to link these elements. Connectors have either multiple incoming and one outgoing arc (join connectors) or one incoming and multiple outgoing arcs (split connectors). As a syntax rule, functions and events have to alternate, either directly or indirectly when they are linked via one or more connectors. The syntax of EPCs can be formally defined as follows (cf. [MS06]):

Notation 1 (Predecessor and Successor Nodes) Let N be a set of nodes and $A \subseteq N \times N$ a binary relation over N defining the arcs. For each node $n \in N$, we define the set of *predecessor nodes* $\bullet n = \{x \in N \mid (x, n) \in A\}$, and the set of *successor nodes* $n \bullet = \{x \in N \mid (n, x) \in A\}$.

Definition 1 (EPC) An EPC = (E, F, C, l, A) consists of three pair wise disjoint sets E, F, C of nodes, a mapping $l: C \rightarrow \{\text{AND, OR, XOR}\}$, and a binary relation $A \subseteq (E \cup F \cup C) \times (E \cup F \cup C)$ such that

- $|\bullet e| \leq 1$ and $|e \bullet| \leq 1$ for each $e \in E$. An element of E is called *event*.
- $|\bullet f| = 1$ and $|f \bullet| = 1$ for each $f \in F$. An element of F is called *function*.
- Either $|\bullet c| = 1$ and $|c \bullet| > 1$ or $|\bullet c| > 1$ and $|c \bullet| = 1$ for each $c \in C$. An element of C is called *connector*.
- The mapping l specifies the *type of a connector* $c \in C$ as AND, OR, or XOR.
- A defines the control flow as a coherent, directed graph. An element of A is called *arc*.

We define the semantics of the different EPC connector types in an informal manner since we only consider the structure of an EPC in the integration process. For an overview of formalisation approaches refer to [Ki06]. An AND split activates all subsequent branches in concurrency while the XOR split activates one of the subsequent branches. The OR split triggers at least one and at most all of multiple branches. The AND join synchronises all incoming branches, then it activates the subsequent EPC element. The OR join synchronises all active branches. This feature has been debated as non-local semantics (see e.g. [Ki06]). The XOR split has also non-local semantics. Either there is one input branch active (which is the expected case) and it activates the subsequent EPC element, or there are multiple branches active and it blocks.

In the remainder of this section, we discuss two example EPCs that both represent a standard procurement process. The labels of the second EPC were translated from German. The first procurement process considered here is taken from [Roh96, p.49]. The start event represents the readiness to calculate procurement requirements on base of a master requirements plan and inventory data. The corresponding calculations are repeated until all open items are handled. Afterwards, these calculated net requirement orders are taken to generate and release an order and the actual procurement is conducted. Figure 2 shows the EPC model of this procurement process.

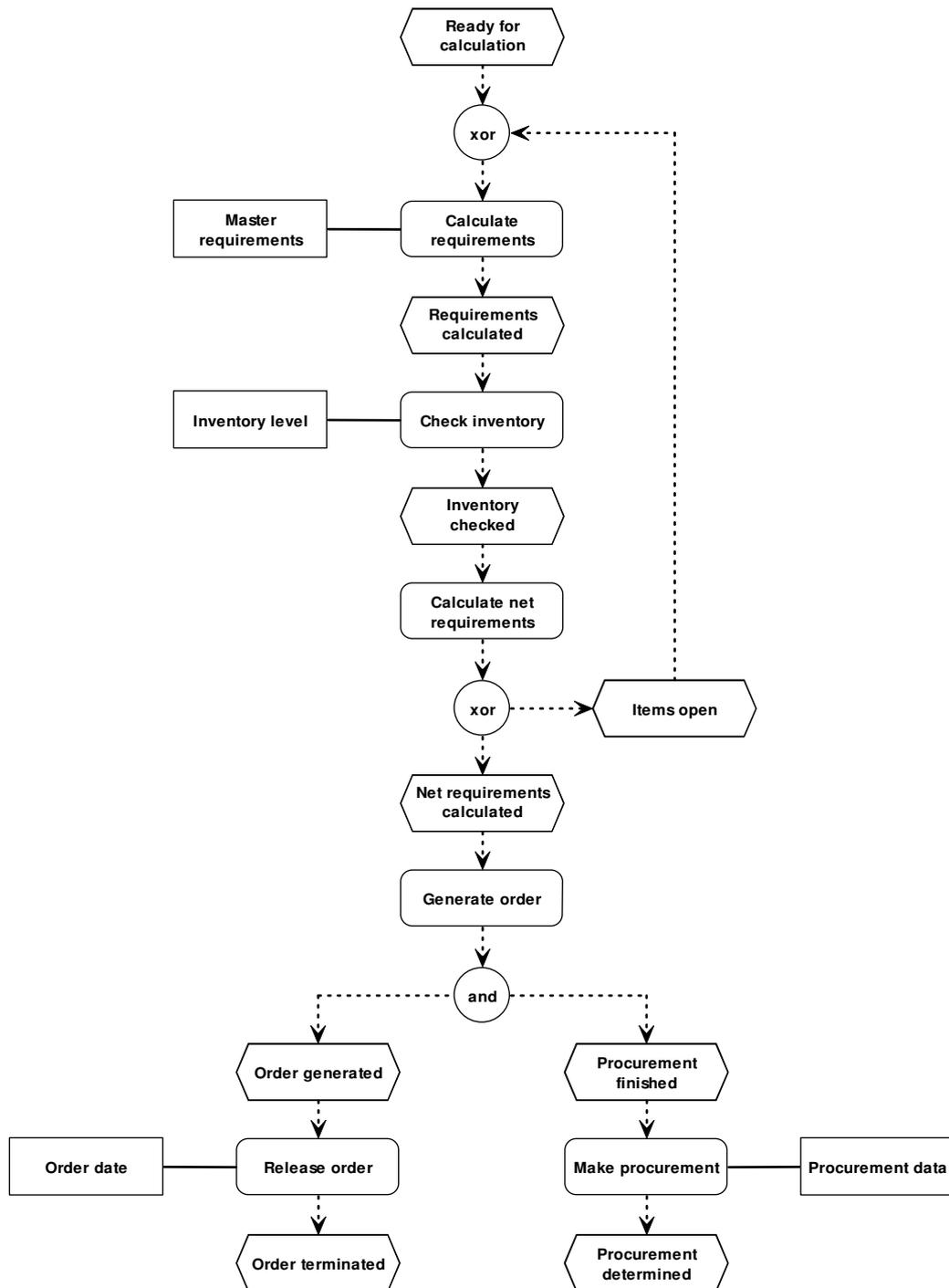


Fig. 2: Procurement process in adoption of [Roh96, p. 49]

The second procurement process is taken from [BS96, p. 65]. A fund manager starts a procurement process on behalf of a recognised demand with stocktaking of the requested products in the warehouse or in sourced out third-party warehouses. On base of the currently available amount of goods the purchase order quantity is calculated and an order is initiated which increases the warehouses. Figure 3 depicts this business process in a diagram.

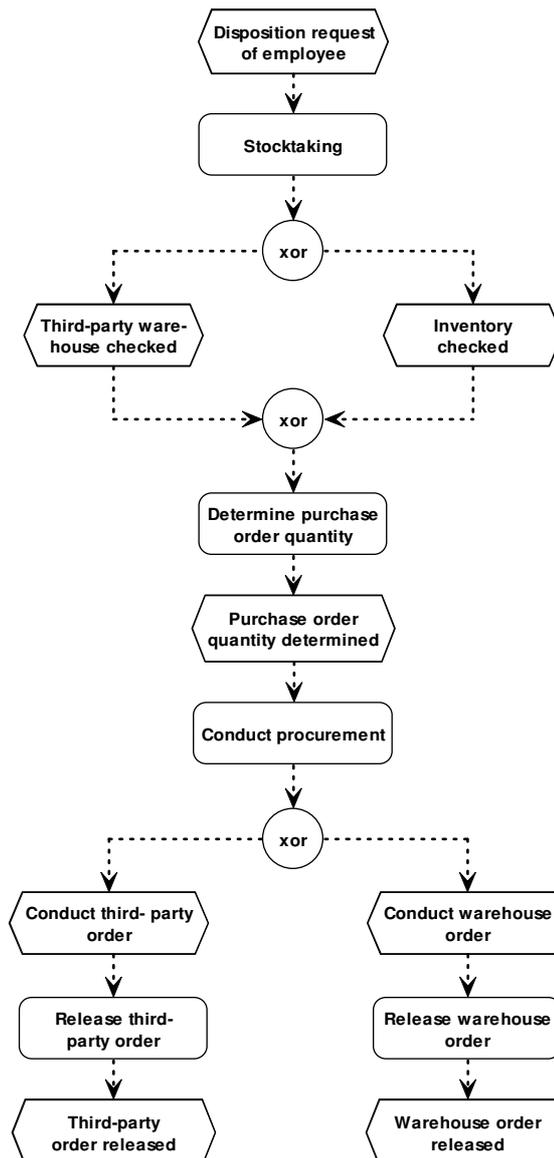


Fig. 3: Procurement process in adoption of [BS96, p. 65]

4 Merging EPCs based on Process Model Integration

Since both models are developed as EPCs, a concept ontology is not needed for an integration if this operation can be applied immediately to the EPC models. In the subsequent section the theoretical foundations will be laid for this and the integration operator is applied to the examples.

Domain ontologies for the examples must explain all used functions and events. As an input for their specification, we utilise both the (narrative) description of the functions in the EPC models

and their sequence relation to other functions. These facts provide adequate information to draw conclusions concerning the similarity of functions in different models. The events of the models play a subordinate role within these examples since they primarily describe intrinsic states of the process progress which can also be derived from the functions and the sequence in which they occur.

Table 1 shows the domain ontology for the model of Rohloff in tabular form explaining the functions and relates them to their predecessor and successor functions.

Action	Predecessor	Successor	Description
<i>Calculate requirements</i>		<i>Check inventory</i> or <i>Calculate net requirements</i>	Determine the need
<i>Check inventory</i>	<i>Calculate requirements</i>	<i>Calculate net requirements</i>	Check current stock amount
<i>Calculate net requirements</i>	<i>Check inventory</i>	<i>Release order</i> and <i>Make procurement</i>	Calculate procurement amount
<i>Generate order</i>	<i>Calculate net requirements</i>		Formulate request
<i>Release orders</i>	<i>Generate order</i>		Distribute order
<i>Make procurement</i>	<i>Generate order</i>		Control procurement

Tab. 1: Domain ontology for the functions in the model of Rohloff

Tabular 2 shows the domain ontology of Becker and Schütte in a similar format. Both representations are used to compare the domain ontologies and to find similarities.

The first function in the model of Rohloff is the calculation of requirements while in the model of Becker and Schütte the process starts with stocktaking followed by determining the purchase order quantity. This indicates that the required amount of goods must have been calculated before a process described by Becker and Schütte starts. And indeed, processes of Becker and Schütte start with an event *Disposition request of employee* which indicates that the demand has already been calculated. It is therefore not possible to identify an analogue function for *Calculate requirements* in the second model.

Action	Predecessor	Successor	Description
<i>Stocktaking</i>		<i>Determine purchase order quantity</i>	Determine current inventory
<i>Determine purchase order quantity</i>	<i>Stocktaking</i>	<i>Conduct procurement</i>	Determine required quantities
<i>Conduct procurement</i>	<i>Determine purchase order quantity</i>	<i>Release third-party order or Release warehouse order</i>	Generate order
<i>Release third-party order</i>	<i>Conduct procurement</i>		Order for external warehouse
<i>Release warehouse order</i>	<i>Conduct procurement</i>		Order for internal warehouse

Tab. 2: Domain ontology for the functions in the model of Becker and Schütte

The next function in the model of Rohloff is *Check inventory* which can be seen as similar to *Stocktaking* due to their descriptions. A difference between these two functions can only be observed, if the following events are considered. Becker and Schütte explicitly distinguish between internal and external warehouses while such a differentiation cannot be identified in the description of Rohloff. Despite of this distinction, both functions in principle describe the same kind of activity.

The next following function in the model of Rohloff is *Calculate net requirements* and in the model of Becker and Schütte *Determine purchase order quantities*. Both functions can be seen as similar due to their description. For the same reasons, also *Generate order* and *Conduct procurement* are diagnosed as similar.

At the end of the procurement process of Rohloff, two functions occur (*Release order* and *Make procurement*) while the process of Becker and Schütte ends with an alternative (between *Release third-party order* and *Release warehouse order*). This differentiation results from the two different warehouses (internal and external) and does not occur in the model of Rohloff. It can therefore be concluded that these two functions are specialisations of *Release order*.

5 Integration of the EPC Models

In this section, a heuristic is developed for the join of EPC models such that the resulting model represents the intersection of the processes of the input models. Due to the absence of a formal state semantics, it cannot be proved that this is formally true like in the Semantic Process Language. Nonetheless, we provide evidence for the usefulness of the chosen definition by explaining the outcome of each integration step. We then apply the heuristic to our example.

Since it is the goal of a merger to integrate the former individual views into a single one, the following join operator is conceptualised as a conjunction of models. The resulting EPC in principle describes the intersection of the input schemas.

Definition 2 (Joined EPC) Let $EPC_1 = (E_1, F_1, C_1, I_1, A_1)$ and $EPC_2 = (E_2, F_2, C_2, I_2, A_2)$ be two EPCs. The *Joined Event-driven Process Chain* $EPC_J = (E_J, F_J, C_J, I_J, A_J)$ – the conjunction of EPC_1 and EPC_2 – is defined in three consecutive steps as follows:

1. Basically, the elements of EPC_1 and EPC_2 are combined in a single diagram:

$$E_J'' := E_1 \cup E_2 \quad \text{and} \quad F_J'' := F_1 \cup F_2,$$

$$C_J'' := C_1 \cup C_2 \quad \text{and} \quad I_J'' := I_1 \cup I_2,$$

$$A_J'' := A_1 \cup A_2$$

2. Each pair (e_1, e_2) of similar event elements $e_1 \in E_1$ and $e_2 \in E_2$ describing the same real-world events is fused into a single one. Former incoming and outgoing control flow arcs are synchronised with the aid of two new connectors c_{split} and c_{join} :

$$E_J' := E_J'' - \{e_2\} \quad \text{and} \quad F_J' := F_J''$$

$$C_J' := C_J'' \cup \{c_{split}, c_{join}\}$$

$$I_J' := I_J'' \cup \{(c_{split}, \text{and}), (c_{join}, \text{and})\} \text{ and}$$

$$\forall x_1 \in \bullet e_1, \forall x_2 \in \bullet e_2, \forall y_1 \in \bullet e_1, \forall y_2 \in \bullet e_2:$$

$$A_J' := A_J'' - \{(x_1, e_1), (x_2, e_2), (e_1, y_1), (e_2, y_2)\} \cup$$

$$\{(x_1, c_{join}), (x_2, c_{join}), (c_{join}, e_1), (e_1, c_{split}), (c_{split}, y_1), (c_{split}, y_2)\}$$

Incomplete tuples due to missing predecessor or successor nodes are omitted.

The result of this operation is that each merged event occurs after all its successor functions and that it triggers all subsequent functions.

3. Each pair (f_1, f_2) of similar function elements $f_1 \in F_1$ and $f_2 \in F_2$ describing the same real-world events is fused into a single one. Former incoming and outgoing control flow arcs are synchronised with the aid of two new connectors c_{split} and c_{join} :

$$E_J := E_J' \quad \text{and} \quad F_J := F_J' - \{f_2\}$$

$$C_J := C_J' \cup \{c_{split}, c_{join}\}$$

$$I_J := I_J' \cup \{(c_{split}, \text{and}), (c_{join}, \text{and})\} \text{ and}$$

$$\forall x_1 \in \bullet f_1, \forall x_2 \in \bullet f_2, \forall y_1 \in \bullet f_1, \forall y_2 \in \bullet f_2:$$

$$A_J := A_J' - \{(x_1, f_1), (x_2, f_2), (f_1, y_1), (f_2, y_2)\} \cup$$

$$\{(x_1, c_{join}), (x_2, c_{join}), (c_{join}, f_1), (f_1, c_{split}), (c_{split}, y_1), (c_{split}, y_2)\}$$

Incomplete tuples due to missing predecessor or successor nodes are omitted.

The result of this operation is that each merged function is only executed if all its successor events have occurred and that it activates all possible follower events.

The first step results in a single EPC model which contains both input EPCs without any connections between them. The second step joins events. Most of the functions have similar events according to the fact that they only describe internal states of the processes. Similarity can be assumed between those events for which a similarity has been identified concerning their adjacent functions (*Requirements calculated* and *Disposition request of employee* as well as for *Net requirements calculated* and *Purchase order quantity determined*). The joined events are labelled using the terminology of Rohloff. Figure 4 exemplarily depicts the join of *Net requirements calculated* and *Purchase order quantity determined*.

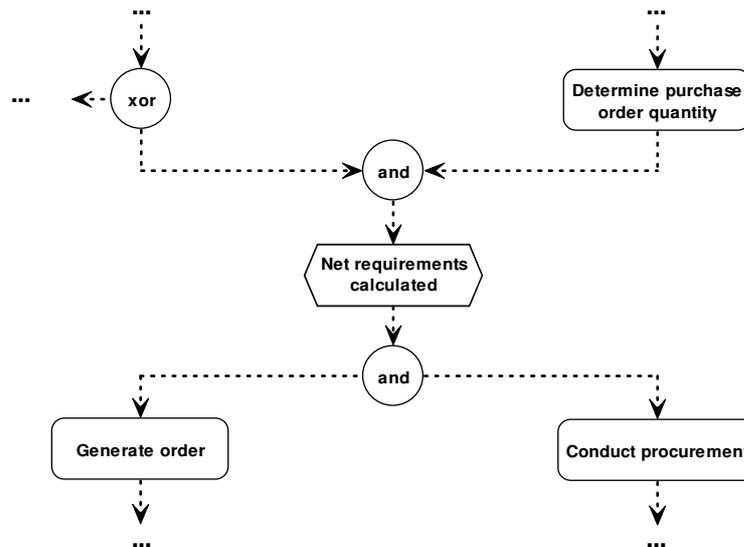


Fig. 4: Result of joining two events

Moreover, *Inventory checked* of Rohloff can be seen as a generalisation of *Third-party warehouse checked* and *Inventory checked* of Becker/Schütte. These events may be joined as well and the resulting events are labelled using the more specific terminology of Becker/Schütte. Definition 2, however, only provides means to join one event of the first EPC with one event of the other. If more than one join makes sense (due to a generalisation/specialisation relationship) each of the specialised joined functions must be enabled with respect to the preceding functions in the more general model i.e., further OR-connectors have to be added. The corresponding formal graph transformation operation can be specified similar to the previously defined operations.

The join of functions uses the ontologies defined in the previous section. Although *Check inventory* and *Stocktaking* are in principle identical, the latter one takes into account the existence of different warehouses which means that this term is a specialisation of *Check inventory*. The joined function is therefore labelled *Stocktaking*. Concerning the in each model following two functions, the join result is labelled using the terminology of Rohloff.

Also concerning the functions a generalisation/specialisation relationship can be observed. The corresponding functions can be merged with the same operation described for events already, i.e. by adding additional OR-connectors.

In the joined model, we observe an inflation of connector symbols. Many of them can be omitted since they are redundant or because they only have one incoming and outgoing arc. Figure 5 shows the resulting Joined EPC of the input schemas.

The resulting EPC model can be interpreted as follows: in the joined model, we still find valid processes. Consequently, a merge is possible. Nonetheless, also deadlocks can be observed at the end of the process. These deadlocks result from the differentiation of warehouse types made in one model but not in the other. A deadlock, however, does not represent a weakness of the presented approach. In contrary, such contradictions are typical for mergers. They help to find those parts of businesses which need to be restructured and where operations have to be aligned.

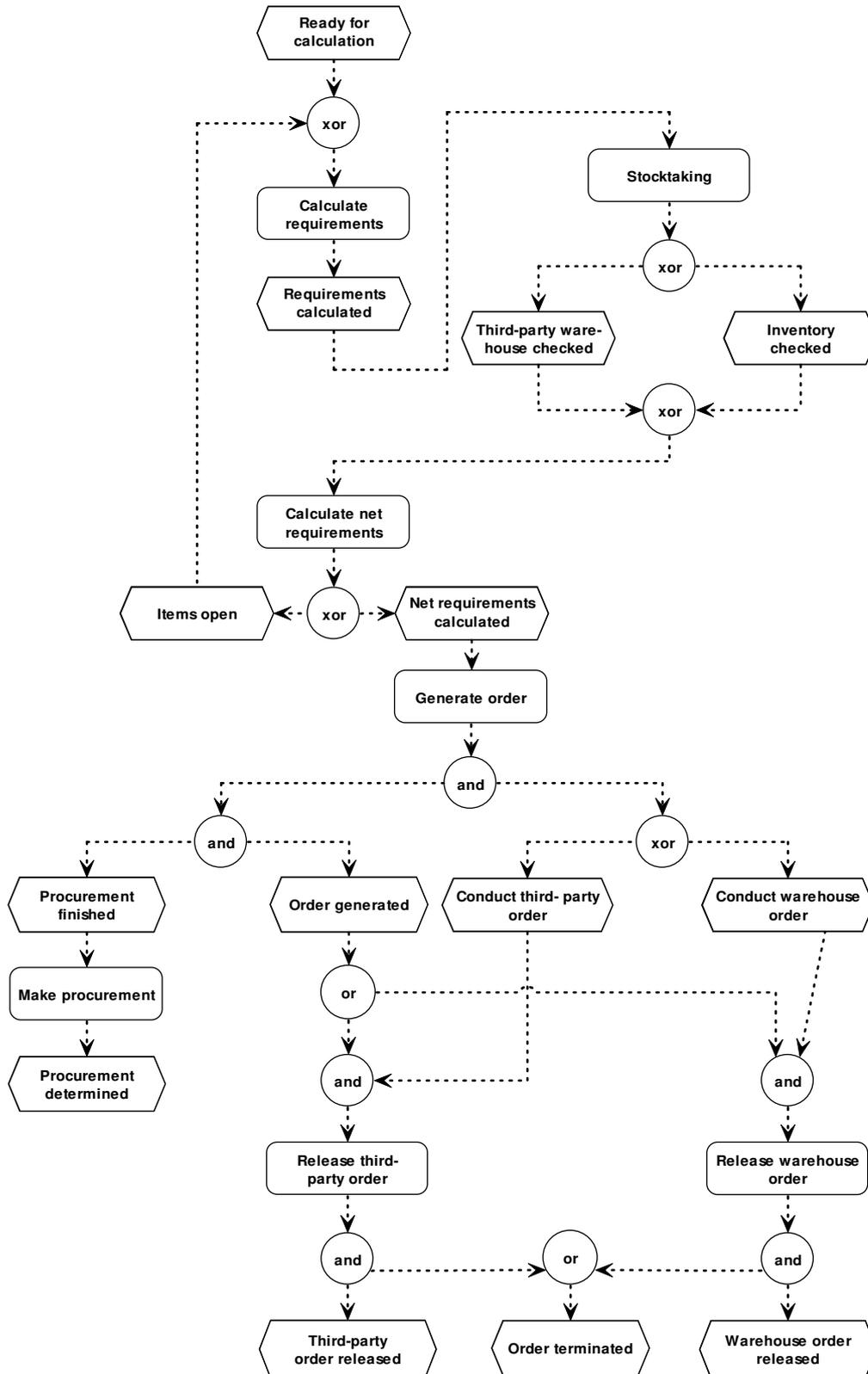


Fig. 5: Joined EPC of Rohloff and Becker/Schütte

6 Conclusion

Support for the conceptual integration of business process models is a key requirement for combining business processes in a merger scenario. Within this paper, we introduce a novel integration operator for conceptual process models (in particular EPCs). We follow an integration process which is adapted from database integration. Within this integration process, a merge operator specific for conceptual business process models is used to combine the models. In order to demonstrate the feasibility of our integration approach, we integrate two reference procurement processes taken from literature. Moreover, the actual merge ontologies are formulated building on the structure of the input EPCs.

The resulting EPC demonstrates that – although the input EPCs have some similarities – in this merger scenario a partial restructuring of the processes would be necessary. The method guides to these critical subsequences and detects those parts of the process models that differ. Since the method is applied immediately to the EPC input models, in opposite to former work [SM06] which required an intermediary transformation into Petri nets, the presented approach takes both into account, functions and events.

The integration operator that we use results in a conjunction of the input schemas. From database integration, also scenarios are known where the result equals a disjunction. The usefulness of a transfer of this integration type to process models will be considered in future research. Furthermore, we aim to provide tool support for the integration process. Beyond that, the process view is only one of many perspectives on information systems of a company. Future work will have to combine the data and the process integration into a multi-perspective integration approach.

References

- [Aal05] Aalst, van der, W. M. P.: Pi calculus versus Petri nets: Let us eat “humble pie” rather than further inflate the “Pi hype”. *BPTrends*, 3(5), 2005, pp. 1-11
- [AKR05] Axenath, B.; Kindler, E.; Rubin, V.: *The Aspects of Business Processes: An Open and Formalism Independent Ontology*. Fachberichte Informatik tr-ri-05-256, Universität Paderborn.

- [BDK01] Best, E.; Devillers, R.; Koutny, M.: Petri net Algebra. In: Brauer, W.; Rozenberg, G.; Salomaa, A. (Edt.): EATCS, Monographs in Theoretical Computer Science, Springer, Berlin, 2001
- [BLN86] Batini, C.; Lenzerine, M.; Navathe, S. B.: A Comparative Analysis of Methodologies for Database Schema Integration. ACM Computing Surveys, 18(4), 1986, pp. 323-364
- [BPS01] Bergstra, J. A.; Ponse, A.; Smolka, A. (Edt.): Handbook of Process Algebra, Elsevier, Amsterdam, 2001
- [BS96] Becker, J.; Schütte, R.: Handelsinformationssysteme, Verlag Moderne Industrie, Landsberg/Lech, 1996
- [EGHS05] Elzenheimer, M.; Grollius, T.; Heinemann, E.; Sternhuber, J.: Vergleichende Buchbesprechung: Ontologien. Wirtschaftsinformatik, 47(4), 2005, pp. 298-309
- [Fok00] Fokkink, W.: Introduction to Process Algebra. Springer, Berlin, 2000
- [Fra94] Frank, U.: Multiperspektivische Unternehmensmodellierung. Theoretischer Hintergrund und Entwurf einer objektorientierten Entwicklungsumgebung. Oldenbourg Verlag, München, 1994
- [GFC04] Gómez-Pérez, A.; Fernández-López, M.; Corcho, O.: Ontological Engineering. Springer, Berlin, 2004
- [GRSS05a] Georg Grossmann, Yikai Ren, Michael Schrefl, and Markus Stumptner. Behavior Based Integration of Composite Business Processes. In 3rd International Conference on Business Process Management, BPM 2005, September 5-8 2005. LNCS(3649/2005), pages 186-204, Springer-Verlag.
- [GRSS05b] Georg Grossmann, Yikai Ren, Michael Schrefl, and Markus Stumptner. Definition of Business Process Integration Operators for Generalization. In 7th International Conference on Enterprise Information Systems, ICEIS 2005, May 24-28 2005.

- [Ki06] Kindler, E.: On the semantics of EPCs: Resolving the vicious circle. *Data & Knowledge Engineering*, Volume 56, Number 1, January 2006, pages 23-40
- [KNS92] G. Keller, M. Nüttgens, and A. W. Scheer. *Semantische Prozessmodellierung auf der Grundlage Ereignisgesteuerter Prozessketten (EPK)*". Heft 89, Institut für Wirtschaftsinformatik, Saarbrücken, Germany, 1992
- [Krc05] Krcmar, H.: *Informationsmanagement (4. ed.)*. Springer, Berlin, 2005
- [MPZ05] J. Mendling, C. Pérez de Laborda, U. Zdun: Towards an Integrated BPM Schema: Control Flow Heterogeneity of PNML and BPEL4WS. In: K.-D. Althoff, A. Dengel, R. Bergmann, M. Nick, T. Roth-Berghofer, eds.: *Post-Proceedings of (WM 2005)*, *Lecture Notes in Artificial Intelligence 3782*, Kaiserslautern, Germany, pages 570–579, 2005.
- [MS06] Mendling, J.; Simon, C.: *Business Process Design by View Integration*. In: Dustdar, S.; Eder, J.: *BPM 2006 Workshops Proceedings*. *Lecture Notes in Computer Science (LNCS)*, Springer 2006
- [PCS01] Preuner, G.; Conrad, S.; Schrefl, M.: View integration of behavior in object-oriented databases. *Data Knowl. Eng.* 36(2), 2001, pp. 153-183
- [ÖBH91] Österle, H.; Brenner, W.; Hilbers, K.: *Unternehmensführung und Informationssysteme: Der Ansatz des St. Galler Informationssystem-Managements*. Teubner, Stuttgart, 2001
- [RB01] Rahm, E.; Bernstein, P. A.: A Survey of Approaches to Automatic Schema Matching. *VLDB Journal*, 10(4), 2001, pp. 334-350
- [RM05] Rizopoulos, N.; McBrien, P.: A General Approach to the Generation of Conceptual Model Transformations. In: Pastor, O.; e Cunha, J. F. (Ed.): *Proceedings: Advanced Information Systems Engineering. 17th International Conference CAiSE 2005*, *Lecture Notes in Computer Science (LNCS)*, 3520, Porto, Portugal, Springer 2005, pp. 326-341

- [Roh9605] Rohloff, M.: Reference Model and Object Oriented Approach for Business Process Design and Workflow Management. In: Proceedings: Information Systems Conference of New Zealand, 1996, pp. 43-52
- [Sch00] Scheer, A.-W.: ARIS – Business Process Modeling (3rd. edt). Springer, Berlin, 2000
- [Sim06] Simon, C.: Integration of Planning and Production Processes. In: Mathmod 2006, Special Session Petrinets: Current Research Topics and their Application in Traffic Safety and Automation Engineering, Wien, Austria, 2006
- [SL90] Sheth, A.P.; Larson, J.A.: Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. *ACM Comput. Surv.* 22(3), 1990, pp. 183-236
- [SM06] Simon, C., Mendling, J.: Verification of Forbidden Behavior in EPCs. In: Mayr, H. C.; Breu, R.: *Modellierung 2006. Lecture Notes in Informatics (LNI P-82)*, Innsbruck, Austria, 2006, pp. 233-242
- [SS05] Schmitt, I.; Saake, G.: A Comprehensive Database Schema Integration Method Based on the Theory of Formal Concepts. *Acta Informatica*, 41(7-8), 2005, pp. 475-524

Realisierungsformen des Geschäftsprozessmanagements

Eine explorative Klassifikationsanalyse

Tobias Bucher, Robert Winter

Institut für Wirtschaftsinformatik

Universität St. Gallen

CH-9000 St. Gallen

{tobias.bucher, robert.winter}@unisg.ch

Abstract

Das Konzept des Geschäftsprozessmanagements (Business Process Management, BPM) lässt sich nicht auf einen einzigen, allgemeingültigen und in jeder Situation anwendbaren Ansatz zurückführen. Deshalb werden im vorliegenden Beitrag die Ergebnisse einer explorativen Analyse bezüglich der Art und Weise, wie Unternehmen das prozessorientierte Managementkonzept adaptieren, vorgestellt. Dabei werden vier Gestaltungsfaktoren des BPM abgeleitet: das Ausmaß der Performance-Messung, die Professionalisierung des BPM, der Einfluss der Prozessverantwortlichen und die Nutzung etablierter Standards und Methoden. Auf dieser Grundlage wird eine Clusteranalyse durchgeführt, die der Identifikation von vier disjunkten Realisierungsformen des BPM dient. Diese Ansätze lassen sich gemäß den drei Beschreibungsdimensionen Reifegrad, Gestaltungstyp und zugemessene Bedeutung in eine BPM-Typologie-Matrix einordnen.

1 Einführung und Grundlagen

Um Wettbewerbsvorteile zu erlangen oder zu bewahren, investieren Unternehmen viel Zeit und Aufwand in die Entwicklung wohldefinierter Geschäftsprozesse, die effizient und flexibel zugleich sind [SmFi02]. Insbesondere seit dem Paradigmenwechsel der späten 1980er Jahre setzen Unternehmen auf prozessorientiertes Management, um den Ineffizienzen und der mangelnden Agilität des traditionellen, funktionalen Managements entgegenzuwirken [DeMc97]. In den vergangenen beiden Dekaden wurde die prozessorientierte Organisation hauptsächlich in

Zusammenhang mit den Konzepten *Business Process Management (BPM)*, *Business Process Reengineering*, *Total Quality Management* und verwandten Ansätzen diskutiert. Der Erörterung dieser Managementansätze wurde in der betrieblichen Praxis ebenso wie in der wissenschaftlichen Literatur große Beachtung geschenkt, und die Überlegenheit des prozessorientierten Managements ist mittlerweile unbestritten.

In diesem Zusammenhang ist die Erkenntnis, dass das Konzept des Geschäftsprozessmanagements nicht auf einen einzigen, allgemeingültigen und in jeder Situation anwendbaren Ansatz zurückzuführen ist, von elementarer Bedeutung. Es ist nicht zu erwarten, dass ein einziger BPM-Ansatz Lösungen für alle erdenklichen Einsatzszenarien bietet. Eine Reihe von Autoren argumentieren, dass die Entwicklung hin zu unternehmerischer Exzellenz durch prozessorientiertes Management in verschiedenen Stadien verläuft, und dass beinahe jedes Unternehmen eigene Vorgehensweisen und Realisierungsformen entwickelt [ArPM99; BBMS04; HoFu94]. Basierend auf Erfahrungen von Unternehmen mit reifem BPM betonen bspw. ARMISTEAD ET AL. [ArPM99]: “[T]he story of the practical use of business process management in different organizations is one of diversity and of effective outcomes.”

Diese situationsbezogenen Aspekte des BPM sind jedoch weitgehend unerforscht. Bevor systematische, methodische Handlungsempfehlungen zur Implementierung und Weiterentwicklung von BPM gegeben werden können, müssen zunächst fundierte Kenntnisse über die bestehenden BPM-Ansätze gewonnen werden. Hierzu werden BPM-Ansätze zunächst in vorzugsweise homogene Klassen eingeteilt. Diese werden benötigt, um generische Methoden der Entwicklung und Gestaltung von Informations- und Geschäftssystemen so zu adaptieren, dass diese auf die spezifischen BPM-Entwicklungssituationen anwendbar sind.

Der Prozess der Konstruktion und der Adaption generischer Methoden, Techniken und Tools bezüglich einer bestimmten Entwicklungssituation wird unter dem Begriff *Methoden-Engineering* subsumiert [Brin96; Harm97]. Die Adaption solcher generischer Artefakte kann einerseits erreicht werden, indem ein Ausgangsartefakt vor dem Hintergrund einer spezifischen Entwicklungssituation angepasst wird [KaÅg04], oder indem Artefaktfragmente zu einer neuen situationsspezifischen Methode komponiert werden [BrSH98; PuLe96]. Unabhängig davon, welcher Adaptionstyp eingesetzt wird, ist die Identifikation und genaue Beschreibung der Charakteristika von Entwicklungssituationen [BuKl06], in denen die Methode eingesetzt werden soll, unerlässlich. Die Klassifikation der BPM-Realisierungsformen, die im vorliegenden Bei-

trag vorgestellt wird, kann bei der Konfiguration bzw. Komposition von Methoden mit BPM-Bezug als Ausgangspunkt für die Analyse der Situationscharakteristika dienen.

Der vorliegende Beitrag gliedert sich wie folgt: Zunächst gibt Abschnitt 2 einen Überblick über den Stand der Forschung im Bereich Geschäftsprozessmanagement unter besonderer Berücksichtigung von methodischen Ansätzen und Rahmenmodellen für das BPM. Trotz umfangreicher Forschungsaktivitäten im Umfeld des BPM sind Fragestellungen bezüglich der Gestaltungsfaktoren des BPM, der Abfolge der Adaption des BPM-Konzepts sowie der resultierenden BPM-Realisierungsformen bislang weitestgehend unerforscht. Aus diesem Grund werden in Abschnitt 3 die Ergebnisse einer explorativen Analyse zur Klassifikation unterschiedlicher BPM-Ansätze präsentiert. Abschnitt 4 ist der Diskussion der Analyseergebnisse sowie der Systematisierung der ermittelten Realisierungsformen mithilfe der *BPM-Typologie-Matrix* gewidmet. Abschnitt 5 fasst die wesentlichen Erkenntnisse zusammen und bietet einen Ausblick auf anschließende Forschungsaktivitäten.

2 Stand der Forschung und verwandte Arbeiten

Das prozessorientierte Paradigma wird in der Literatur als systematischer, kundenorientierter Managementansatz beschrieben, bei dem die betrieblichen Aktivitäten und Abläufe eines Unternehmens mithilfe von funktionsübergreifendem Teamwork und Delegation von Verantwortung kontinuierlich analysiert, verbessert und überwacht werden [LeDa98; Zair97]. Durch die Betrachtung von Aktivitäts-Ablaufolgen ergibt sich das Bild eines Unternehmens als eine Sequenz funktionaler Prozesse, die organisationsweit vernetzt bzw. verbunden sind [DeMc97].

Seit Mitte der 1990er-Jahre sind umfangreiche Forschungsaktivitäten im Bereich methodischer Ansätze und Rahmenkonzepte des BPM zu beobachten. So entwickelte bspw. HARRINGTON die *Process Breakthrough Methodology*, die 27 Aktivitäten der Adaption des prozessorientierten Paradigmas im Unternehmen vorschlägt [Harr95]. Dieser Beitrag stellt die Grundlage für umfangreiche weitere Forschungsarbeiten dar. Obgleich sich die nachfolgenden Untersuchungen auf die Betrachtung unterschiedlicher Teilaspekte des BPM konzentrierten, besteht allgemeiner Konsens darüber, dass das Geschäftsprozessmanagement vier Komponenten (Definition und Modellierung, Implementierung und Ausführung, Monitoring und Controlling, Optimierung und Weiterentwicklung) umfasst, die eng miteinander in Beziehung stehen und sich wechselseitig bedingen:

Prozessdefinition und -modellierung. Zunächst sind die kritischen Geschäftsprozesse zu identifizieren. Viele Autoren vertreten in diesem Zusammenhang die Ansicht, dass Unternehmen eine geringe Anzahl möglichst umfassender Geschäftsprozesse definieren sollten [Dave93; HaCh93]. SMITH und FINGAR betonen hingegen, dass diese Kernprozesse eine Vielzahl von unterstützenden Abläufen und unterschiedlichen Prozessvarianten erfordern [SmFi02]. Weiterhin ist eine genaue Abstimmung unter den beteiligten Akteuren bezüglich der angestrebten Abläufe unabdingbar [HSFM06; MSHM05]. Die Prozessmodellierung, z.B. auf Grundlage von KUENG und KAWALEK's zielbasiertem Ansatz [KuKa97], ermöglicht u.a. auch die Identifikation von Modifikations- und Verbesserungspotenzialen, selbst über funktionale Grenzen hinweg. Darüber hinaus können Simulationstechniken für die Feinabstimmung und Justierung der Prozessmodelle eingesetzt werden.

Prozessimplementierung und -ausführung. Die Implementierung schließt die Anpassung all derjenigen Aktivitäten, Arbeitsabläufe, Ressourcen und unterstützenden Systeme ein, die erforderlich sind, damit der Prozess reibungslos ablaufen kann. Sowohl die frühzeitige Miteinbeziehung als auch die Schulung von Mitarbeitern werden in diesem Zusammenhang als kritische Erfolgsfaktoren angesehen [MSHM05]. Aufgrund des dynamischen Markt- und Wettbewerbumfelds bieten statisch definierte, schlecht anpassbare Prozesse heute allenfalls unzureichende Unterstützung für die Geschäftstätigkeit. Geschäftsprozesse müssen die situations- und kontextspezifischen Charakteristika berücksichtigend implementiert werden, so dass sie gemäß der sich verändernden Einflussfaktoren flexibel anpassbar sind [WaWa06].

Prozessmonitoring und -controlling. Sowohl automatisiert ablaufende Prozesse als auch solche mit Benutzerinteraktionen müssen zeitnah oder vorzugsweise gar in Echtzeit überwacht werden, um bei Planabweichungen oder Prozessfehlern unmittelbar korrigierende Eingriffe bzw. Massnahmen vornehmen zu können. Darüber hinaus können Messgrößen der Prozessausführung auch zur Unterstützung von Führungsentscheidungen und zur kontinuierlichen Prozessverbesserung herangezogen werden. Nach KAPLAN und NORTON [KaNo92] sollten prozessorientierte Kennzahlen möglichst direkt mit übergeordneten, strategischen Unternehmenszielen in Verbindung stehen. Die Evaluation der Prozessleistung kann mithilfe verschiedener Performance Measurement-Systeme bzw. Ansätzen mit Fokus auf finanzielle und/oder nicht-finanzielle Kennzahlen und quantitative und/oder qualitative Messgrößen wie z.B. dem *Balanced Scorecard* (BSC)-Ansatz [KaNo92], dem *EFQM-Modell für Exzellenz* (EFQM) [EFQM03] oder mit KUENG und KRAHN's *Process Performance Measurement System* [KuKr99] erfolgen. Überraschenderweise

stellten KUENG und KRAHN fest, dass die Messung der Prozessleistungen in vielen Unternehmen auch heute noch eine eher untergeordnete Rolle spielt [KuKr99].

Prozessoptimierung und -weiterentwicklung. Obgleich bereits die ersten Schritte in Richtung Prozessorientierung eine weit reichende Veränderung der Unternehmensgestaltung bedeuten, darf die beständige Weiterentwicklung und Optimierung der Prozessabläufe nicht aus dem Fokus geraten. Dieser Aspekt des prozessorientierten Managementparadigmas wird insbesondere von HARRINGTON [Harr95] und HILL ET AL. [HSFM06] betont. Andere Autoren gehen auf den Aspekt der fortlaufend erforderlichen Prozessoptimierung ein, indem sie BPM als „continuous approach to optimization“ [Zair97] bezeichnen oder indem sie diesen Aspekt der Prozessorientierung dem Konzept des Prozessmonitoring und -controlling zurechnen.

Wie bereits im einleitenden Abschnitt 1 erwähnt, setzen verschiedene Unternehmen den prozessorientierten Managementansatz auf unterschiedliche Art und Weise um. Folglich besteht Grund zur Annahme, dass in der betrieblichen Realität verschiedene Ausprägungen des BPM-Konzepts existieren, die abhängig vom Reifegrad der Prozessorientierung und dem Fokus auf einen oder mehrere der vorgenannten Aspekte sind.

Derzeit wird in der wissenschaftlichen Literatur eine Vielzahl verschiedener Reifegradmodelle für das Geschäftsprozessmanagement vorgeschlagen und diskutiert. Während sich einige dieser Ansätze wie z.B. [Harm06] auf das *Capability Maturity Model* [PCCW83] stützen, kritisieren andere Autoren die in ihren Augen unzulässige Reduktion der Reifegradanalyse auf eine einzige Betrachtungsdimension [DeMc97; MaTM03; PrAr99; RoBr05].

Der vorliegende Beitrag soll zur Ermittlung des Entwicklungsstands bzw. des Umsetzungsgrads des prozessorientierten Managementparadigmas in Unternehmen, zur Identifikation von Gestaltungsfaktoren des BPM und zur Klassifikation von BPM-Realisierungsformen dienen. Hierfür wird ein zweistufiges Vorgehen verfolgt:

1. Bestimmung der Gestaltungsfaktoren des BPM. Es ist unabdingbar, als Ausgangspunkt für die Ermittlung der verschiedenen Realisierungsformen des prozessorientierten Paradigmas zunächst die zugrunde liegenden Gestaltungsfaktoren des BPM zu bestimmen.

2. Identifikation und Klassifikation von BPM-Realisierungsformen. Die real zu beobachtenden Ausprägungen des Geschäftsprozessmanagements unterscheiden sich von Unternehmen zu Unternehmen. Auf Grundlage der vorgenannten Charakteristika können der Entwicklungsstand bzw. die Reife sowie die der konkreten Ausprägung zugrunde liegenden Gestaltungsprinzipien abgeleitet werden. Indem verschiedene Gestaltungsprinzipien zu situationsspezifischen Ansätzen

zusammengefasst werden, können Handlungsmuster identifiziert werden, die darüber Aufschluss geben, wie Unternehmen das prozessorientierte Managementparadigma adaptieren. Diese Handlungsmuster werden im Folgenden als Realisierungsformen des BPM bezeichnet.

Zur Erkenntnisgewinnung bezüglich der vorstehend beschriebenen Fragestellungen wurde eine explorative Untersuchung durchgeführt. Dazu wurde zunächst eine Faktorenanalyse durchgeführt, um die Gestaltungsfaktoren des BPM zu ermitteln. Mithilfe eines clusteranalytischen Algorithmus wurden die untersuchten Fälle anschließend auf Grundlage der berechneten Faktorenwerte in möglichst homogene, wechselseitig disjunkte Klassen unterteilt, welche die unterschiedlichen Realisierungsformen der Geschäftsprozessorientierung widerspiegeln.

3 Explorative Analyse zur Klassifikation von BPM-Ansätzen

Der der explorativen Analyse zugrunde liegende Datensatz wurde mithilfe eines Fragebogens erhoben, der im Rahmen von zwei im Jahr 2005 in Deutschland und der Schweiz durchgeführten Fachtagungen mit BPM-Fokus gestreut wurde. Teilnehmer der Veranstaltungen waren Fach- und Führungskräfte, die im IT-Umfeld oder im Geschäftsprozessmanagement tätig sind.

Der Fragebogen war darauf ausgerichtet, den Umsetzungs- bzw. Reifegrad sowie die Gestaltungsfaktoren des BPM zu erfassen. Bezüglich dieser Umfrageziele wurden Aussagen formuliert, deren Zutreffen bzw. Nichtzutreffen die Befragten für ihr Unternehmen auf einer fünfstufigen Likert-Skala bewerten sollten. Vor dem Einsatz wurden die Fragen im Rahmen eines Pre-Tests sowie mithilfe von Experteninterviews auf Vollständigkeit und Verständlichkeit geprüft.

Insgesamt wurden 47 ausgefüllte Fragebögen retourniert. Blieben ein oder mehrere der in Abschnitt 3.2 beschriebenen Statements unbeantwortet, wurde der jeweilige Fragebogen verworfen. Nach Anwendung dieses Selektionskriteriums verblieben 38 vollständig beantwortete Fragebögen, die in die Analyse einbezogen wurden. Obgleich diese Stichprobe eher klein ist, stellt sie dennoch eine ausreichende empirische Grundlage für eine explorative Analyse dar.

Die in der Untersuchung berücksichtigten Unternehmen sind primär dem Mittelstand und den Großunternehmen aus dem deutschsprachigen Raum zuzurechnen (71% haben mehr als 1000 Mitarbeiter, und weitere 20% haben mehr als 200 Mitarbeiter). Die hauptsächlich vertretenen Branchen sind Banken, Versicherungen und Finanzdienstleister (32%), die produzierende Gewerbe und die Konsumgüterindustrie (15%), die öffentliche Verwaltung und Organisationen

des Gesundheitswesens (11%), Unternehmen, die im IT-nahen Umfeld tätig sind (9%), sowie Energie- und Wasserversorger (ebenfalls 9%).

3.1 Beschreibung des Datensatzes

Zusätzlich zu demografischen Angaben und zu Informationen über den Entwicklungsstand des BPM umfasst der Datensatz 31 Statements, die die Ausgestaltung bzw. Umsetzung des BPM-Konzepts beschreiben. Diese Elemente können fünf Kategorien zugeordnet werden:

Kommunikation des Prozessmanagements. Die gewissenhafte Dokumentation und Kommunikation der Aktivitäten, Arbeitsabläufe, Prozessleistungen und Prozesskennzahlen sind erfolgskritische Faktoren in jeder BPM-Initiative [MSHM05]. Der Datensatz umfasst Informationen darüber, zu welchem Grad die wichtigsten Prozesse grafisch modelliert sind, und ob die Dokumentationen der Abläufe, der Prozessleistungen, der Ziele sowie der finanziellen und nicht-finanziellen Kennzahlen für alle beteiligten Mitarbeiter zugänglich sind.

Rolle der Prozessverantwortlichen. Um den prozessorientierten Managementansatz effektiv im Unternehmen vorantreiben zu können, muss die Rolle der Prozessmanager derart ausgestaltet sein, dass der Wandel hin zur prozessorientierten Denkweise nachhaltig unterstützt wird. Eine dedizierte BPM-Organisationseinheit, die auf einer hinreichend hohen Hierarchieebene angesiedelt ist, eine spezielle Ausbildung für Prozessverantwortliche, die Ausstattung mit ausreichenden fachlichen Entscheidungsbefugnissen in der Prozessgestaltung sowie die aktive Einbindung in allfällige Veränderungsvorhaben tragen zur Erreichung der Ziele bei. Aus diesem Grund umfasst der Datensatz Angaben zu den vorgenannten Variablen.

Prozessgestaltung. Ausgehend von der Tatsache, dass die Definition und Modellierung von Prozessen, die Prozessimplementierung und -ausführung, das Monitoring und Controlling sowie die Weiterentwicklung und Optimierung von Geschäftsprozessen (vgl. Abschnitt 2) auf unterschiedliche Arten adressiert werden können, wurden Informationen zu insgesamt zehn Faktoren der Prozessgestaltung erhoben. Dazu gehören u.a. Angaben bezüglich der Top-down- vs. Bottom-up-Ableitung von Performance-Indikatoren, der Verwendung von Vorgehensmodellen für die Entwicklung des prozessorientierten Performance Managements, der Nutzung von bestehenden Prozessen oder Referenzmodellen als Ausgangspunkt für die Prozessanalyse, der Einbeziehung von Mitarbeitern und Kunden in die Prozessdefinition und das Prozessdesign, die Toolunterstützung und Nutzung von Simulationstechniken zu Design- und Rationalisierungszwecken sowie die Analyse der Kundenzufriedenheit zur Prozessoptimierung.

Performance-Messung in Prozessen. Sowohl Prozessmonitoring als auch Prozesscontrolling spielen bei BPM-Initiativen eine wichtige, gleichwohl häufig vernachlässigte Rolle. Fehlende Performance-Messung auf Prozessebene kann jedoch zu ineffizienten Ressourcenallokationen, erschwerter Identifikation von Planabweichungen, Schwächen in der Prozessausführung und übermäßig aufwändigen Entscheidungsprozessen hinsichtlich korrigierender Eingriffe in die Prozessabläufe führen [KuKr99]. Aus diesem Grund umfasst der Datensatz Informationen darüber, welche Messgrößen erhoben werden (z.B. Ablaufzeiten, Outputs, Prozessleistungen, andere Performance-Indikatoren wie etwa Kosten), und wie diese Messungen stattfinden (in Echtzeit oder zeitverzögert, mit oder ohne Unterstützung eines Workflow-Management-Systems).

Andere Initiativen mit BPM-Bezug. Vielfach sehen sich Unternehmen gezwungen, Qualitätsstandards und/oder etablierten Methodologien der Performance-Messung zu adaptieren. Insbesondere der BSC-Ansatz und das EFQM-Modell sind weit verbreitet. Der Datensatz enthält deshalb Angaben darüber, ob das Unternehmen (bzw. die entsprechende Organisationseinheit) ISO9001-zertifiziert ist, ob und zu welchem Grad EFQM-, Six Sigma- und BSC-Ansätze verwendet werden, und ob diese Zertifizierungen die Gestaltung der Geschäftsprozesse nachhaltig beeinflussen.

3.2 Auswertung und Ergebnisse der Analyse

Um die dominierenden Gestaltungsfaktoren des Geschäftsprozessmanagements zu extrahieren, wurde zunächst eine Faktorenanalyse durchgeführt. Dieses Verfahren kann eingesetzt werden, um eine geringe Anzahl von relevanten, wechselseitig voneinander unabhängigen Faktoren aus einer Vielzahl von Variablen eines Datensatzes zu extrahieren.

Die Technik der Faktorenanalyse geht von der Prämisse aus, dass den einzelnen Variablen gemeinsame Beeinflussungsgrößen (sog. Faktoren) zugrunde liegen. Aus diesem Grund ist ein Datensatz nur dann für die Anwendung dieser Technik geeignet, wenn die Anti-Image-Kovarianz der Variablen, d.h. der Anteil der Varianz jeder einzelnen Variablen, der unabhängig von den Ausprägungen der anderen Variablen des Datensatzes ist, so klein wie möglich ausfällt. Nach DZIUBAN und SHIRKEY ist ein Datensatz für die Anwendung faktorenanalytischer Verfahren geeignet, wenn der Anteil der nicht-diagonalen Elemente der Anti-Image-Kovarianz-Matrix, der ungleich Null ist (d.h. der um mehr als 0.09 von Null abweicht), weniger als 25% beträgt [BEPW06; DzSh74]. Im betrachteten Fall liegt dieser Parameter für einen reduzierten, aus 18 Variablen bestehenden Datensatz bei circa 17.6%. Darüber hinaus liegt das Kaiser-Meyer-Olkin Kriterium, welches die Zusammengehörigkeit zwischen den Variablen ausdrückt und so Auf-

schluss darüber gibt, ob eine Faktorenanalyse sinnvoll durchgeführt werden kann, bei einem Wert 0.753. KAISER und RICE bewerten Ausprägungen, die über einem Schwellwert von 0.7 liegen, als „ziemlich gut“ [KaRi74].

Aufgrund dieser Kriterien wurde eine Faktorenanalyse auf Basis eines auf 18 Variablen reduzierten Datensatzes durchgeführt. Die in die Faktorenanalyse eingehenden und auf die Faktoren ladenden Variablen sind nachfolgend in Tab. 1 bis Tab. 4 dargestellt und erläutert. Mithilfe der Technik der Hauptkomponentenanalyse wurden vier Faktoren extrahiert, die zusammen genommen mehr als 69% der Varianz in den Variablen erklären. Zur Sicherstellung einer besseren Interpretierbarkeit der Ergebnisse wurde die Komponentenmatrix anschließend mittels der Varimax-Methode mit Kaiser-Normalisierung rotiert.

Im Ergebnis finden sich sechs Variablen, die stark auf den ersten Faktor laden. Fortan wird dieser als *Ausmaß der Performance-Messung* bezeichnet (vgl. Tab. 1). Die Notwendigkeit des Monitoring und des Controllings in der Umsetzung des BPM wurde bereits in den vorangehenden Abschnitten betont. Etwa die Hälfte der befragten Unternehmen weist hohe Umsetzungsgrade in den diesem Faktor zugrunde liegenden Variablen auf und misst demnach der Performance-Messung von Geschäftsprozessen hohe Bedeutung bei.

Variable	Beschreibung
1.1	Bei der Prozessgestaltung werden bei Bedarf Simulationen verwendet (z.B. zur Identifikation von Engpässen oder zur Kapazitätsplanung)
1.2	Um die Zufriedenheit mit den Prozessen zu messen, werden Befragungen bei den Prozesskunden durchgeführt
1.3	Im Zuge der Performance-Messung werden Durchlaufzeiten erfasst
1.4	Im Zuge der Performance-Messung werden Outputs bzw. Prozessleistungen gemessen
1.5	Die Messwerte sind ohne unerwünschten Zeitverzug verfügbar
1.6	Ein verwendetes Workflow-Management-System unterstützt bei der Messung von Performance-Indikatoren

Tab. 1: Darstellung der Variablen, die einen signifikanten Einfluss auf Faktor 1 („Ausmaß der Performance-Messung im Unternehmen“) ausüben.

Weiterhin konnten vier Variablen identifiziert werden, die einen signifikanten Einfluss auf den zweiten, im Folgenden als *Professionalisierung des Prozessmanagements* bezeichneten Faktor ausüben (vgl. Tab. 2). Den Ergebnissen der Analyse zufolge ist professionelles BPM von vergleichsweise detaillierter und ausführlicher Dokumentation der Geschäftsprozesse und der Performance-Indikatoren sowie von Bestrebungen, die Prozessverantwortlichen in der Ausübung ihrer Aufgaben so gut wie möglich zu unterstützen, geprägt.

Variable	Beschreibung
2.1	Die Dokumentation der Prozessleistungen und Ziele ist allen beteiligten Mitarbeitern zugänglich (z.B. über Intranet)
2.2	Die Dokumentation der nicht-finanziellen Messwerte ist allen beteiligten Mitarbeitern zugänglich (z.B. über Intranet)
2.3	Es gibt eine Organisationseinheit, die dediziert für strategisches Prozessmanagement zuständig ist (z.B. Entscheidungen über Standards und Werkzeuge und/oder über die organisatorische Einbettung des BPM)
2.4	Es gibt eine dedizierte Ausbildung bzw. Fort- und Weiterbildungsmaßnahmen für Prozessverantwortliche

Tab. 2: Darstellung der Variablen, die einen signifikanten Einfluss auf Faktor 2 („Professionalisierung des Prozessmanagements“) ausüben.

Vier weitere Variablen laden stark auf den dritten Faktor, der den *Einfluss der Prozessverantwortlichen* beschreibt (vgl. Tab. 3). Dieser Faktor betont die hervorgehobene Bedeutung bzw. den Einfluss derjenigen Mitarbeiter, die den Übergang zum prozessorientierten Paradigma anstoßen und verantworten, sowie die wichtige Rolle derjenigen Personen, die für die kontinuierliche Weiterentwicklung und Verbesserung der Prozesse zuständig sind.

Variable	Beschreibung
3.1	Das Prozessmanagement ist auf einer ausreichend hohen Hierarchieebene im Unternehmen angesiedelt, um wirksam werden zu können
3.2	Die Prozessverantwortlichen haben hohes Ansehen in der Organisation
3.3	Die Prozessverantwortlichen haben ausreichende fachliche Entscheidungsbefugnisse, um auf die Prozessgestaltung und -ausführung Einfluss nehmen zu können
3.4	Die Prozessverantwortlichen sind aktiv in Veränderungsvorhaben eingebunden (z.B. Organisationsentwicklung, Softwareprojekte)

Tab. 3: Darstellung der Variablen, die einen signifikanten Einfluss auf Faktor 3 („Einfluss der Prozessverantwortlichen im Unternehmen“) ausüben.

Schließlich konnten die verbleibenden vier Variablen dem vierten Faktor zugerechnet werden. Dieser repräsentiert die *Nutzung etablierter Standards und Methoden* des Prozess- und Qualitätsmanagements sowie der Performance-Messung (vgl. Tab. 4). Unternehmen, die sich auf derartige Standards und Methoden abstützen, nutzen bspw. Vorgehens- bzw. Referenzmodelle wie etwa den BSC-Ansatz zur Gestaltung der Systeme für die prozessorientierte Performance-Messung, greifen auf bewährte Prozessabläufe als Referenz- und Ausgangspunkt für die Prozessanalyse und -gestaltung zurück, und streben ISO-Zertifizierungen sowie die Implementierung von Exzellenz-Ansätzen wie z.B. EFQM an.

Variable	Beschreibung
4.1.	Bei der Gestaltung des Performance Management kommen bestimmte Vorgehensmodelle zum Einsatz
4.2	Bei der Prozessanalyse werden Prozesse von Mitbewerbern oder Referenzmodelle einbezogen
4.3	Das Unternehmen bzw. der Unternehmensteil ist ISO-zertifiziert
4.4	Das Unternehmen bzw. der Unternehmensteil verwendet den EFQM-Ansatz

Tab. 4: Darstellung der Variablen, die einen signifikanten Einfluss auf Faktor 4 („Nutzung etablierter Standards und Methoden des Prozess-, Qualitäts- und Performance-Managements“) ausüben.

Nachfolgend wurden die 38 ausgewählten Fälle auf Grundlage der berechneten Faktorenwerte der vier identifizierten Faktoren mithilfe eines hierarchischen Clusteralgorithmus klassifiziert. Clusteranalytische Ansätze dienen einerseits der Datenreduktion, andererseits auch der Klassifikation von Objekten bzw. der Erkennung von Strukturen und Mustern in komplexen Datensätzen [Ade73].

Agglomerative Clusteralgorithmen wie die hierarchische Clusteranalyse starten mit der feinsten möglichen Partitionierung der Objektgesamtheit, d.h. derjenigen Klassifikation, nach der jedes Objekt ein eigenes Cluster repräsentiert. In aufeinander folgenden Schritten werden jeweils diejenigen beiden Cluster miteinander vereinigt, deren Distanz minimal ist. Aus der Wahl eines Distanzmaßes und eines Fusionierungsalgorithmus ergibt sich eine Vielzahl an Freiheitsgraden in der Ausgestaltung dieser Klassifikationsverfahren. Hierarchisch-agglomerative Verfahren brechen ab, sobald alle Untersuchungsobjekte in einem einzigen Cluster vereinigt sind (größte mögliche Partitionierung). Vom Methodenanwender ist anschließend zu bestimmen, welche Lösung (d.h. welche Clusteranzahl) die für das Klassifikationsproblem am besten geeignete Lösung darstellt. Hilfestellung hierfür bieten heuristische Kriterien wie bspw. Heterogenitätsmaße oder das sog. Elbow-Kriterium [BEPW06; Deim86]; das Hauptaugenmerk sollte jedoch auf der sinnvollen Interpretierbarkeit der Clusterlösung liegen. Hierarchische Verfahren stellen für kleine bis mittelgroße Datensätze effiziente Klassifikationsmöglichkeiten zur Verfügung, weisen jedoch den Nachteil auf, dass einmal getroffene Zuordnungen von Objekten zu Clustern in nachfolgenden Schritten nicht mehr veränderbar sind [EvLL01].

Im vorliegenden Fall erfolgte die Durchführung der hierarchischen Clusteranalyse auf Grundlage des Ward'schen Fusionierungsalgorithmus unter Zuhilfenahme der quadrierten euklidischen Distanz als Abstandsmaß zwischen den Clustern. Das Dendrogramm der Clusteranalyse legt die Vermutung nahe, dass eine Klassifikation der Realisierungsformen des BPM in vier disjunkte

Klassen die mutmaßlich beste Lösung des Clusteringproblems darstellt. Darüber hinaus deutet auch das Elbow-Kriterium auf diese Lösung hin.

4 Interpretation der Ergebnisse und Implikationen für Forschung und Praxis

Abb. 1 stellt die standardisierten arithmetischen Mittelwerte der 18 in die Faktorenanalyse eingeflossenen Variablen für jede der vier im Rahmen der Clusteranalyse identifizierte Realisierungsformen des BPM dar. Zur besseren Übersichtlichkeit wurden die Variablen gemäß ihrer Faktorenzugehörigkeit gruppiert.

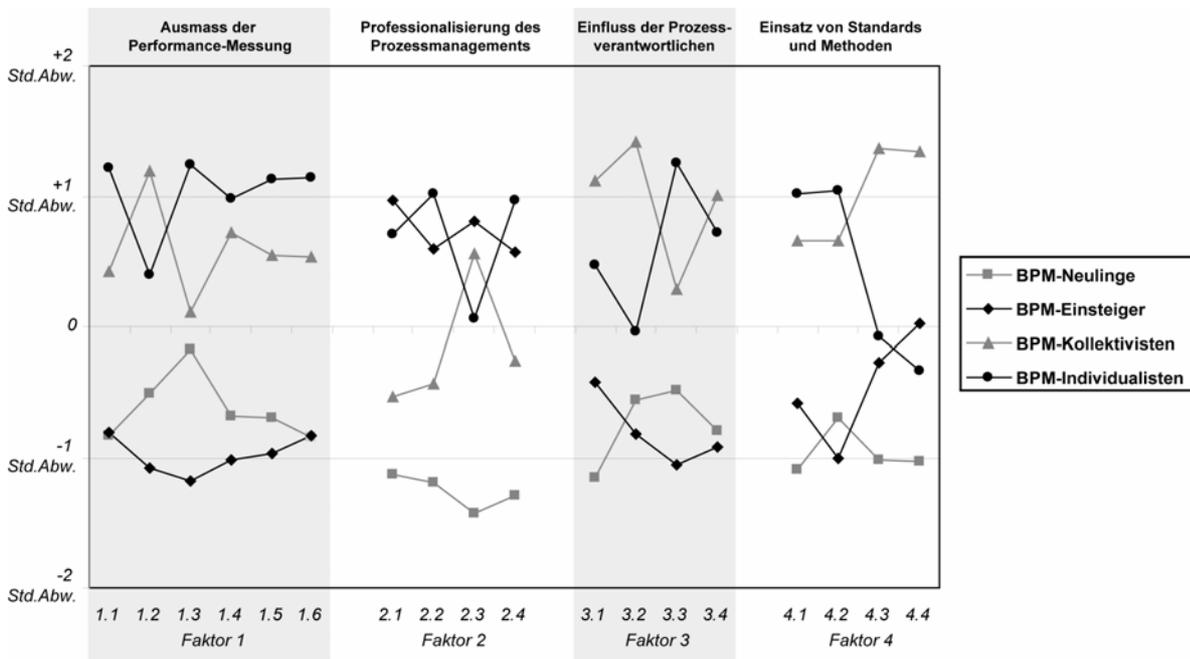


Abb. 1: Profillinien der vier Realisierungsformen des BPM.

Die Profillinien illustrieren eine offenkundige Separation zwischen jeweils zwei BPM-Realisierungsformen. Im Folgenden werden diese Cluster als *BPM-Neulinge* und *BPM-Einsteiger* einerseits sowie als *BPM-Kollektivisten* und *BPM-Individualisten* andererseits bezeichnet. Unternehmen, die in den beiden erstgenannten Gruppen vertreten sind, zeichnen sich durch eher schwach ausgeprägtes Bewusstsein für die Bedeutung der prozessorientierten Performance-Messung, des Einflusses der Prozessverantwortlichen sowie der Nutzung etablierter

Standards und Methoden (Faktoren 1, 3 und 4) aus, wohingegen die in den beiden letztgenannten Gruppen vertretenen Unternehmen signifikant höhere Umsetzungsgrade hinsichtlich der genannten Faktoren aufweisen. Dementsprechend sind die Realisierungsformen der BPM-Kollektivisten und der BPM-Individualisten als „reife“ Formen des Geschäftsprozessmanagements zu bezeichnen. Der Reifegrad des BPM bestimmt sich folglich über drei verschiedene Dimensionen, die durch die Faktoren 1, 3 und 4 beschrieben werden.

BPM-Neulinge sind von einer außergewöhnlich geringen Professionalisierung des Prozessmanagements (Faktor 2) geprägt. Durch ebendieses Kriterium unterscheiden sich BPM-Neulinge von der Gruppe der BPM-Einsteiger. Unternehmen, die sich in dieser Realisierungsform des BPM befinden, zeichnen sich dadurch aus, dass sie dem prozessorientierten Management hohe Bedeutung zumessen und deshalb verstärkte Aufmerksamkeit schenken, z.B. durch die Etablierung dedizierter Aus- und Weiterbildungsprogramme für die Prozessverantwortlichen.

Im Gegensatz zu dieser Klassifikation hinsichtlich der dem BPM zugemessenen Bedeutung vollzieht sich die Unterscheidung zwischen BPM-Kollektivisten und BPM-Individualisten auf der Ebene der konkreten Ausgestaltung des prozessorientierten Managementkonzepts. Unternehmen, welche der erstgenannten Realisierungsform zuzurechnen sind, stützen sich primär auf etablierte Standards sowie auf Vorgehens- und Referenzmodelle, wohingegen Unternehmen, die den BPM-Individualisten zuzurechnen sind, eher die Implementierung eines „maßgeschneiderten“ BPM-Ansatzes verfolgen. Diese beiden als reif zu bezeichnenden Realisierungsformen des BPM unterscheiden sich folglich primär hinsichtlich der Faktoren 2 und 4, d.h. hinsichtlich der Professionalisierung des BPM sowie der Nutzung etablierter Standards und Methoden.

Aufbauend auf dieser empirischen Grundlage ist es möglich, die vier Realisierungsformen des BPM gemäß den drei nachfolgend dargestellten Dimensionen zu klassifizieren:

Reifegrad des Prozessmanagements. Als wichtigstes Differenzierungs- bzw. Klassifikationskriterium der identifizierten BPM-Realisierungsformen ist der Reifegrad des Prozessmanagements anzuführen. Diese Betrachtungsdimension ergibt sich auch aus der offenkundigen Trennung der vier Cluster in die Gruppe der BPM-Neulinge und der BPM-Einsteiger einerseits und diejenige der BPM-Kollektivisten und der BPM-Individualisten andererseits.

Zugemessene Bedeutung. Wenn der Reifegrad der Prozessorientierung eher gering ist, kann davon ausgegangen werden, dass BPM in der Vergangenheit eine eher untergeordnete Rolle im Unternehmen gespielt hat. Die Differenzierung anhand der Aufmerksamkeit bzw. der Bedeutung, die dem prozessorientierten Managementkonzept gegenwärtig beigemessen wird, ermög-

licht die Unterscheidung zwischen den BPM-Neulingen (geringe Aufmerksamkeit) und den BPM-Einsteigern (hohe Aufmerksamkeit). Der letztgenannte Ansatz kann als Übergangsphase hin zum prozessorientierten Paradigma charakterisiert werden.

Gestaltungstyp des BPM. Ist der Reifegrad des BPM hingegen eher hoch (d.h. haben sich die Unternehmen bereits längere Zeit mit dem prozessorientierten Denken beschäftigt), so kann zwischen zwei Gestaltungstypen des BPM unterschieden werden. Die BPM-Kollektivisten stützen sich auf etablierte Standards sowie auf Vorgehens- und Referenzmodelle ab, wohingegen die BPM-Individualisten einen eher „maßgeschneiderten“ Ansatz des Prozessmanagements verfolgen. Aus diesem Grund bieten Unternehmen, die sich als BPM-Individualisten positionieren, den Prozessmanagern breit gefächerte Aus- und Weiterbildungsmöglichkeiten und räumen ihnen weit reichende Entscheidungsbefugnisse bei der Prozessgestaltung und -ausführung ein.

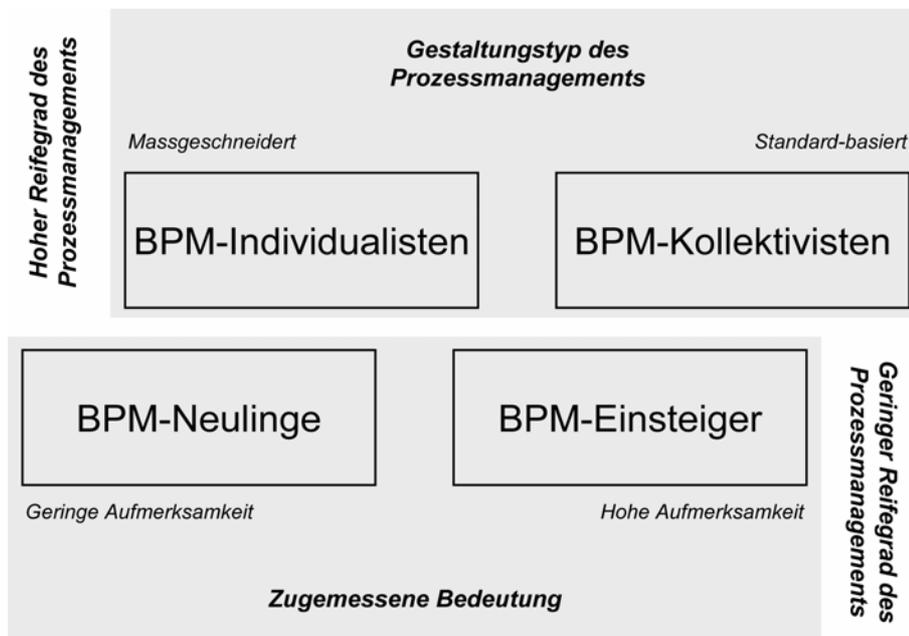


Abb. 2: Typologie-Matrix des BPM.

Die Klassifikation der vier disjunkten Realisierungsformen des BPM hinsichtlich der genannten Betrachtungsdimensionen ist in der so genannten Typologie-Matrix des BPM (vgl. Abb. 2) dargestellt. Die Differenzierung zwischen geringem und hohem Reifegrad des Prozessmanagements ist auf der vertikalen Achse abgetragen, während auf den horizontalen Achsen die dem prozessorientierten Paradigma zugemessene Bedeutung (für die untere Hälfte der Darstellung) bzw. der Gestaltungstyp des Prozessmanagements (für die obere Hälfte) wiedergegeben wird.

5 Zusammenfassung und Ausblick auf weitere Forschungsaktivitäten

Die Ergebnisse der explorativen Klassifikationsanalyse zeigen, dass bezüglich der Herangehensweise, mit der Unternehmen das Konzept des Geschäftsprozessmanagements adaptieren, grundlegende Unterschiede existieren. Ausgehend von vier Gestaltungsfaktoren des BPM konnten vier disjunkte Realisierungsformen des prozessorientierten Managementparadigmas identifiziert und nach drei Beschreibungsdimensionen klassifiziert werden. Diese Forschungsergebnisse werden durch Vorarbeiten im Bereich des Geschäftsprozessmanagements gestützt, die zeigen, dass BPM-Realisierungen meist sehr spezifisch auf das Unternehmen, die jeweilige Entwicklungssituation und die Kontextfaktoren zugeschnitten sind.

Auf Grundlage des vorliegenden Beitrags zur Weiterentwicklung des prozessorientierten Paradigmas können methodische Handlungsempfehlungen für die situationsbezogene Implementierung und Weiterentwicklung des BPM abgeleitet werden. Mithilfe der vorgeschlagenen Klassifikation von BPM-Realisierungsformen ist es möglich, generische Methoden für das Prozessmanagement so zu adaptieren bzw. zu konfigurieren, dass diese im Kontext spezifischer BPM-Entwicklungssituationen anwendbar werden.

Aus dem vorliegenden Beitrag ergeben sich zwei Anknüpfungspunkte für weiterführende Forschungsarbeiten: zum einen sind bestehende Methoden für die Implementierung und Weiterentwicklung des BPM so zu modifizieren bzw. um situative Charakteristika zu erweitern, dass diese auf Grundlage von situations- bzw. kontextspezifischen Faktoren unterschiedlichen BPM-Entwicklungssituationen angepasst werden können [BuKl06]; zum anderen können auf Grundlage der gewonnenen Erkenntnisse weiterführende empirische Studien durchgeführt werden:

Konstruktion situativer Methoden zur Implementierung und Weiterentwicklung des BPM. Die Konstruktion derartiger situations- bzw. kontextspezifischer Methoden sollte auf Grundlage bestehender, etablierter Methoden des BPM erfolgen. Die vier in Abschnitt 2 vorgestellten Komponenten können als Ausgangspunkt und Referenz für die Konstruktion von Methoden dienen, die spezifisch auf ein Szenario bzw. eine Realisierungsform des BPM zugeschnitten sind. Die Klassifikation der BPM-Entwicklungssituationen hinsichtlich des Reifegrads der Prozessorientierung, der dem prozessorientierten Paradigma zugemessenen Bedeutung und des Gestaltungstyps des Prozessmanagements versetzt Methodenentwickler in die Lage, wichtige Implikationen für die Konfiguration bzw. Komposition generischer Methoden bzw. Methodenfragmente abzuleiten. Zweifellos erfasst die im vorliegenden Beitrag vorgeschlagene Klassifikation nicht alle denkbaren Entwicklungssituationen der Informationssystemgestaltung und der

Organisationsentwicklung im Umfeld des BPM. Gleichwohl erlaubt die Differenzierung der vier Realisierungsformen der BPM-Neulinge, der BPM-Einsteiger, der BPM-Kollektivisten und der BPM-Individualisten die Ableitung einiger wertvoller Erkenntnisse über situations- bzw. kontextbezogene Charakteristika, die Methodenentwickler im BPM-Umfeld beherrschen sollten. *Weiterführende empirische Forschung.* In Übereinstimmung mit zahlreichen Reifegradmodellen für das BPM [DeMc97; Harm06; MaTM03; PrAr99; RoBr05] zeigen die Ergebnisse der Klassifikationsanalyse, dass Unternehmen bezüglich der Reife des Prozessmanagements klassifiziert werden können. Jedoch waren Fragen in Bezug auf die Reihenfolge, in der die vier identifizierten Klassen durchlaufen werden, nicht Gegenstand der vorliegend beschriebenen Untersuchung. Interessante Forschungsfragen in diesem Zusammenhang sind beispielsweise: Welche der vier Realisierungsformen BPM-Neulinge, BPM-Einsteiger, BPM-Kollektivisten und BPM-Individualisten sind aufeinander folgende Phasen, die ein Unternehmen im Zuge des Wandels hin zur Prozessorientierung durchläuft? Muss ein Unternehmen all diese Phasen durchlaufen, um eine hohe Reife des Geschäftsprozessmanagements zu erreichen? Oder stehen die Entscheidungsträger eines Unternehmens vielmehr in der Pflicht, nach Durchlaufen der initialen Phasen (BPM-Neulinge und BPM-Einsteiger) eine Entscheidung bezüglich des zu adaptierenden BPM-Gestaltungstyps zu treffen?

Sowohl quantitativ-empirische Methoden der Hypothesenprüfung als auch qualitative Methoden wie z.B. die Fallstudienforschung stellen viel versprechende Möglichkeiten dar, die formulierten Forschungsfragen zu adressieren. Insbesondere bezüglich der Evaluierung möglicher Entwicklungspfade empfiehlt sich die Durchführung einer Zeitreihenanalyse. Gleichzeitig kann auf diese Art und Weise auch der langfristige und nachhaltige Erfolg der verschiedenen Realisierungsformen bzw. Ansätze des BPM untersucht werden. Nicht zuletzt erscheint es auch sinnvoll, die beiden Gestaltungstypen des reifen BPM näher zu betrachten. Es ist anzunehmen, dass zusätzliche Faktoren identifiziert werden können, aufgrund derer weitere Gestaltungsansätze zu unterscheiden sind.

Literatur

[Ande73] Andenberg, M.R.: Cluster Analysis for Applications, Academic Press, New York 1973.

- [ArPM99] Armistead, C.; Pritchard, J.-P.; Machin, S.: Strategic Business Process Management for Organizational Effectiveness, in: Long Range Planning, 32(1999)1, S. 96-106.
- [BEPW06] Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: Multivariate Analysemethoden - Eine anwendungsorientierte Einführung, Springer, Berlin et al. 2006.
- [BBMS04] Balzarova, M.A.; Bamber, C.J.; McCambridge, S.; Sharp, J.M.: Key Success Factors in Implementation of Process-based Management - A UK Housing Association Experience, in: Business Process Management Journal, 10(2004)4, S. 387-399.
- [Brin96] Brinkkemper, S.: Method Engineering - Engineering of Information Systems Development Methods and Tools, in: Information and Software Technology, 38(1996), S. 275-280.
- [BrSH98] Brinkkemper, S.; Saeki, M.; Harmsen, F.: Assembly Techniques for Method Engineering, in: Pernici, B.; Thanos, C. (Hrsg.): Advanced Information Systems Engineering - Proceedings of the 10th International Conference CAiSE'98, Springer, Berlin et al. 1998, S. 381-400.
- [BuKl06] Bucher, T.; Klesse, M.: Contextual Method Engineering: Working Paper, Institute of Information Management, University of St. Gallen, St. Gallen 2006.
- [Dave93] Davenport, T.H.: Process Innovation - Reengineering Work through Information Technology, Harvard Business School Press, Boston (MA) 1993.
- [Deim86] Deimer, R.: Unscharfe Clusteranalysemethoden - Eine problemorientierte Darstellung zur unscharfen Klassifikation gemischter Daten, Schulz-Kirchner, Idstein 1986.
- [DeMc97] DeToro, I.; McCabe, T.: How to Stay Flexible and Elude Fads, in: Quality Progress, 30(1997)3, S. 55-60.
- [DzSh74] Dziuban, C.D.; Shirkey, E.C.: When is a Correlation Matrix Appropriate for Factor Analysis?, in: Psychological Bulletin, 81(1974)6, S. 358-361.
- [EFQM03] EFQM: The Fundamental Concepts of Excellence, <http://www.efqm.org> 2003.
- [EvLL01] Everitt, B.S.; Landau, S.; Leese, M.: Cluster Analysis, Arnold, London 2001.
- [HaCh93] Hammer, M.; Champy, J.: Reengineering the Corporation - A Manifesto for Business Revolution, Harper Business, New York (NY) 1993.
- [Harm06] Harmon, P.: BPM Methodologies and Process Maturity, www.bpmtrends.com 2006.
- [Harm97] Harmsen, F.: Situational Method Engineering, Moret Ernst & Young Management Consultants, Utrecht 1997.
- [Harr95] Harrington, H.J.: Total Improvement Management - The Next Generation in Performance Improvement, McGraw-Hill, New York (NY) et al. 1995.
- [HSFM06] Hill, J.B.; Sinur, J.; Flint, D.; Melenovsky, M.J.: Gartner's Position on Business Process Management 2006, <http://www.gartner.com> 2006.
- [HoFu94] Ho, S.K.M.; Fung, C.K.H.: Developing a TQM Excellence Model, in: The TQM Magazine, 6(1994)6, S. 24-30.

- [KaRi74] Kaiser, H.F.; Rice, J.: Little Jiffy, Mark IV, in: Educational and Psychological Measurement, 34(1974), S. 111-117.
- [KaNo92] Kaplan, R.S.; Norton, D.P.: The Balanced Scorecard - Measures that Drive Performance, in: Harvard Business Review, 70(1992)1, S. 71-79.
- [KaÅg04] Karlsson, F.; Ågerfalk, P.J.: Method Configuration - Adapting to Situational Characteristics while Creating Reusable Assets, in: Information and Software Technology, 46(2004), S. 619-633.
- [KuKa97] Kueng, P.; Kawalek, P.: Goal-based Business Process Models - Creation and Evaluation, in: Business Process Management Journal, 3(1997)1, S. 17-38.
- [KuKr99] Kueng, P.; Krahn, A.J.W.: Building a Process Performance Measurement System - Some Early Experiences, in: Journal of Scientific and Industrial Research, 58(1999)3/4, S. 149-159.
- [LeDa98] Lee, R.G.; Dale, B.G.: Business Process Management - A Review and Evaluation, in: Business Process Management Journal, 4(1998)3, S. 214-225.
- [MaTM03] Maull, R.S.; Tranfield, D.R.; Maull, W.: Factors Characterising the Maturity of BPR Programmes, in: International Journal of Operations & Production Management, 23(2003)6, S. 596-624.
- [MSHM05] Melenovsky, M.J.; Sinur, J.; Hill, J.B.; McCoy, D.W.: Business Process Management - Preparing for the Process-Managed Organization, <http://www.gartner.com> 2005.
- [PCCW83] Paulk, M.C.; Curtis, B.; Chrissis, M.B.; Weber, C.V.: Capability Maturity Model for Software, Version 1.1 (Technical Report CMU/SEI-93-TR-024), Software Engineering Institute, Carnegie Mellon University, Pittsburgh (PA) 1993.
- [PrAr99] Pritchard, J.-P.; Armistead, C.: Business Process Management - Lessons Learned from European Business, in: Business Process Management Journal, 5(1999)1, S. 10-32.
- [PuLe96] Punter, T.; Lemmen, K.: The MEMA-Model - Towards a New Approach for Method Engineering, in: Information and Software Technology, 38(1996), S. 295-305.
- [RoBr05] Rosemann, M.; de Bruin, T.: Towards a Business Process Management Maturity Model, in: Bartmann, D. et al. (Hrsg.): Proceedings of the Thirteenth European Conference on Information Systems, Regensburg 2005.
- [SmFi02] Smith, H.; Fingar, P.: Business Process Management - The Third Wave, Meghan-Kiffer Press, Tampa (FL) 2002.
- [WaWa06] Wang, M.; Wang, H.: From Process Logic to Business Logic - A Cognitive Approach to Business Process Management, in: Information & Management, 43(2006), S. 179-193.
- [Zair97] Zairi, M.: Business Process Management - A Boundaryless Approach to Modern Competitiveness, in: Business Process Management Journal, 3(1997)1, S. 64-80.

Individualisierung von Prozessen und E-Services mithilfe von Case Based Reasoning

Günter Schicker, Carolin Kaiser, Freimut Bodendorf

Lehrstuhl für Wirtschaftsinformatik II

Universität Erlangen-Nürnberg

Lange Gasse 20

90403 Nürnberg

bodendorf@wiso.uni-erlangen.de

Abstract

Mit steigender Wettbewerbsdynamik, kürzeren Strategie- und Produktlebenszyklen sowie zunehmender Leistungsindividualisierung steigen die Anforderungen an die Flexibilisierung der Prozess- und IT-Landschaft, um eine effiziente Implementierung neuer Geschäftsstrategien sicherzustellen. Gleichzeitig ermöglichen neue Technologien und Standards wie Web Services und Serviceorientierte Architekturen (SOA) die IuK-technische Umsetzung flexibler Wertschöpfungssysteme. Es wird ein Konzept zur Individualisierung von Prozessen und E-Services mittels Case Based Reasoning vorgestellt. Anhand von Prozesskontext-Informationen werden Vorschläge für individualisierte Prozesse und unterstützende E-Services automatisch erstellt. Der Ansatz erlaubt die effiziente fachliche Modellierung und E-Service-Unterstützung individueller Prozesse auf Basis einer Serviceorientierten Architektur.

1 Motivation

In den vergangenen Jahren dominierten in den Prozessmanagement-Abteilungen von Unternehmen Diskussionen über die Standardisierung und Harmonisierung von Prozessen. Aufgrund des internen Kostendrucks wurden Anstrengungen unternommen, um Synergien auf fachlicher Ebene z.B. durch Organisation von Shared Services oder Outsourcing von Unterstützungsprozessen zu realisieren [FeSe05, S.14ff]. Grundlage hierfür war die Definition

geeigneter Prozesse und eine an den neu gestalteten Prozessen ausgerichtete IuK-Infrastruktur. Neben fachlichen Treibern wurde die Standardisierung von Prozessen auch vorangetrieben, um die Konsolidierung der Applikationslandschaft zu befördern. Eine Vielzahl heterogener Anwendungen (z.B. unterschiedliche Softwareprodukte oder Systeminstanzen) für vergleichbare Einsatzbereiche führte zu hohen Lizenz-, Wartungs- und Betriebskosten und erforderte – als Basis einer Applikationsharmonisierung – zunächst ein gemeinsames Prozessverständnis sowie abgestimmte Prozesse.

Die Standardisierung von Prozessen stößt jedoch an Grenzen. Der Hauptgrund hierfür liegt in der strategiekonformen Ausrichtung der Prozesslandschaft, die einer rein effizienzorientierten Ausrichtung – zumindest teilweise - entgegen läuft. Um sich vom Wettbewerb zu differenzieren, müssen sich Kernprozesse von Wettbewerbern unterscheiden. Prozesse setzen Strategie um. Die IT-Landschaft setzt Prozesse um. Diese zentralen Aussagen von Porter geben die logische Abhängigkeit vor: Strategie – Prozesse – IT [Port96]. Rückkopplungen sind dabei möglich, z.B. können IT-Innovationen wie RFID Auswirkungen auf Prozesse und Strategie haben. Prozesse sind daher gemäß den strategischen Vorgaben zu modellieren und zu implementieren, um die Unternehmensziele zu erreichen. Aus diesem Grund ist sowohl die Adaption von Referenzprozessmodellen, die Übertragung von Best Practices auf das eigene Unternehmen als auch die Standardisierung von Prozessen mit Bedacht vorzunehmen. Gleichzeitig steigt der Bedarf an Flexibilisierung oder sogar Individualisierung von Prozessen:

- Die Vernetzung nimmt zu und verstärkt Arbeitsteilung und Spezialisierung, um strategische Wertschöpfungsnetze oder Virtuelle Unternehmen zu gestalten [Flei01, S. 17ff].
- Fusionen und Akquisitionen erfordern eine flexible Anpassung der Prozess- und IT-Landschaft, um Synergien im Zuge von Transitionen zügig zu realisieren.
- Die Intensivierung des Wettbewerbs sowie kürzere Produktlebenszyklen führen in dynamischen Branchen (z.B. Consumer Electronics) zu kürzeren Strategiezyklen und zur Notwendigkeit, Strategien prozessseitig zügig umzusetzen [Moor95].
- Die zunehmende Produkt- und Dienstleistungsindividualisierung als strategisches Wettbewerbsinstrument erfordert eine Flexibilisierung der Leistungserstellungsprozesse.

Die Flexibilisierung der Prozesse ist zunehmend wettbewerbskritisch.¹ Die IT-Architektur und die zugehörige Applikationslandschaft ist daher auf eine effiziente und zugleich flexible Unterstützung der Prozesse auszurichten. Dies wird u.a. durch Serviceorientierte Architekturen (SOA) technisch ermöglicht [KrBS04]. Standards zur dynamischen Orchestrierung elektronischer Dienste (Web Services) zur Prozessunterstützung (z.B. BPML, BPEL) etablieren sich, kommerzielle Softwareprodukte zur technischen Konfiguration und Ausführung von Web-Service-basierten Workflows und Orchestrierungen gewinnen an Reife [Chan06, S. 157ff].

Allerdings bleiben zahlreiche Herausforderungen in Bezug auf die fachliche Gestaltung flexibler Prozesse bestehen. Im Folgenden wird ein Ansatz zur Spezifikation individueller Prozesse sowie zur Konfiguration unterstützender E-Services mithilfe von Case Based Reasoning vorgestellt. Ziel ist es, das Design individueller Prozesse zumindest teilweise zu automatisieren und darüber hinaus im Rahmen der Ausführung der Prozessbeschreibung zur richtigen Zeit die richtigen E-Services² für die Prozessbeteiligten zur Verfügung zu stellen.

2 Bestehende Individualisierungsansätze

Zur Individualisierung von Prozessen und E-Services existieren verschiedene Konzepte, die nach Art des eingesetzten Instruments unterschieden werden können.

Beim klassischen Ansatz der Prozessindividualisierung wird jeder Prozess für den jeweiligen Kontext passend unter Verwendung von Richtlinien neu modelliert [Lang 97, S. 2]. Diese permanente Neugestaltung von individuellen Prozessen ist sehr zeitintensiv und verhindert die Wiederverwendung von Erfahrungen aus früheren Prozessen [Rupp02, S. 2]. Im Gegensatz dazu ermöglichen die Modellierungsansätze mithilfe von Referenzmodellen [Brock03, S. 31ff], Prozessskeletten [Remme97, S. 114ff] und Prozessbausteinen [Lang97, S. 4ff] die Nutzung von Prozesswissen. Problematisch bei allen Ansätzen ist jedoch die aufwändige Suche und Anpassung an den aktuellen Kontext.

Die Methoden zur Individualisierung von E-Services sind den Verfahren der Prozessindividualisierung sehr ähnlich. Analog zur Prozessindividualisierung besteht der

¹ Flexibilität ist die Fähigkeit, sich auf geänderte Anforderungen und Gegebenheiten der Umwelt einstellen zu können. Im Prozesskontext bezieht sich die Änderungsfähigkeit auf Ablauforganisation (Struktur), Geschäftsregeln, Prozessakteure sowie E-Services und deren Provider zur Design- und Laufzeit von Prozessen.

² E-Services sind Softwarekomponenten, die aus Servicekontrakt, Schnittstellen und Implementierung bestehen und fachliche Funktionen auf grobgranularer Ebene kapseln (Informations-/Anwendungsdienste) [KrBS04].

klassische Ansatz der E-Service-Individualisierung darin, für jeden Kontext einen E-Service anhand vorgegebener Programmierrichtlinien neu zu erstellen. Diese Vorgehensweise ist jedoch sehr aufwändig und erlaubt keine Wiederverwendung von E-Service-Wissen. Eine Wiederverwendung ermöglichen die Ansätze zur Erzeugung eines E-Services durch den Einsatz von Entwurfsmustern, Frameworks und Bibliotheken.

Abstrakt betrachtet besteht die Vorgehensweise aller Ansätze zur Prozess- und E-Service-Individualisierung aus der Suche eines geeigneten Instruments und der daran anschließenden Anpassung an den individuellen Kontext. Beide Schritte führen bei manueller Durchführung oft zu einem hohen Zeitaufwand. Intelligente Ansätze zur Individualisierung beschäftigen sich deshalb mit der Entwicklung von automatischen Such- und Adaptionenverfahren. Die Systeme „WorkBrain“ [WaWT97, S. 3ff] und „Flexware“ [WaWe97, S. 52f] ermöglichen es zum Beispiel mittels eines fallbasierten Retrieval-Systems bestehende Prozessmodelle zu finden, die einen zur aktuellen Situation ähnlichen Prozesskontext aufweisen. Der „ARIS Process Generator“ [HaRS99, S. 21ff] und das Konzept zur projektspezifischen Individualisierung [Rupp02, S. 67ff] erlauben die automatische Adaption von Prozessmodellen an den Prozesskontext mithilfe von benutzerdefinierten Regeln. Problematisch bei diesen Ansätzen ist, dass nur einer der beiden Schritte, Suche oder Adaption, automatisiert wird, während der andere weiterhin manuell durchzuführen ist. Ziel dieses Projektes ist es daher ein System zu entwickeln, das sowohl intelligente Automatismen zur Suche bestehender Prozessmodelle und ihrer E-Services als auch zur Adaption gefundener Prozessmodelle und ihrer E-Services an den aktuellen individuellen Kontext bietet. Dabei kommt die Methodik der Künstlichen Intelligenz Case Based Reasoning zum Einsatz.

3 Case Based Reasoning

Case Based Reasoning (CBR) ermöglicht das Lösen von Problemen auf Basis von Erfahrungswissen, das in Form von Fällen in einer Fallbasis abgelegt ist [MaTS01, S. 1]. Ein Fall besteht aus zwei grundsätzlichen Teilen, einer Problembeschreibung und der dazugehörigen Lösung. Um ein neues Problem zu lösen, wird in der Fallbasis das dazu ähnlichste Problem gesucht und dessen Lösung wieder verwendet [MaTS01, S. 1]. Im Rahmen der Prozess- und E-Service-Individualisierung können CBR-Systeme dazu verwendet werden, um zu einem gegebenen Prozess- und E-Service-Kontext (Problembeschreibung) den

ähnlichsten in der Fallbasis enthaltenen Prozess- und E-Service-Kontext aufzufinden und dessen Prozess mit E-Services (Lösung) wieder zu verwenden und an den neuen Kontext anzupassen. CBR erfüllt damit die Grundanforderungen der automatischen Selektion und Adaption.

Die Funktionsweise eines CBR-Systems kann als zyklisches Phasenmodell [AaPI94, S. 46] dargestellt werden. In der Retrieve-Phase wird zu dem gegebenen neuen Kontext der Prozess mit E-Services in der Fallbasis gesucht, der den dazu ähnlichsten Kontext enthält. Dieser dient in der Reuse-Phase als Basis zur Lösungsfindung. Im Rahmen der Revise-Phase erfolgt die tatsächliche Ausführung des neuen Prozesses und seiner E-Services. In Abhängigkeit des Erfolges erhält er eine Bewertung, die innerhalb der Retain-Phase über seine Aufnahme in die Fallbasis entscheidet. Abb. 1 veranschaulicht den CBR-Zyklus.

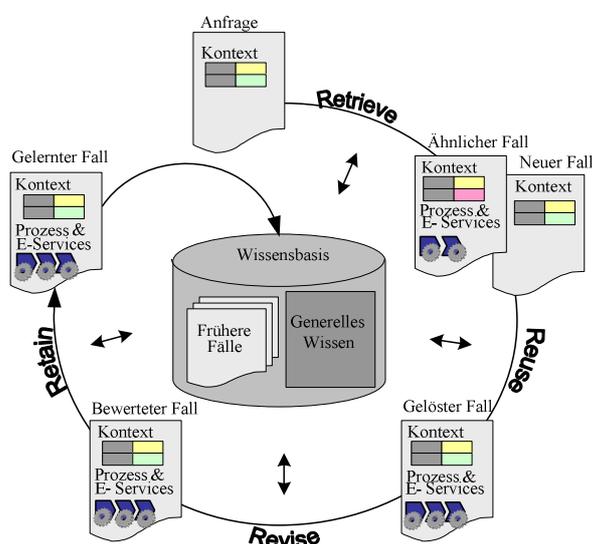


Abb. 1: CBR-Zyklus nach [AaPI94, S. 46]

Bei der Entwicklung eines CBR-Systems zur Prozess- und E-Service-Individualisierung sind eine geeignete Form der Repräsentation der Fälle sowie passende Verfahren zur Ähnlichkeitsbestimmung, Suche, Adaption und Wartung zu bestimmen.

4 Prozess- und E-Service-Individualisierung mittels Case Based Reasoning

4.1 Definition und Repräsentation von Kontext, Prozess und E-Services

Ein Fall setzt sich aus dem Prozess- und E-Service-Kontext sowie dem ausgeführten Prozess und seinen E-Services zusammen. Der Prozess- und E-Service-Kontext spiegelt die Rahmenbe-

dingungen wider, unter denen der Prozess und seine E-Services sinnvoll eingesetzt werden können (z.B. Branche, Kundenanforderungen, strategisches Umfeld, Phase im Produktlebenszyklus, Prozesskomplexität). Die Repräsentation des Kontextes erfolgt mit Hilfe von Attribut-Werte-Vektoren [Rich03, S. 412]. Hierbei wird der Prozess- und E-Service-Kontext durch eine Anzahl n von Attributen A_1, A_2, \dots, A_n näher charakterisiert. Ein konkreter Prozess- und E-Service-Kontext P wird durch den Vektor der Attributwerte a_1, a_2, \dots, a_n aus den Wertebereichen W_1, W_2, \dots, W_n spezifiziert: $P = (a_1, a_2, \dots, a_n) \in W_1 \times W_2 \times \dots \times W_n$. Hierbei können metrische, ordinale und nominale Attribute verwendet werden. Um unbekannte Attributwerte zuzulassen, wird der Wertebereich jedes Attributs um die Ausprägung „unbekannt“ erweitert.

Auf Grund ihrer komplexen Struktur wird für Prozesse und E-Services die ausdrucksstarke objektorientierte Repräsentationsform herangezogen. Hierbei werden die Prozesse und E-Services mit ihren Elementen durch Klassen näher charakterisiert. Ein konkreter Prozess mit E-Services besteht aus einer Menge von Objektinstanzen dieser Klassen. Das Klassendiagramm in Abb. 2 zeigt auf, wie Prozesse und E-Services definiert sind.

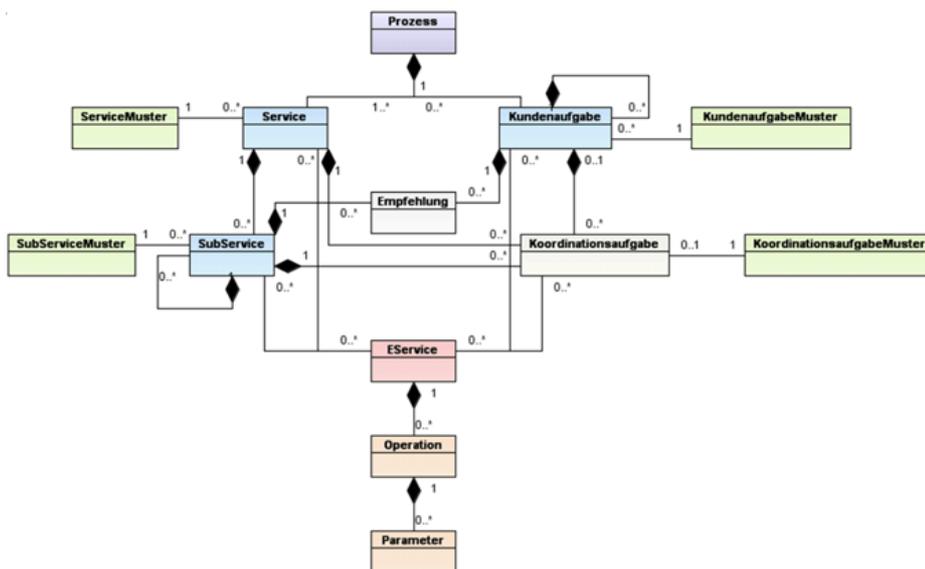


Abb. 2: Prozess- und E-Service-Modell

Ein Prozess besteht aus einem oder mehreren Services, die von einem Serviceanbieter für einen Kunden erbracht werden. Er kann verschiedene Aufgaben beinhalten, die der Kunde selbst zu übernehmen hat. Jeder Service kann mehrere Sub-Services umfassen. Kundenaufgaben und Sub-Services können sich selbst als Element enthalten und ermöglichen somit eine rekursive hierarchische Schachtelung. Reihenfolgebedingte Beziehungen werden nur auf Serviceebene modelliert. Die Prozesselemente Service, Kundenaufgabe und Sub-Service können weiterhin

Empfehlungen und Koordinationsaufgaben beinhalten. Alle Services, Kundenaufgaben, Sub-Services und Koordinationsaufgaben beruhen auf Mustern, die allgemeine Informationen zu den Prozesselementen enthalten (z.B. Name, Beschreibung, Funktion, Ziel). Die Ausführung von Services, Kundenaufgaben, Sub-Services und Koordinationsaufgaben kann durch E-Services unterstützt werden. Jeder E-Service kann mehrere Funktionalitäten in Form von Operationen umfassen. Operationen können wiederum mehrere Parameter enthalten, die ihre Funktionalität näher spezifizieren.

4.2 Bestimmung ähnlicher Prozess- und E-Service-Kontexte

Zielsetzung der CBR-Retrieve-Phase ist es zu einem neuen Kontext den Prozess mit E-Services in der Fallbasis zu ermitteln, der den dazu ähnlichsten Kontext enthält. Zur Bestimmung der Ähnlichkeit zweier Kontexte kommen lokale und globale Ähnlichkeitsmaße zum Einsatz. Während lokale Ähnlichkeitsmaße die Ähnlichkeit zwischen einzelnen Attributausprägungen messen, ermitteln globale die Ähnlichkeit zwischen zwei gesamten Kontexten durch Aggregation der lokalen Ähnlichkeiten [Stah03, S. 50ff; Wess95, S. 125ff].

Lokale Ähnlichkeitsmaße werden in Abhängigkeit des Skalenniveaus des jeweiligen Attributs definiert. Zur lokalen Ähnlichkeitsmessung nominaler Attribute werden Ähnlichkeitstabellen [Goos96, S. 90f] herangezogen. Diese enthalten für jede Kombination aus Attributwerten a_i des Anfragefalles a und des Vergleichfalles f aus der Fallbasis einen Ähnlichkeitswert $x_{ij} \in [0;1]$.

a/f	a_1	a_2	...	a_n
a_1	I	x_{12}		x_{1n}
a_2	x_{21}	I		x_{2n}
...				
a_n	x_{n1}	x_{n2}		I

Abb. 3: Ähnlichkeitstabelle

Bei metrischen Attributen erfolgt die lokale Ähnlichkeitsbestimmung unter Verwendung von Distanzmaßen, die die Differenz zweier Attributwerte berücksichtigen. Zur lokalen Ähnlichkeitsbestimmung metrischer Attribute stehen verschiedene Maße m zur Wahl. Als Maße m können schwellenwertbasierte, lineare, exponentiale und sigmoide Funktionen gewählt werden

[Stah03, S. 54ff]. Nachstehende Abbildung zeigt die Graphen dieser Funktionen für den Wertebereich $f > a$ auf.

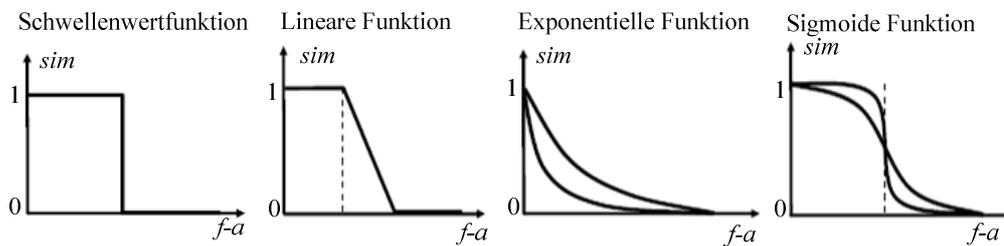


Abb. 4: Ähnlichkeitsfunktionen nach [Stah03, S.56]

Zur Ähnlichkeitsmessung ordinaler Attribute existieren keine eigenen Ähnlichkeitsmaße. Es können jedoch die Maße metrischer Attribute verwendet werden [Goos96, S. 91]. Hierbei müssen die Ausprägungen des ordinalen Attributs auf Zahlenwerte abgebildet werden, über die die Ähnlichkeitsmessung erfolgt. Bei unbekanntem Attributwert erfolgt die Ähnlichkeitsmessung für Attribute aller Skalenniveaus durch eine Schätzung.

Als globales Ähnlichkeitsmaß wird die gewichtete Summe der lokalen Ähnlichkeiten herangezogen. Hierbei kann der Domänenexperte (z.B. Fachprozess-Owner) jedem Attribut ein Gewicht zuweisen, welches die Relevanz des Attributs widerspiegelt.

Die manuelle Definition der Ähnlichkeitsmaße ist zeitaufwändig und erfordert ein hohes Maß an explizitem Domänenwissen. Daher kommt ein Lernalgorithmus zum Einsatz, der die Definition des globalen Ähnlichkeitsmaßes vereinfacht und das Maß optimiert. Mit Hilfe des Lernalgorithmus werden die initialen uniformen oder benutzerdefinierten Gewichte so modifiziert, dass das globale Ähnlichkeitsmaß das Auffinden von nützlichen und adaptierbaren Prozessen und E-Services ermöglicht. Als Lernalgorithmus wird das Gradientenabstiegsverfahren eingesetzt [Stah03, S. 105ff]. Dieses benötigt Trainingsdatensätze, die Informationen vom Benutzer des CBR-Systems über die Rangfolge der Nützlichkeit adaptierten Prozesse und E-Services zu einem gegebenen Problem enthalten. Diese Informationen können während der Benutzung des Systems erfasst werden. Auf Basis des Vergleichs der vom Benutzer angegebenen Rangfolge der adaptierten Lösungen und der durch das Ähnlichkeitsmaß berechneten Rangfolge ist eine Fehlerfunktion definiert, die das Gradientenabstiegsverfahren durch Modifikation der Gewichte minimiert.

4.3 Suche ähnlicher Prozesse und E-Services-Kontexte

Ziel der Suchverfahren, die im Rahmen der CBR-Retrieve-Phase eingesetzt werden, ist es, zu einer Anfrage n Fälle aus der Fallbasis zu finden, die bezüglich des definierten Ähnlichkeitsmaßes die höchste Ähnlichkeit aufweisen [Wess95, S. 159]. Die Suche der ähnlichsten Fälle kann mittels verschiedener Strategien realisiert werden, die alle einen Kompromiss aus Qualität der gefundenen Fälle, Zeitverhalten und Flexibilität erfordern. Zur Prozess- und E-Service-Individualisierung werden zwei alternative Suchverfahren, die sequenzielle Suche und die wissensarme Indexierung mittels erweitertem k-d-Baum [Wess95, S. 209ff], bereitgestellt. Beide Verfahren finden mit Sicherheit die ähnlichsten Fälle in der Fallbasis, unterscheiden sich aber in Effizienz und Flexibilität. Während die sequenzielle Suche beliebige Ad-hoc-Anfragen erlaubt, ermöglicht der erweiterte k-d-Baum eine schnelle standardisierte Suche. Bei der sequenziellen Suche wird die Ähnlichkeit zwischen der Anfrage und jedem Fall der Fallbasis nacheinander mit Hilfe des definierten Ähnlichkeitsmaßes ermittelt und die Fälle werden mit aufsteigender Ähnlichkeit zur Anfrage geordnet [Wess95, S. 163ff].

Die Grundidee des erweiterten k-dimensionalen Baumes besteht darin, die Fallbasis mithilfe eines Baumes nach k Dimensionen zu partitionieren [Wess95, S. 180f]. Die Dimensionen entsprechen dabei den Attributen des Prozess- und E-Service-Kontextes [Goos96, S. 69f]. Abb. 5 zeigt einen erweiterten 2-d-Baum.

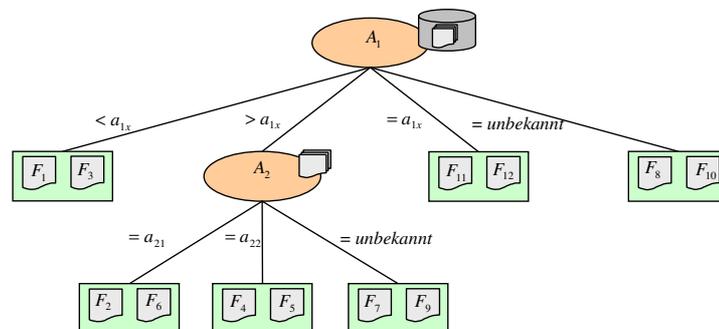


Abb. 5: Erweiterter 2-d-Baum

Der erweiterte k-d-Baum wird automatisch generiert [Wess95, S. 198ff]. Hierbei wird die Baumstruktur so gestaltet, dass zur Ermittlung der ähnlichsten Fälle nur für wenige Fälle die Ähnlichkeit zur Anfrage berechnet werden muss und somit eine effiziente Suche möglich ist [Goos96, S. 71]. Bei der Suche im erweiterten k-d-Baum [Wess95, S. 188, S. 234f] werden zunächst ausgehend von dem Wurzelknoten die Pfade im Baum verfolgt, deren Kantenbedingungen die Attributwerte der Anfrage erfüllen, bis ein Blattknoten erreicht ist. Die

im Blattknoten enthaltenen Fälle werden gemäß ihrer Ähnlichkeit zur Anfrage geordnet. Mithilfe der Attribut-Werte-Bedingungen der Knoten wird getestet, ob noch ähnlichere Fälle existieren, und die Suche gegebenenfalls rekursiv am Vaterknoten fortgesetzt.

4.4 Adaption bestehender Prozesse und E-Services

Ziel der Adaption, die im Rahmen der CBR-Reuse-Phase stattfindet, ist es, ein aktuelles Problem mithilfe des in der CBR-Retrieve-Phase gefundenen ähnlichsten Falles zu lösen. Die Lösung des ähnlichsten Falles dient dabei als Basis für die Bestimmung der Lösung des aktuellen Problems [FLMN99, S. 105].

Die Adaption verläuft, abhängig von dem Ziel der Anfrage, in verschiedenen Schritten. Als Anfrage-Ziele können die Erstellung eines neuen individuellen Prozesses mit E-Services und die individuelle Erweiterung eines bestehenden Prozesses mit seinen E-Services unterschieden werden. Bei der Neuerstellung wird eine Kopie des Prozesses und der E-Services des ähnlichen Falles als Grundlage für die Anfrage verwendet. Im Falle einer Erweiterung wird aus dem Prozess und den E-Services der Anfrage und des ähnlichen Falles ein neuer Prozess mit E-Services zusammengesetzt. Anschließend erfolgt durch die substitutionale und strukturelle Adaption sowie die Konsistenzsicherung eine Anpassung des kopierten bzw. konfigurierten Prozesses und seiner E-Services an den aktuellen Kontext. Nachstehende Grafik veranschaulicht den Ablauf der Adaption.

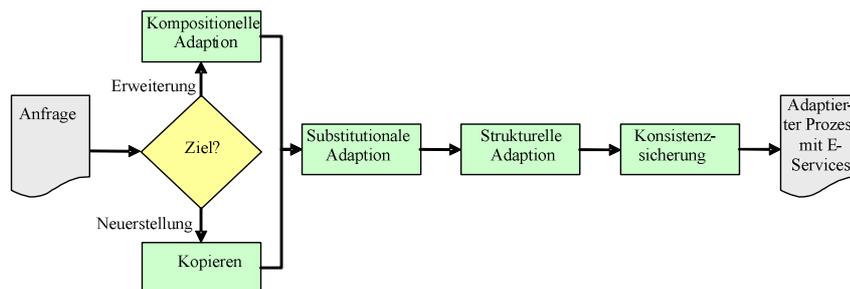


Abb. 6: Ablauf der Adaption

Ziel der kompositionellen Adaption ist es, einen neuen Prozess mit E-Services aus den Prozess- und E-Serviceelementen der Anfrage und des ähnlichen Falles zusammenzustellen. In den neuen Prozess werden alle abgeschlossenen und in Bearbeitung befindlichen Prozess- und E-Serviceelemente der Anfrage übernommen und Prozess- und E-Servicebestandteile des ähnlichen Falles hinzugefügt. Um eine sinnvolle Reihenfolge auf Serviceebene zu ermöglichen, wer-

den in der Retrieve-Phase nur solche Fälle selektiert, deren Serviceabfolge mit den Services, die in der Anfrage abgeschlossen sind oder sich in Bearbeitung befinden, übereinstimmt.

Im Rahmen der substitutionalen Adaption erfolgt die Anpassung der Attributwerte der Prozess- und E-Serviceelemente, die aus dem ähnlichen Fall entnommen sind. Hierbei wird zum Beispiel bei allen Services der Status auf „geplant“ gesetzt, als Editierungsdatum das aktuelle Datum eingetragen und ein Serviceanbieter ausgewählt, für den Präferenzen des Kunden vorliegen.

Zielsetzung der strukturellen Adaption ist es, die Struktur des Prozesses und der E-Services des ähnlichen Falles an den Prozess- und E-Service-Kontext des Anfragefalles anzupassen. In Abhängigkeit von den Gemeinsamkeiten und Unterschieden der Kontextattribute werden Komponenten des Prozesses und der E-Services des ähnlichen Falles entfernt, neue Komponenten hinzugefügt und die Reihenfolge der Komponenten verändert.

Zur Realisierung der strukturellen Adaption wird ein zusätzliches CBR-System, im Folgenden A-CBR-System genannt, verwendet. Die Fälle des A-CBR-Systems werden mit Hilfe der Fallbasis des Haupt-CBR-Systems generiert [JaCR01, S. 1013]. Dadurch werden die Adaptionfälle genau auf die Adaptionsbedürfnisse des Haupt-CBR-Systems abgestimmt. Ein Adaptionsfall setzt sich aus den Anwendbarkeitsbedingungen und dem Adaptionsbedarf (Problembeschreibung) sowie den Adaptionsaktionen (Lösung) zusammen [WiCR02, S. 426ff]. Die Anwendbarkeitsbedingungen beschreiben die Kontextattributwerte und den Prozessablauf, bei dem der Einsatz der Adaptionsaktionen sinnvoll ist. Der Adaptionsbedarf gibt die Unterschiede zwischen den Kontextattributen des Anfragefalles und des ähnlichen Falles an, die die Durchführung von Adaptionsaktionen bedingen. Die Adaptionsaktionen beinhalten die Modifikationen, die an dem ähnlichen Prozess durchgeführt werden müssen, um ihn an den neuen Kontext anzupassen. Hierbei stehen *FügeHinzu-*, *Lösche-*, *ÄndereReihenfolge-* und *ÄndereStrukturelleBeziehung-*Aktionen zur Verfügung.

Im Rahmen der strukturellen Adaption wird für einen Anfragefall und einen dazu ähnlichen Fall der Adaptionsfall gesucht, der die am besten geeigneten Adaptionsaktionen enthält. Zu diesem Zweck erzeugt das Haupt-CBR-System aus dem Anfragefall und seinem ähnlichen Fall einen Adaptionsanfragefall, den es dem A-CBR-System übergibt. Dieses selektiert aus seiner Fallbasis den dazu ähnlichsten Adaptionsfall und übermittelt ihn an das Haupt-CBR-System, das daraufhin die Adaptionsaktionen des Adaptionsfalles ausführt. Abb. 7 veranschaulicht das Zusammenspiel zwischen Haupt-CBR-System und A-CBR-System.

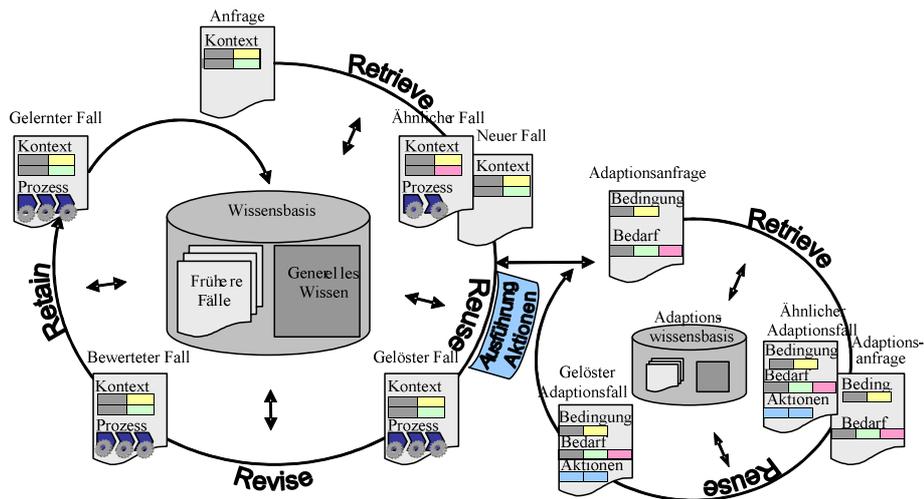


Abb. 7: Strukturelle Adaption

Da es sich bei den Adaptionsmechanismen um Heuristiken handelt, ist die Konsistenz des so erzeugten Prozesses und seiner E-Services nicht garantiert. Um die Konsistenz und damit auch die Qualität des Prozesses mit seinen E-Services zu erhöhen, kommen lokale und globale Konsistenzsicherungsmaßnahmen zur Anwendung. Im Rahmen der lokalen Konsistenzsicherung wird die Eignung einzelner Prozess- und E-Serviceelemente bezüglich des Prozess- und E-Service-Kontextes überprüft und gegebenenfalls werden Korrekturmaßnahmen vorgenommen. Die globale Konsistenzsicherung hingegen beschäftigt sich mit den Interdependenzen zwischen den einzelnen Prozess- und E-Serviceelementen und führt Maßnahmen zur Verbesserung der Konsistenz des gesamten Prozesses und seiner E-Services durch.

4.5 Ausführung des Prozesses und seiner E-Services

Ergebnis der Adaption ist ein individueller Prozess- und E-Servicevorschlag auf Basis gesammelten Prozesswissens. Ziel ist es jedoch, nicht nur die Definition individualisierter Prozesse zu unterstützen, sondern vielmehr den Ansatz einer prozessbasierten E-Service-Logistik zu realisieren. Die richtigen E-Services sollen zur richtigen Zeit am richtigen Ort im Prozess zur Verfügung gestellt werden und so die Koordinations- und Informationsbedarfe der Prozessbeteiligten befriedigen. Die dazu notwendige Ausführung der durch das CBR-System vorgeschlagenen Prozess- und E-Service-Modelle wird durch die Meta-Orchestration unterstützt, die in Abschnitt 5 skizziert wird. Zu diesem Zweck wird der Prozessvorschlag des CBR-Systems in eine XML-basierte Prozessbeschreibung transformiert und ausgeführt.

4.6 Wartung des CBR-Systems

Veränderungen in der Umwelt, dem Aufgabenfokus, den Benutzeranforderungen und den Wissenscontainern des CBR-Systems können zu einer Verschlechterung der Ergebnisqualität und Effizienz des CBR-Systems führen [Wils01, S. 1]. Die Zielsetzung der Wartung, die im Rahmen der CBR-Retain-Phase stattfindet, besteht darin, das Wissen eines CBR-Systems (Prozess- und E-Service-Wissen) zu bewahren und gegebenenfalls zu korrigieren, um eine hohe Qualität und Effizienz zu gewährleisten [Roth02, S. 30]. Im Rahmen der Individualisierung von Prozessen und E-Services werden die Wissenscontainer „Ähnlichkeitswissen“, „Adaptionswissen“ und „Fallbasis“ mit jeweils eigenen Verfahren automatisch gewartet. Der Wissenscontainer „Vokabular“, bestehend aus Kontextattributen und Prozess- und E-Serviceelementen, wird manuell von einem Domänenexperten gewartet. Das globale Ähnlichkeitswissen wird manuell initialisiert und anschließend fortlaufend interaktiv mittels der benutzerpräferierten Rangfolgen der adaptierten Prozesse und E-Services optimiert. Die Generierung des strukturellen Adaptionswissens aus Fällen der Fallbasis ermöglicht die Erweiterung und Optimierung des Adaptionswissens. Die Wartung der Fallbasis besteht aus den Retain- sowie Review- und Restore-Schritten [Roth02, S. 55ff]. Im Rahmen des Retain-Schrittes entscheiden Intrafallqualitätsmaße über die Aufnahme eines Prozesses mit seinen E-Services in die Fallbasis. Innerhalb des Review-Schrittes werden die Prozesse und E-Services hinsichtlich ihrer Intrafallqualität, Interfallkonsistenz und Redundanz überprüft und gegebenenfalls im Restore-Schritt gelöscht.

5 Systemarchitektur

Die CBR-Lösung ist Teil der Gesamtarchitektur „Individual Value Web System (IVWS)“ zur prototypischen Realisierung des Forschungsansatzes einer prozessbasierten E-Service-Logistik anhand interorganisatorischer Behandlungsprozesse in Gesundheitsnetzen [ScBo06]. Die Architektur unterstützt als Prozess- und Integrationsplattform die Konfiguration und Ausführung individueller Prozesse. Sie wurde auf Basis von .NET realisiert und gliedert sich in vier Ebenen. Präsentations-Ebene: Application Frontends initiieren und steuern alle Aktivitäten des IVWS. Typischerweise werden im Projekt Application Frontends über graphische Benutzeroberflächen

implementiert, welche die direkte Interaktion der Endanwender im Rahmen rollenspezifischer Prozessportale mit dem System ermöglichen (in C# realisierte WebParts des Portal Server).

Customization- und Prozess-Steuerungs-Ebene: Diese Ebene besteht aus drei Komponenten – der Prozess- und E-Service-Customization, der Meta Orchestration sowie dem Service Bus, welcher u.a. Funktionen für die Orchestration und Ausführung von Web Services beinhaltet.

- Service Bus: Diese Komponente verbindet alle Netzwerkakteure. Sie ermöglicht die Kommunikation zwischen Application Frontends und E-Services. Die erforderliche Funktionalität wird durch den MS Biztalk Server zur Verfügung gestellt (u.a. Connectivity, Integration Services, Communication). Der Service Bus unterstützt auch die technische Konfiguration und Ausführung feingranularer, Nachrichtenfluss-orientierter Web Service-Orchestrations (WSO), z.B. repräsentiert in BPEL oder XLANG.
- Prozess- und E-Service-Customization: Für die Individualisierung von Prozessen und der Anpassung der zugeordneten E-Service-Logistik wird Case Based Reasoning genutzt (vgl. Kapitel 4). Die vom CBR-System adaptierte Prozess- und E-Service-Spezifikation wird zur Ausführung an die Meta-Orchestration-Engine übermittelt.
- Meta Orchestration: Die Architektur basiert auf dem Prinzip Web-Service-basierter Workflows und dem Einsatz der WSO. Defizite bei der Anwendung der WSO und der dafür verfügbaren Systeme sind geringe Flexibilität, unzureichende Individualisierbarkeit und fehlende Benutzernähe. Die Meta Orchestration liefert einen Beitrag zur Behebung dieser Defizite, indem sie als Zwischenschicht zur Orchestrierungsebene (z.B. Biztalk Server) fungiert. Der Prozess-Owner kann die aus der CBR-Komponente übermittelten Prozessmodelle und E-Service-Parameter modifizieren (z.B. E-Services entfernen/hinzufügen) und das Modell instanziiieren. Die technische Konfiguration wird in der Laufzeitumgebung ausgeführt, d.h. die Meta-Orchestration-Engine stößt die Ausführung „normaler“ WSO des Biztalk-Servers an [ScBo06].

Applikations-Ebene: Neben den geschilderten IVWS-spezifischen CBR- und Orchestrierungsdiensten kann jede Form existierender Web Services Informationen und Anwendungen den Netzwerk-Teilnehmern zur Verfügung stellen, über SOAP und WSDL über den Service Bus aufrufen und in die E-Service-Logistik integrieren.

Daten-Ebene: Diese Ebene stellt die für die Umsetzung notwendige Datenbasis bereit. In einer MS SQL-Datenbank werden die Prozess- und E-Service-Daten gespeichert und verwaltet.

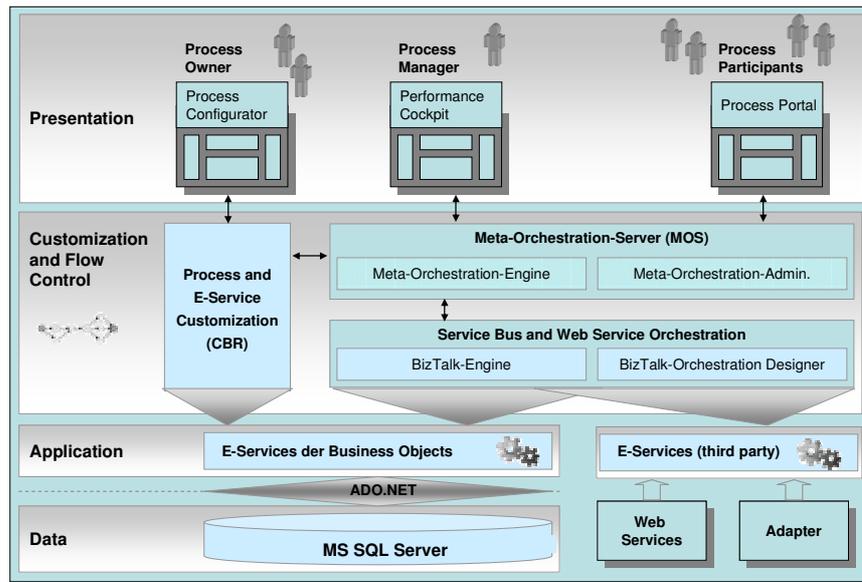


Abb. 8: Individual Value Web System

6 Zusammenfassung und Ausblick

Um Strategien in der Prozess- und IT-Landschaft effizient und effektiv umzusetzen, nimmt der Bedarf zur Flexibilisierung bzw. Individualisierung von Prozessen zu. Gleichzeitig ermöglichen neue Technologien und Standards wie Web Services und Serviceorientierte Architekturen die IuK-technische Umsetzung flexibler Wertschöpfungssysteme. Ausgehend von den bestehenden Ansätzen der Individualisierung und deren Defizite wurde ein System zur Individualisierung von Prozessen und E-Services mithilfe von Case Based Reasoning entwickelt. Zielsetzung war es, den manuellen Aufwand für die Definition und Modellierung von Prozessen für den Benutzer gering zu halten und den Vorschlag automatisch an den gegebenen Prozesskontext anzupassen. Das Konzept basiert auf intelligenten Automatismen zur Selektion geeigneter bestehender Prozesse und ihrer E-Services sowie zur Adaption gefundener Prozesse und ihrer E-Services. Die Adaption erfolgt anhand automatischer transformationsorientierter Verfahren. Hierbei werden nacheinander kompositionelle, substitutionale und strukturelle Adaptionsmechanismen sowie eine Konsistenzsicherung ausgeführt.

Das vorgestellte CBR-System zur Individualisierung von Prozessen und E-Services lässt sich dort einsetzen, wo Serviceorientierte Architekturen und Möglichkeiten der technischen Konfiguration von E-Services geschaffen wurden. Der Ansatz unterstützt die anwenderbezogene, fachliche Prozessdefinition anhand des Prozesskontextes und ermöglicht so die dynamische

Konfiguration von Prozessen, die Auswahl relevanter Prozessakteure und die Orchestrierung der E-Services. Der Ansatz ergänzt daher technisch orientierte Ansätze der Prozessrepräsentation und -konfiguration wie BPEL oder BPML. Eine aufwändige Modellierung zahlreicher Prozessvarianten oder gar individueller Prozessinstanzen entfällt.

Die Einrichtung und Wartung solcher Systeme ist mit Aufwand verbunden. So müssen vor der Nutzung des Systems einheitliche Kontextattribute, Prozesselement-Muster und E-Serviceelemente definiert sowie bestehende Prozesse mit ihren Kontexten und E-Services in dieses Format transformiert werden. Während der Nutzungsphase des Systems sind diese Daten stets an veränderte Umweltbedingungen und Benutzeranforderungen anzupassen. Gleichzeitig erfolgt damit jedoch die Explizierung und Formalisierung von Prozesswissen. Das CBR-System stellt ein Basisverfahren zur Individualisierung von Prozessen und E-Services dar, das sich noch durch zusätzliche Konzepte (z.B. Prognose unbekannter Kontextattribute) erweitern lässt. Auch die Kombination mit Prozess Repositories, Modellierungssystemen oder die Integration in Business Process Management Systeme (z.B. SAP XI, Intalio n³) weist Forschungspotenzial auf, um Wertschöpfungssysteme zu flexibilisieren und die Möglichkeiten Serviceorientierter Architekturen aus fachlicher Sicht auszuschöpfen.

Literaturverzeichnis

- [AaP194] Aamodt, Angar; Plaza, Enric: Case Based Reasoning: Foundational Issues, Methodological Variations and System Approaches. In: AI Communications 7 (1994) 1, S. 39-59.
- [Chan06] Chang, James F.: Business Process Management Systems: Strategy and Implementation. Taylor & Francis, Boca Raton 2006.
- [FeSe05] Feldmayer, Johannes; Seidenschwarz, Werner: Marktorientiertes Prozessmanagement – Wie Process Mass Customization Kundenorientierung und Prozessstandardisierung integriert. Vahlen, München 2005.
- [Flei01] Fleisch, Elgar: Das Netzwerkunternehmen: Theorien, Strategien und Prozesse zur Steigerung der Wettbewerbsfähigkeit in der “Networked economy”. Springer, Berlin et. al. 2001.

- [FLMN99] Fuchs, Béatrice; Lieber, Jean; Mille, Alain; Napoli, Amedeo: Towards a Unified Theory of Adaptation in Case Based Reasoning. In: Althoff, K.-D.; et. al. (Hrsg.): Case Based Reasoning Research and Development. Springer, Berlin 1999, S. 104-117.
- [Goos96] Goos, Klaus: Fallbasiertes Klassifizieren: Methoden, Integration und Evaluation. Infix Verlag, Sankt Augustin 1996.
- [HaRS99] Hagemeyer Jens, Rolles Roland, Scheer August-Wilhelm: Der schnelle Weg zum Sollkonzept: Modellgestützte Standardsoftwareeinführung mit dem ARIS Process Generator. In: Scheer, August-Wilhelm (Hrsg.): Veröffentlichungen des Instituts für Wirtschaftsinformatik an der Universität Saarbrücken, Heft 152, Saarbrücken 1999.
- [JaCR01] Jarmulak, Jacek; Craw, Susan; Rowe, Ray: Using Case-Based Data to Learn Adaptation Knowledge for Design. In: Proceedings of the 17th IJCAI Conference. Morgan Kaufmann, o.O. 2001, S. 1011-1016.
- [KrBS04] Krafzig, Dirk; Banke, Karl; Slama, Dirk: Enterprise SOA, Service-Oriented Architecture Best Practices. Prentice Hall PTR, Maryland 2004.
- [Lang97] Lang, Klaus: Gestaltung von Geschäftsprozessen mit Referenzbausteinen. Universität Erlangen-Nürnberg, Dissertation. 1997.
- [MaTS01] Main, Julie; Dillon, Tharam S.; Shiu, Simon C. K.: A Tutorial on Case Based Reasoning. In: Pal, Sankar K.; Dillon, Tharam S.; Yeung, Daniel S. (Hrsg.): Soft Computing in Case Based Reasoning. Springer, London 2001, S. 1-28.
- [Moor95] Moore, Geoffrey A.: The tornado – marketing strategies from Silicon Valley’s cutting edge. Harper Collins, New York 1995.
- [Port96] Porter, Michael E.: What is strategy? In: Harvard Business Review, (1996) 11.
- [Remm97] Remme, Markus: Konstruktion von Geschäftsprozessen – Ein modellgestützter Ansatz durch Montage genetischer Prozesspartikel. Gabler, Wiesbaden 1997.

- [Rich03] Richter, Michael M.: Fallbasiertes Schließen. In: Görz, Günther; Rollinger, Claus-Rainer; Schneeberger, Josef (Hrsg.): Handbuch der Künstlichen Intelligenz. 4. Aufl. Oldenbourg Wissenschaftsverlag, München 2003, S. 407-430.
- [Roth02] Roth-Berghofer, Thomas: Knowledge Maintenance of Case Based Reasoning Systems - The SIAM Methodology. Universität Kaiserslautern, Dissertation. 2002.
- [Rupp02] Rupprecht, Christian: Ein Konzept zur projektspezifischen Individualisierung von Prozessmodellen. Universität Karlsruhe, Dissertation. 2002.
- [ScBo06] Schicker, Günter; Bodendorf, Freimut: Process-based E-Service-Logistics for Healthcare Networks. In: Gesellschaft für Informatik (Hrsg.): Lecture Notes in Informatics. European Conference on eHealth. Fribourg 2006, S. 37-46.
- [Stah03] Stahl, Armin: Learning of Knowledge-Intensive Similarity-Measures in Case Based Reasoning. Universität Kaiserslautern, Dissertation. 2003.
- [WaWe97] Wargitsch, Christoph; Wewers, Thorsten: Flexware: Fallorientiertes Konfigurieren von komplexen Workflows – Konzepte und Implementierungen. In: Müller, M.; et. al. (Hrsg.): Beiträge zum 11. Workshop Planen und Konfigurieren im Rahmen der 4. Tagung Wissensbasierte Systeme. Erlangen 1997, S. 45-55.
- [WaWT97] Wargitsch, Christoph; Wewers, Thorsten; Theisinger, Felix: WorkBrain: Merging Organizational Memory and Workflow Management Systems. In: Workshop ‚Knowledge Based Systems for Knowledge Management in Enterprises‘ im Rahmen der 21. KI-Jahrestagung. Freiburg 1997.
- [Wess95] Wess, Stefan: Fallbasiertes Problemlösen in wissensbasierten Systemen zur Entscheidungsunterstützung und Diagnostik. Infix Verlag, Sankt Augustin 1995.
- [WiCR02] Wiratunga, Nirmalie; Craw, Susan; Rowe, Ray: Learning to Adapt for Case-Based Design. In: Lecture Notes in Computer Science, Band 2416. Springer Verlag, o.O. 2002, S. 421-435.
- [Wils01] Wilson, David C.: Case-Base Maintenance: The Husbandry of Experience. Indiana University, Dissertation. 2001.

Plädoyer für die Entwicklung perspektivenspezifischer Problemlösungskomponenten zur Unterstützung der Prozessverbesserung

Ralf Knackstedt, Martin Pellengahr

European Research Center for Information Systems (ERCIS)

Institut für Wirtschaftsinformatik

Westfälische Wilhelms-Universität Münster

Leonardo-Campus 3

48149 Münster

{ralf.knackstedt, martin.pellengahr}@wi.uni-muenster.de

Abstract

Ein wesentliches Charakteristikum des Business Process Engineering stellt die modellbasierte Prozessanalyse und -verbesserung dar. Zur Modellerstellung selbst liegt eine Vielzahl von Modellierungstechniken vor. Die methodische Unterstützung der Analyse von Ist-Modellen und die darauf basierende Entwicklung von Soll-Modellen weist Entwicklungspotenziale auf. Der Beitrag stellt einen multiperspektivischen Ansatz vor, mit dem Prozessgestalter bei ihrer Analysetätigkeit detailliert unterstützt werden können.

1 Modellbasierte Unterstützung von Prozessverbesserungen als Gegenstand von Design Science Research

Business Process Engineering (BPE) formuliert mit der Bezugnahme auf „Engineering“ den Anspruch einer ingenieurmäßigen Gestaltung von Geschäftsprozessen, die durch eine methodenbasierte Problemlösung gekennzeichnet ist [Teub97]. Probleme lassen sich dabei als das Empfinden einer Diskrepanz zwischen einem Ist und einem Soll charakterisieren [Zeile99]. Methoden beschreiben systematische Verfahren zur Überwindung von Problemen [Lore95, 876; Stah95, 239]. Im Wesentlichen lassen sich Methoden anhand von Aktivitäten, Ergebnissen und Techniken beschreiben [Gutz94]. Innerhalb vorgegebener Aktivitäten werden unter Einsatz

bestimmter Techniken Ergebnisse erzielt. Die Ergebnisse einzelner Aktivitäten können als Input in die Bearbeitung nachfolgender Aktivitäten einfließen, wodurch sich eine Vernetzung einzelner Aktivitäten ergibt. Techniken beschreiben, wie im Rahmen der Aktivitäten die Ergebnisdokumente zu erstellen sind.

Eine für das BPE übliche Abfolge von Aktivitäten ist die Repräsentation bestehender Prozesse in Form von Ist-Modellen, die Analyse dieser Modelle hinsichtlich Verbesserungspotenzialen bzw. Schwachstellen, die Entwicklung von Soll-Modellen, in welche die Analyseergebnisse einfließen, und die Umsetzung der Soll-Modelle [BeBK05].

Das wissenschaftliche Erkenntnisinteresse in Bezug auf die beschriebene modellbasierte Vorgehensweise des BPE kann gemäß einer in der Betriebswirtschaftslehre und Wirtschaftsinformatik üblichen Unterscheidung Erklärungsziele oder Gestaltungsziele verfolgen [Zele99]. Eine erklärungszielorientierte Forschung würde beispielsweise die Verbreitung und Akzeptanz verschiedener Techniken im Rahmen des BPE untersuchen. Die gestaltungszielorientierte Forschung, die im Information Systems Research zurzeit unter dem Begriff des Design Science Research diskutiert wird, zielt dagegen auf die Entwicklung von Artefakten, die einen Beitrag zur Problemlösung leisten [HMPR04]. Design Science Research lässt sich dabei in zwei Phasen untergliedern. In der ersten Phase werden die Artefakte, wie z. B. Sprachkonstrukte, Modelle, Methoden und Implementierungen, entwickelt, wobei auf Ergebnisse erklärungszielorientierter Forschung zurückgegriffen werden kann. In der zweiten Phase wird der Problemlösungsbeitrag der geschaffenen Artefakte evaluiert [HMPR04].

Der vorliegende Beitrag schlägt ein den Vorgehensprinzipien des Design Science Research entsprechendes Verfahren zur systematischen Entwicklung von Methodenbestandteilen vor, die im Rahmen des BPE Gestalter dabei unterstützen sollen, auf der Basis von Ist-Modellen gezielt Verbesserungspotenziale bzw. Schwachstellen zu identifizieren. Dafür wird zunächst auf der Basis einer Betrachtung des State-of-the-Art der modellbasierten Prozessanalyse die leitfragengestützte Analyse als entwicklungsfähiger Gegenstand von Design Science Research ausgewählt (Kapitel 2). Für die als Untersuchungsgegenstand gewählte Prozessmodellanalyse wird ein Vorgehensmodell zur Ausgestaltung von Design Science Research entwickelt (Kapitel 3). Der Ansatz wird selbst einer Evaluation unterzogen, indem er für die Entwicklung einer leitfragengestützten Prozessmodellanalyse zur Kundenintegration angewendet wird (Kapitel 4). Forschungsmethodische Bemerkungen bilden den Abschluss des Beitrags (Kapitel 5).

2 State-of-the-Art der modellbasierten Prozessanalyse

2.1 Bestehende modellbasierte Ansätze zur Unterstützung von Prozessverbesserungen

Zur Unterstützung der Identifikation von Verbesserungspotenzialen bzw. Schwachstellen in Ist-Prozessmodellen als Grundlage für die Entwicklung von Soll-Modellen werden in Literatur und Praxis vor allem die folgenden Ansätze verfolgt:

- *Prozessmodellierungstechniken*: Einen großen Schwerpunkt bisheriger gestaltungszielorientierter Forschung im Rahmen der Prozessanalyse stellte bislang die Entwicklung von Prozessmodellierungstechniken dar. Die verschiedenen Ansätze unterscheiden sich insbesondere darin, wie detailliert jeweils die Kontroll-, Informations-, Material- und Finanzflüsse in Geschäftsprozessen abgebildet werden können und welche betrieblichen Objekte zum Gegenstand der Modellierung gemacht werden. Prozessmodellierungstechniken tragen wesentlich dazu bei, den betrieblichen Gegenstandsbereich zu strukturieren und damit das Gestaltungsproblem in eine leichter lösbare Fassung zu überführen.
- *Referenzmodellierung*: Referenzprozessmodelle stellen die Common Practice bzw. Best Practice der Gestaltung betrieblicher Abläufe dar [Sche99]. Sie unterstützen die Prozessverbesserung, indem sie eine Vergleichsbasis für die Ist-Prozessmodelle zur Verfügung stellen. Referenzprozessmodelle liegen für verschiedene Branchen und innerbetriebliche Funktionsbereiche vor [FeLo04].
- *Benchmarking*: Der Aussagegehalt von Referenzprozessmodellen lässt sich durch die zusätzliche Hinterlegung von Benchmarks erweitern [Schü98]. Neben den Prozessstrukturen können dann auch die Ausprägungen wichtiger Prozesskennzahlen, wie z. B. Durchlaufzeiten, Kosten, Fehlerhäufigkeiten etc., verglichen werden. Das Supply Chain Operations Reference Model [SCC05] stellt ein verbreitetes Beispiel für ein Prozessreferenzmodell dar, für das umfangreiche Benchmarkingdaten kommerziell angeboten werden [Harm03, 8].
- *Simulation und Animation*: Die Simulationsfähigkeit von Prozessmodellen stellt ein weiteres wichtiges Kriterium zur Unterscheidung von Prozessmodellierungstechniken dar [Uthm01]. Durch die Modellierungswerkzeug-unterstützte Simulation von Prozessdurchläufen lassen sich zu Modellvarianten Prozesskennzahlenausprägungen

gen generieren, auf deren Basis eine Auswahl von Gestaltungsalternativen erfolgen kann. Bei der Prozessmodellanimation steht anstelle der Erhebung von Prozesskennzahlen die Visualisierung exemplarischer Prozessdurchläufe im Vordergrund.

Die unterschiedenen Ansätze zur Unterstützung von Prozessverbesserungen sind alle von Bedeutung und sollten zum Gegenstand weiterer Forschungen gemacht werden. In der Praxis wird die referenzmodellbasierte Prozessverbesserung durch eine zu geringe Verfügbarkeit geeigneter Referenzmodelle eingeschränkt. Auch das Benchmarking scheitert häufig am Fehlen der Daten vergleichbarer Unternehmen [ScLa05]. Voraussetzung für eine systematische Simulation ist unter anderem die angeleitete Entwicklung der zu untersuchenden Prozessmodellvarianten. Diese Umstände sprechen dafür, dass die Entwicklung von Soll-Modellen allein auf der Basis von Ist-Modellen ohne Rückgriff auf Referenzmodelle, Benchmarking und Simulation weiterhin von erheblicher Bedeutung sein wird. Daher wird der allein Prozessmodellierungstechnikbasierte Ansatz im Folgenden näher untersucht.

2.2 Konzeptionalisierungen von Prozessmodellierungstechniken

Zur Konzeptionalisierung von Modellierungstechniken im Allgemeinen und von Prozessmodellierungstechniken im Speziellen wird vorgeschlagen, den konzeptionellen Aspekt, den repräsentationellen Aspekt und eine Handlungsanleitung zu unterscheiden [Holt01]. Der konzeptionelle Aspekt legt die grundlegenden Konzepte und deren Beziehungen fest, mittels derer betriebliche Abläufe strukturiert und repräsentiert werden. Die grundlegenden Konzepte Ereignisgesteuerter Prozessketten (EPK) sind z. B. die Funktion, das Ereignis, der Kontrollfluss und verschiedene logische Konnektoren. Der repräsentationelle Aspekt ordnet den Elementen des konzeptionellen Aspekts Symbole und Topologien zu. Im Falle der EPK werden den Ereignissen z. B. Rauten als Symbole zugeordnet. Die Handlungsanleitung beschreibt das Vorgehen bei der Konstruktion von Modellen auf der Basis der eingeführten sprachlichen Mittel. Beispielsweise wird empfohlen, dass die Modellierung von EPKs mit der Repräsentation der Startereignisse begonnen wird. Durch ihre Unterstützung bei der Strukturierung des betrieblichen Geschehens stellen die genannten Bestandteile eine wichtige Grundlage zur Problemlösung dar. Wird der Leitgedanke einer umfassenden methodischen Unterstützung des Business Process Engineering zu Grunde gelegt, wird deutlich, dass Modellierungstechniken allein für die Unterstützung der Entwicklung von Soll-Prozessmodellen aus Ist-Modellen ohne Rückgriff auf Referenzmodelle, Benchmarkingdaten und Simulationen bzw. Animationen nicht ausreichen. Der Prozessgestalter benö-

tigt zusätzliche Problemlösungstechniken, die ihn bei der Identifikation von Prozessverbesserungspotenzialen bzw. Schwachstellen anleiten [BKHH01].

Hinsichtlich der Explizierung entsprechender Problemlösungstechniken zeigt sich zurzeit ein deutliches Missverhältnis. Während eine beachtliche Vielzahl von Modellierungstechniken vorgeschlagen wird und deren Komponenten in Form von sprach- und prozessorientierten Metamodellen zunehmend umfangreicher und formalisierter dokumentiert werden, finden sich als Anleitungen zur Analyse von Ist-Modellen lediglich sehr allgemein gehaltene Empfehlungen. Im Wesentlichen wird der Prozessgestalter angehalten nach Möglichkeiten zu suchen, Funktionen zu parallelisieren, zu eliminieren, zu beschleunigen, Kontrollflüsse zu vereinfachen sowie Rücksprünge und Medienbrüche zu vermeiden [ScLa05, 173].

Im Folgenden wird auf dieses Missverhältnis reagiert, indem ein Ansatz zur Ausgestaltung von Design Science Research vorgestellt wird, der darauf zielt, Artefakte in Form von perspektiven-spezifischen leitfragengestützten Problemlösungstechniken zu entwickeln, die Prozessgestalter bei der Entwicklung von Soll-Prozessmodellen auf der Basis von Ist-Prozessmodellen unterstützen.

3 Ein Ansatz zur Entwicklung von Problemlösungstechniken zur Unterstützung von Prozessverbesserungen

Das hier vorgestellte Konzept einer multiperspektivischen Prozessanalyseunterstützung zielt darauf ab, auf unterschiedliche Anwendungskontexte ausgerichtete, leitfragenbasierte Aufbereitungen von Prozessmodellen zur Verfügung zu stellen. Für die methodische Ausgestaltung von Design Science Research und Entwicklung dieser Problemlösungstechniken zur Unterstützung von Prozessverbesserungen wird ein Phasenmodell vorgeschlagen, das fünf Aufgabenbereiche unterscheidet (vgl. Abb. 1(a)). Die in den verschiedenen Phasen entwickelten Ergebnisse und deren Beziehungen werden anhand eines Entity-Relationship-Modells beschrieben (vgl. Abb. 1(b)).

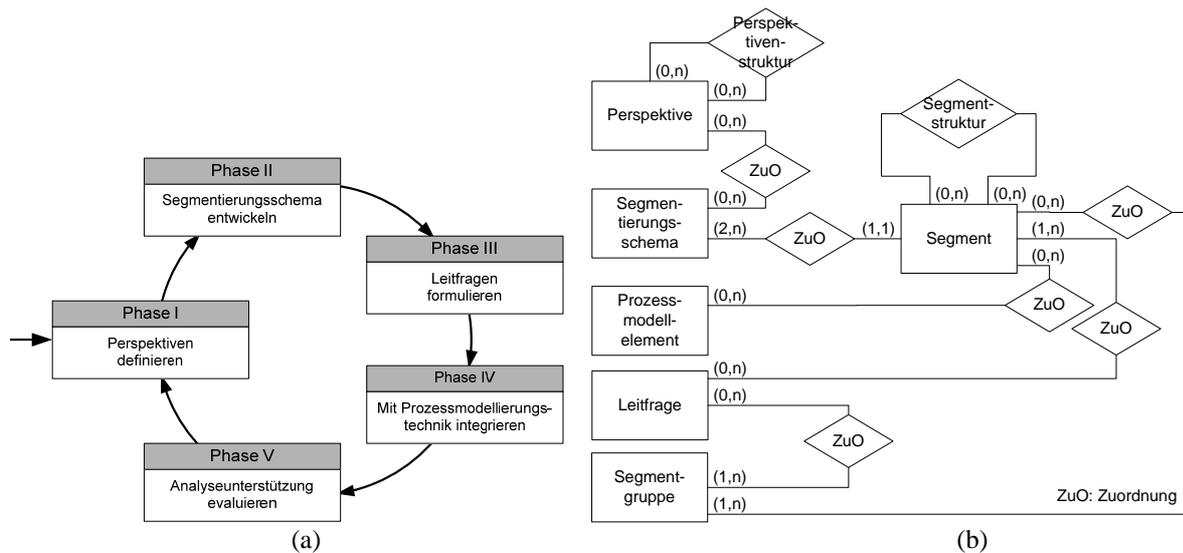


Abb. 1: Phasen- und Entity-Relationship-Modell zur Entwicklung perspektivenspezifischer leitfragengestützter Problemlösungstechniken zur Unterstützung von Prozessverbesserungen

Phase I sieht vor, dass spezielle *Perspektiven* definiert werden, aus deren Sicht die Prozessanalyse vorgenommen wird. Mit Hilfe von Perspektiven werden dementsprechend die Anwendungskontexte spezifiziert, innerhalb derer die im Weiteren zu konzipierenden Problemlösungstechniken eingesetzt werden sollen. Eine mögliche Perspektive stellt z. B. die Analyse des Prozesses aus Sicht des Wissensmanagements dar. Verfeinerungen und Zusammensetzungen von Perspektiven spiegeln sich in der Perspektivenstruktur wieder. Mit der Berücksichtigung der Perspektiven wird der Leitgedanke der multiperspektivischen Modellierung auf die Unterstützung der Prozessverbesserung übertragen, der davon ausgeht, dass sich die adäquaten Subjektivierungen von Problemlösern je nach Anwendungskontext bzw. Analysefragestellung unterscheiden [Luhm99, 181 ff.]. Die Berücksichtigung der Perspektiven gewährleistet, dass im Vergleich zu den bisher vorliegenden allgemeinen Empfehlungen zur Prozessanalyse detaillierte und spezielle Anleitungen entwickelt werden.

In Phase II ist für jede Perspektive – möglichst auf der Basis erklärungszielorientierter Forschungsergebnisse – ein *Segmentierungsschema* für die Prozessmodellelemente zu entwickeln. Das Segmentierungsschema soll die für die gewählte Perspektive relevanten Unterscheidungen von Prozessmodellelementen wiedergeben und ist von wesentlicher Bedeutung für die theoretische Fundierung der zu entwickelnden Problemlösungstechnik. Beispielsweise interessiert aus der Sichtweise des Wissensmanagements die Unterscheidung, ob in einer Funktion Daten erzeugt, gespeichert, verteilt oder genutzt werden [Heis02, 53]. Ein Segmentierungsschema muss mindestens eine Unterscheidung in zwei solche Segmente vornehmen. Jedes Segment gehört

dabei genau einem Segmentierungsschema an. Ein Segmentierungsschema kann für mehrere Perspektiven geeignet sein, und eine Perspektive kann auf mehrere Segmentierungsschemata zurückgreifen. Segmente können in andere Segmente eingehen bzw. andere Segmente umfassen, was in der Segmentstruktur festgehalten wird. Während die Segmentstruktur der Abbildung von Verfeinerungsbeziehungen dient, erlauben Segmentgruppen sonstige inhaltlich begründete Zusammenstellungen von Segmenten.

In Phase III werden zu den Elementen der Segmentierungsschemata so genannte *Leitfragen* formuliert, die den Prozessanalysten gezielt auf perspektiven- und segmentspezifische Gestaltungsoptionen hinweisen. Im Kontext des Wissensmanagements ist z. B. zu fragen, ob alle von Funktionen gespeicherten Daten von anderen Funktionen genutzt werden. Leitfragen werden Segmenten oder Segmentgruppen zugeordnet. Für die Zuordnung zu Segmentgruppen sind solche Leitfragen prädestiniert, welche die gleichzeitige Berücksichtigung unterschiedlicher Prozessmodellbereiche erfordern. Leitfragen aller Segmentierungsschemata einer Perspektive ergeben die perspektivenspezifische Checkliste zur Prozessverbesserung.

Im Anschluss an die Entwicklung dieser leitfragenbasierten Problemlösungstechniken ist in Phase IV eine *Integration* mit einer Modellierungstechnik vorzunehmen. Die Integration bestehender Modellierungstechniken ist dabei einer grundsätzlichen Neuentwicklung im Allgemeinen vorzuziehen. Aufgabe der Integration ist es, die konzeptionellen und repräsentationellen Aspekte der Modellierungstechnik so zu erweitern, dass die in den vorangegangenen Phasen definierten Prozessmodellsegmente ausgewiesen werden können. Die Zuordnung kann z. B. durch eine entsprechende direkte oder indirekte Attributierung von Funktionen Ereignisgesteuerter Prozessketten vorgenommen werden. Bei der direkten Attributierung werden die Segmentbezeichnungen als Attributwerte verwendet. Bei der indirekten Zuordnung werden dagegen Regeln definiert und ausgewertet, die aus den Prozessmodellelementen hinterlegten Attributwerten die jeweils zugehörigen Segmente ableiten. Da mit den Attributen für die Zuordnung von Prozessmodellelementen zu Segmenten gegebenenfalls in die Modelle Informationen eingeführt werden, die nicht für sämtliche Perspektiven relevant sind, werden die im Rahmen der multiperspektivischen Modellierung entwickelten Mechanismen zur Modelladaptation in das vorgestellte Konzept integriert [BeDK04]. Auf diese Weise sollten diejenigen Attribute, die nicht für die jeweilige Perspektive von Bedeutung sind, ausgeblendet werden können. Die Attributierung, dass Daten erzeugt, gespeichert, verteilt oder genutzt werden, ist z. B. spezifisch für die Prozessanalyse aus Sicht der Perspektive Wissensmanagement. Je nach Segmentierung sind

geeignete Darstellungen zu entwickeln. Für eine Wissensmanagement-spezifische Segmentierung eines Prozessmodells könnte z. B. vorgesehen werden, dass die Funktionen je nach Prozesszugehörigkeit unterschiedlich eingefärbt oder mit einer zusätzlichen Beschriftung versehen werden.

Dem grundsätzlichen Vorgehen des Design Science Research folgend ist in Phase V eine *Evaluation* der um die Problemlösungskomponente erweiterten Prozessmodellierungstechnik vorzunehmen. Im Rahmen der Evaluation können unterschiedliche Forschungsmethoden zum Einsatz kommen, die den Prozess bzw. das Ergebnis der Anwendung der Problemlösungstechniken untersuchen [Hein00; BWHW05]. Die Anwendung der Analyseunterstützung umfasst die Modellierung von Prozessen mit der gewählten Prozessmodellierungstechnik und die perspektiven-spezifische Analyse des Prozessmodells mit Hilfe der adaptierten Modelle und zugeordneten Leitfragenkataloge sowie die durch die Leitfragen unterstützte Ableitung von Verbesserungspotenzialen und Schwachstellen. Die jeweiligen Ansatzpunkte zur Prozessverbesserung sind zu konsolidieren und münden in der Modellierung eines Soll-Modells. Während sich mit qualitativen Forschungsmethoden wie z. B. Fallstudien, Action Research, narrativen Interviews und Beobachtungen der Prozess kritisch nachvollziehen lässt, lassen sich über Experimente Ergebnisse der Methodenanwendung sammeln, die sich auch quantitativ auswerten lassen.

Die Ergebnisse der Evaluation können motivieren, Teile der methodischen Erweiterungen zu ergänzen, zu ändern oder zu eliminieren. Entsprechend dieser Lernzyklen sind wiederholte Durchläufe durch die Phasen der Methodenerweiterung vorzunehmen.

4 Umsetzung des Konzepts für die Perspektive Kundenintegration

4.1 Perspektiven definieren

KLEINALTENKAMP erkennt die adäquate Einbindung des Kunden und seiner externen Faktoren in den Leistungserstellungsprozess, die so genannte *Kundenintegration*, als wesentlichen Erfolgsfaktor für eine kundenorientierte Prozessgestaltung [Klei00]. Der allgemeine Trend zur Produktindividualisierung und kundenindividuellen Leistungserstellung bewirkt, dass bei jeder Form der Dienstleistungsproduktion vom Kunden zur Verfügung gestellte Produktionsfaktoren, mindestens Informationen, in den Wertschöpfungsprozess integriert werden müssen [Klei00]. Das betriebswirtschaftliche Kernproblem der Kundenintegration besteht folglich in der zielge-

richteten Steuerung der Zusammenarbeit zwischen Anbieter und Nachfrager. Die Analyse der Integration der externen Faktoren des Kunden stellt daher eine wesentliche Perspektive der modellbasierten Prozessverbesserung dar. Sie durch die Formulierung geeigneter Leitfragen zu unterstützen, ist das Ziel der nachfolgenden Schritte.

4.2 Segmentierungsschema entwickeln

Zur Analyse des Leistungserstellungsprozesses zwischen Kunde und Anbieter wurden in der Marketing-Literatur seit den 1980er-Jahren mehrere aufeinander aufbauende Ansätze zur Unterscheidung verschiedener kunden- und anbieterseitiger Aktivitätsbereiche entwickelt [Shos82; King89]. Die Herleitung der entsprechenden Unterscheidungen kann dabei über betriebs- und volkswirtschaftliche, verhaltensklärende Theorien abgestützt werden. Hierbei wird u. a. auf die Principal Agent Theorie, Informationsökonomik und Transaktionskostentheorie zurückgegriffen. Die entsprechenden Argumentationen können hier nicht wiedergegeben werden, weshalb z. B. auf [KIMa95] verwiesen werden muss. KLEINALTENKAMP fasst die wichtigsten Ergebnisse dieser Forschungen in einem umfassenden Ansatz zusammen [Klei00; Klei99]. Kern dessen ist die Unterteilung des Dienstleistungsprozesses in Aktivitätsbereiche, die sich hinsichtlich ihrer Nähe zum Kunden unterscheiden (vgl. im Folgenden [Klei99]):

- *Aktivitäten des Kunden:* Dieser Bereich weist die Aktivitäten aus, die vom Kunden in den Leistungserstellungsprozess eingebracht werden.
- *Onstage-Aktivitäten:* Der Bereich der Onstage-Aktivitäten umfasst alle für den Kunden sichtbaren Anbieteraktivitäten.
- *Backstage-Aktivitäten:* Diejenigen Anbieteraktivitäten, die zwar für den Kunden nicht sichtbar aber von Personal durchgeführt werden, das im Rahmen anderer Aktivitäten durchaus Kundenkontakt besitzt und entsprechend über spezielle Erfahrungen im Umgang mit Kunden verfügt, werden als Backstage-Aktivitäten bezeichnet.
- *Support-Aktivitäten:* Wenngleich für den Kunden nicht unmittelbar sichtbar, so kennzeichnet Support-Aktivitäten dennoch die Eigenschaft, dass sie unmittelbar kundenindividuell ausgeführt werden und damit als einzelkundeninduziert zu bezeichnen sind.

- *Preparation-Aktivitäten:* Maßnahmen der Vorkombination, die nicht unmittelbar von konkreten Kundenaufträgen abhängen, aber die der Markterschließung dienen bzw. auf die im Rahmen der einzelkundeninduzierten Prozesse zurückgegriffen wird, gelten als Preparation-Aktivitäten.
- *Facility-Aktivitäten:* Um die prinzipielle Leistungsfähigkeit des Unternehmens zu gewährleisten, stellen Facility-Aktivitäten grundlegende Ressourcen bereit.

4.3 Leitfragen formulieren

Werden die Funktionen von Kundenintegrationsprozessen im Rahmen der Istmodellierung gemäß dem im vorigen Abschnitt entwickelten Segmentierungsschema unterschieden, so eröffnet sich die Möglichkeit, zu jedem dieser Bereiche spezielle Leitfragen zu formulieren, die Gestaltungsoptionen zur Erreichung einer verbesserten Kundenintegration aufzeigen.

Aktivitäten des Kunden	<ul style="list-style-type: none"> • An welchen Prozessen des Kunden will sich das Unternehmen beteiligen? • Sollen Aktivitäten, die bisher vom Kunden durchgeführt werden, vom Anbieterunternehmen übernommen werden? • An welchen Stellen treten Probleme auf, wenn der Kunde seine Leistungsbeiträge (z. B. Produktanforderungen oder andere Informationen) nicht in der erforderlichen zeitlichen, qualitativen und quantitativen Form in den Prozess einbringt? • Wie kann eine adäquate Einbindung des Kunden zur Erleichterung der Planung aus Anbietersicht forciert werden (z. B. durch rechtzeitige Abforderung von Leistungsbeiträgen)? • ...
Onstage-Aktivitäten	<ul style="list-style-type: none"> • Welche Aktivitäten, die für den Kunden bisher sichtbar sind, sollen ihm zukünftig verborgen bleiben? • Wie beurteilt der Kunde (vermutlich) die sichtbaren Aktivitäten? Welche Daten sind in welcher Form zu erfassen, um eine Einschätzung der für den Kunden sichtbaren Aktivitäten aus dessen Sicht zu ermitteln? • Welche Kanäle (Internet, Telefon etc.) sollten genutzt werden, um dem Kunden den gewünschten Einblick zu gewähren? Soll hierbei das Push- oder Pull-Prinzip zur Anwendung kommen? • ...
Backstage-Aktivitäten	<ul style="list-style-type: none"> • Sollten zusätzliche Aktivitäten, die für den Kunden relevant sind, für den Kunden sichtbar gemacht werden? • Wie zufrieden sind die Kunden mit den einzelnen Personalbereichen mit Kundenkontakt? Welche Daten sind zu erfassen, um entsprechende Informationen zu erheben? • ...
Support-Aktivitäten	<ul style="list-style-type: none"> • Sollten bisherige Support-Aktivitäten besser von Bereichen mit direktem Kundenkontakt übernommen werden? • Ergeben sich Vorteile durch eine Dezentralisierung verschiedener Aufgaben, indem sie von Kundenkontaktpersonal vorgenommen werden, das näher am Standort des Kunden ist? Welche Organisationseinheiten kommen hierfür in Frage? • Können (Teil-)Prozesse standardisiert werden? • ...
Preparation-Aktivitäten	<ul style="list-style-type: none"> • Sollten zusätzliche Aktivitäten kundenindividuell ausgeführt werden? • Werden die notwendigen Kundendaten zur Planung von Markterschließungsmaßnahmen erhoben? • ...
Facility-Aktivitäten	<ul style="list-style-type: none"> • Sollten Teilaufgaben auf Sublieferanten verlagert oder bislang ausgelagerte Aktivitäten selbst übernommen werden? • Wie kann die Koordination mit Lieferanten optimiert werden? • ...

Tab. 1: Ausschnitt aus einem Leitfragenkatalog zur Gestaltung der Kundenintegration [KnDa04]

Ein speziell für die Kundenintegration entwickelter Leitfragenkatalog ist in Ausschnitten in Tab. 1 abgebildet [KnDa04].

4.4 Mit Prozessmodellierungstechnik integrieren

Um die leitfragengestützte Problemlösungstechnik mit Prozessmodellierungstechniken verknüpfen zu können, muss die Unterscheidung der Aktivitätsbereiche innerhalb der Prozessmodelle ermöglicht werden. Für Ereignisgesteuerte Prozessketten bietet es sich dabei beispielsweise an, die Modellelemente der EPK gemäß ihrer Segmentzugehörigkeit in Spalten anzuordnen. Dieses Vorgehen profitiert im Falle der EPK davon, dass die Spaltendarstellung im Rahmen der prozessorientierten Reorganisation bereits als ein gängiges Mittel zur Steigerung der Übersichtlichkeit Verbreitung gefunden hat [RoSD05, 69 f.].

4.5 Analyseunterstützung evaluieren

Die vorgestellte Problemlösungstechnik wurde bisher im Rahmen verschiedener Business Process Engineering Projekte in verschiedenen Unternehmen eingesetzt und dabei kontinuierlich weiterentwickelt. Beispielsweise wurde mittels des vorgestellten Verfahrens das Projektgeschäft eines IT-Service-Providers analysiert [BeKK05]. Auch bei einem Entsorgungsdienstleister konnten mit Hilfe der vorgestellten Problemlösungstechnik Ansätze zur Verbesserung seiner Prozesse systematisch entwickelt werden [Saue05].

Neben diesen Evaluationen auf Basis qualitativer Untersuchungsmethoden wurde auch ein quantitativ ausgerichtetes Laborexperiment durchgeführt. Das Evaluationsziel bestand dabei darin, Erkenntnisse darüber zu gewinnen, ob das vorgestellte modellierungstechnische Konzept seinem Anspruch tatsächlich gerecht wird, Schwachstellen leichter auffindig zu machen. Das Laborexperiment wurde im Rahmen einer Hauptstudiumsvorlesung am Institut für Wirtschaftsinformatik der Westfälischen Wilhelms-Universität Münster durchgeführt. Da die Einführung in die Modellierung Ereignisgesteuerter Prozessketten fester Bestandteil dieser Veranstaltung ist, konnte davon ausgegangen werden, dass alle am Laborexperiment teilnehmenden Probanden mit der EPK-Modellierung hinreichend vertraut sind. Da die Anzahl gefundener Schwachstellen im Evaluationsmodell als (quantitative) abhängige Variable gewählt und den wesentlichen Einflussfaktor die Verwendung des Leitfragenkataloges in Kombination mit der Segmentierung des Prozessmodells in Form einer (qualitativen) unabhängigen Variablen ausmacht, liegt ein einfaktorieller Versuchsplan vor [BoDö, 531]. Die unabhängige Variable ist zweifach gestuft (Analy-

se mit / ohne methodische Unterstützung), sodass das Laborexperiment mit zwei Versuchsgruppen, d. h. Experimental- und Kontrollgruppe, auskommt.

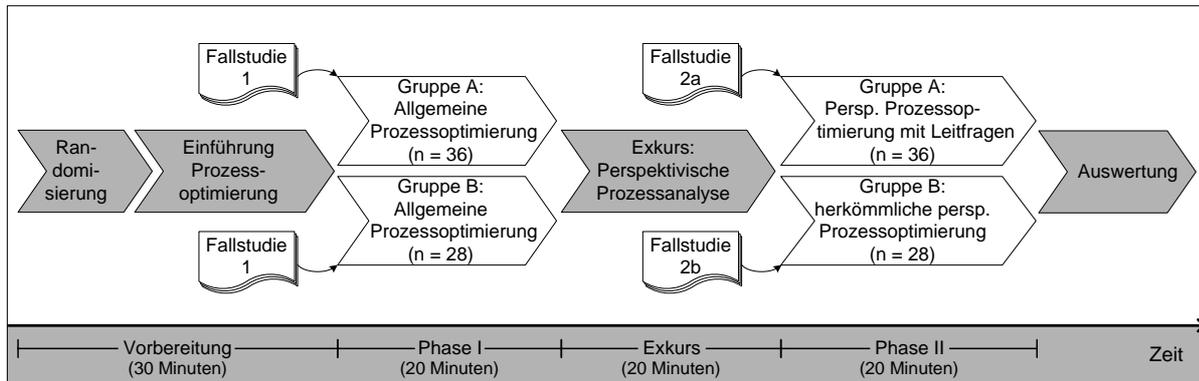


Abb. 2: Ablauf des Experiments

Der Ablauf des Experiments (vgl. Abb. 2) begann für die Probanden mit der Auslosung ihrer Gruppenzugehörigkeit. Diese Randomisierung verfolgte den Zweck, die interne Validität des Experiments zu erhöhen und Störvariablen, insbesondere personenbezogene, wie z. B. unterschiedliche Vorkenntnisse unter den Probanden, zu neutralisieren [BoDö02, 58 ff.]. Als Konsequenz hieraus ergab sich eine zufallsbedingt unterschiedliche Gruppenstärke.

Die eigentliche Durchführung des Experiments vollzog sich in zwei Phasen. Inhalt der ersten Phase war eine Vorstudie, in der die Probanden unabhängig von ihrer Gruppenzugehörigkeit mit dem später in Phase II aus der Perspektive der Kundenintegration zu analysierenden Prozessmodell konfrontiert wurden. Im Anschluss an eine Einführung in herkömmliche Ansätze zur Unterstützung von Prozessverbesserungen (vgl. Abschnitt 2.1) durch den Versuchsleiter wurde den Studierenden hierzu das in eine Fallstudie integrierte Prozessmodell ausgehändigt. Um inhaltlichen Missverständnissen vorzubeugen, wurde das Szenario des betrachteten Geschäftsprozesses erläutert, ehe die Probanden den Auftrag erhielten, das Modell in einer 20-minütigen Analyse auf sich allein gestellt hinsichtlich aller von ihnen für relevant erachteter Gestaltungsoptionen zu untersuchen und Schwachstellen sowie Verbesserungsideen in einer dafür vorgesehenen Liste zu dokumentieren.

Im Anschluss an das Einsammeln der Ergebnislisten wurde die Informationsbasis der Versuchspersonen mit einem Exkurs zur speziellen Perspektive der Prozessanalyse Kundenintegration erweitert. Dabei wurden die in Abschnitt 4.2 eingeführten Aktivitätengruppen visuell unterstützt vorgestellt und wesentliche auf diesen fußende Gestaltungsoptionen aus Abschnitt 4.3 verbal erläutert. In der darauf folgenden Versuchsphase II wurde von der Gruppeneinteilung

Gebrauch gemacht: Die Probanden erhielten nun unterschiedliches Untersuchungsmaterial. Aufgrund ihrer Unkenntnis über die Zugehörigkeit zur Versuchs- oder Kontrollgruppe handelte es sich hierbei um einen Blindversuch. Während sich bei der Kontrollgruppe (im Folgenden *Gruppe B*) mit der Analyseperspektive der Kundenintegration lediglich die Zielsetzung der Prozessanalyse änderte, erhielt die Experimentalgruppe (im Folgenden *Gruppe A*) neben dem Leitfragenkatalog ein den Vorschlägen in Abschnitt 4.4 entsprechendes, in Spaltennotation aufbereitetes Prozessmodell sowie eine Erläuterung zu deren Nutzung. Diese letzte Analysephase dauerte wiederum 20 Minuten.

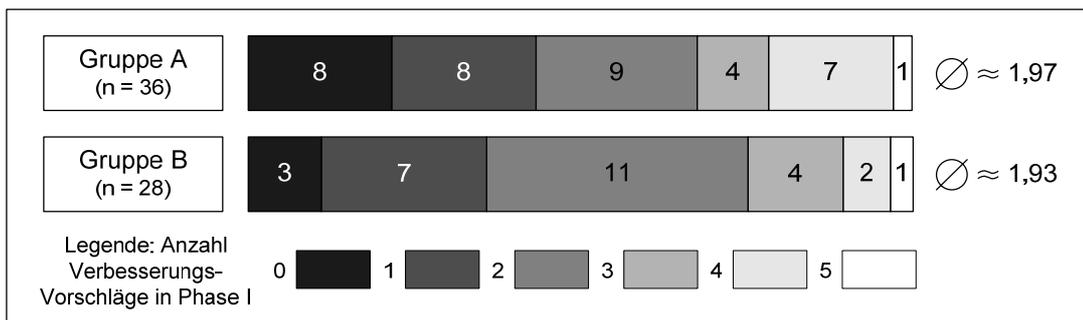


Abb. 3: Verteilung der Anzahl gemachter Verbesserungsvorschläge der beiden Versuchsgruppen in Phase I

Die Auswertung der ausgefüllten Schwachstellenlisten aus Phase I zeigt, dass die Verteilung der Anzahl gefundener Fehler als Indiz für die relative Homogenität der beiden Versuchsgruppen gedeutet werden kann (vgl. Abb. 3). Um die Bewertungsgrundlage zu systematisieren, gingen lediglich konstruktive Verbesserungsvorschläge in die Auswertung ein, zudem wurde versucht, die Verbesserungen in einem möglichst einheitlichen Granularitätsgrad zu werten.

Abb. 4 visualisiert die am häufigsten angemerkteten Verbesserungsvorschläge aus Sicht der Kundenintegration im Detail, wobei die erste Spalte die absolute Häufigkeit unter allen Teilnehmern aufsummiert und die zweite Spalte dies hinsichtlich der Gruppen differenziert. Es zeigt sich, dass sämtliche Probanden unabhängig von ihrer Gruppenzugehörigkeit in Phase II deutlich mehr Verbesserungsideen entwickelten. Dies kann als Wirksamkeit einer perspektivenspezifischen Herangehensweise bei der Prozessverbesserung gewertet werden. Im direkten Vergleich der beiden Gruppen offenbart sich ferner, dass die durchschnittlich gefundene Fehleranzahl in der Experimental- (mit ausgeteiltem Leitfragenkatalog 3,8) höher ist als in der Kontrollgruppe (ohne ausgeteilten Leitfragenkatalog 2,6). Aufgrund der hinsichtlich der gefundenen Fehler unterschiedlichen Ergebnisse zeigt sich bezüglich der Wirksamkeit der Prozessaufbereitung in Spaltennotation und der Bereitstellung eines detaillierten und nicht nur kurz erläuterten Leitfragenkatalogs kein eindeutiges Ergebnis.

Verbesserungsvorschläge aus Sicht der Kundenintegration	Anzahl konstruktiver Verbesserungsvorschläge in Phase I bzw. Phase II	Vergleich der Experimental- (Gruppe A) und Kontrollgruppe (Gruppe B) in der jeweiligen Phase
Projektleiter gemeinsam „on stage“ mit dem Kunden bestimmen	<p>Phase II</p>	Gruppe A 16,6 % Gruppe B 7,1 %
Gemeinsame Auftragskorrektur mit dem Kunden	<p>Phase I</p>	Gruppe A 30,5 % Gruppe B 10,7 %
	<p>Phase II</p>	Gruppe A 58,3 % Gruppe B 60,7 %
Vorprojekt durchführen: Informationen vom Kunden anfordern	<p>Phase II</p>	Gruppe A 36,1 % Gruppe B 10,7 %
Vorprojektauswertung an den Kunden übermitteln	<p>Phase II</p>	Gruppe A 36,1 % Gruppe B 17,8 %
Vorprojektauswertung gemeinsam mit dem Kunden durchführen	<p>Phase II</p>	Gruppe A 16,6 % Gruppe B 14,3 %
Übermittlung des Projektplans an den Kunden	<p>Phase II</p>	Gruppe A 69,4 % Gruppe B 50 %
Einbezug des Kunden in die Projektplanerstellung	<p>Phase II</p>	Gruppe A 50 % Gruppe B 32,1 %

Abb. 4: Gruppenübergreifende Verteilung der Verbesserungsvorschläge aus Sicht der Kundenintegration sowie deren Verteilung auf die beiden Versuchsgruppen

Über die bisherigen quantitativen Ergebnisse hinaus kann zudem festgehalten werden, dass die Experimentalgruppe die gefundenen Schwachstellen mittels des Vokabulars aus dem Leitfragenkatalog ausdrucksstärker beschreibt.

5 Schlussbemerkungen

Basierend auf einer Betrachtung des State-of-the-Art wurde die Unterstützung der Entwicklung von Soll-Modellen auf der Basis von Ist-Modellen als ein bisher unzureichend methodisch unterstützter Bereich des Business Process Engineering identifiziert. Einen Beitrag zur Schließung dieser Lücke zu leisten, wurde auf zweifache Weise verfolgt: Einerseits wurde ein forschungsmethodischer Ansatz zur Konstruktion von Artefakten in Form von Prozessverbesserungstechniken entwickelt, andererseits wurde dieses Konzept angewendet, um für die Kundenintegration

eine konkrete Prozessverbesserungstechnik zu schaffen. In beiden Fällen handelt es sich um gestaltungszielorientierte Forschungsfragen, die sich der Design Science Research zuordnen lassen. Die abschließende Zusammenfassung und Kommentierung der Ergebnisse dieses Beitrags werden daher gemäß der von HEVNER et al. formulierten Design Science Research Guidelines strukturiert (vgl. Tab. 2) [HMPR04].

Guideline	Forschungsmethodischer Ansatz (Kapitel 3)	Prozessverbesserungsunterstützung für die Perspektive Kundenintegration (Kapitel 4)
1. Design as an artefact	Artefakt in Form einer Forschungsmethode dokumentiert durch Phasenmodell (Aktivitäten) und Entity-Relationship-Modell (Ergebnisse)	Artefakt in Form einer Methode zur Prozessverbesserung aus der Perspektive Kundenintegration bestehend aus einer um eine Problemlösungstechnik ergänzten Modellierungstechnik
2. Problem relevance	Entwicklung von Problemlösungstechniken zur Unterstützung der Entwicklung von Soll-Modellen als relevantes Problem	Verbesserung der Kundenintegration als relevantes Problem
3. Design evaluation	Fallstudienbasierte Evaluation des forschungsmethodischen Ansatzes (Ausweitung der Evaluation durch Berücksichtigung weiterer Anwendungsfälle)	Evaluation des Ansatzes zur Kundenintegration mittels Durchführung eines Modellierungsexperiments mit 64 Probanden unter Bildung einer Experimental- und einer Kontrollgruppe
4. Research contribution	Schema zur Entwicklung von Problemlösungstechniken für die Prozessverbesserung, das auf unterschiedliche Perspektiven angewendet werden kann und dabei der Explizierung und praktischen Verbreitung von Gestaltungswissen dient	Operationalisierter Ansatz zur Entwicklung von Prozessverbesserungen hinsichtlich der Kundenintegration, der bezüglich seiner Auswirkungen auf das Analyseverhalten von Prozessgestaltern weiter untersucht werden sollte
5. Research rigour	Übertragung der Grundlagen konstruktivistischer, multiperspektivischer Modellierung auf die Prozessverbesserung; ausbaufähige fallstudienbasierte Evaluation	Übernahme von in der Marketing-Literatur langjährig entwickelten Unterscheidungskriterien zur Prozesssegmentierung; um qualitative Untersuchungen ergänzbare experimentelle Evaluation
6. Design as a search process	Weitere Verbesserung und Detaillierung des Ansatzes im Zuge der Anwendung auf weitere Analyseperspektiven und im Zuge der Entwicklung einer Integration des Ansatzes in Modellierungswerkzeuge	Anpassung des Segmentierungsschemas und der Leitfragen in Folge einer ausgeweiteten Untersuchung der Methodenanwendung als Prozess
7. Communication of research	Konzentration auf die Präsentation des Phasenmodells und des Entity-Relationship-Modells aus Platzgründen	Konzentration auf Ausschnitte des Leitfragenkatalogs, auf Grundzüge der Modellierungstechnikintegration und der Evaluationsergebnisse aus Platzgründen; Archivierung der Daten des durchgeführten Experiments

Tab. 2: Ergebnisse vor dem Hintergrund der Design Science Research Guidelines nach HEVNER et al. [HMPR04]

Literaturverzeichnis

- [BeBK05] Becker, J.; Berning, W.; Kahn, D.: Projektmanagement. In: Becker, J.; Kugeler, M.; Rosemann, M. (Hrsg.): Prozessmanagement. Ein Leitfaden zur prozessorientierten Organisationsgestaltung. Berlin u.a. 2005, S. 17-44.
- [BeDK04] Becker, J.; Delfmann, P.; Knackstedt, R.: Konstruktion von Referenzmodellierungssprachen. Ein Ordnungsrahmen zur Spezifikation von Adaptionsmechanismen für Informationsmodelle. In: Wirtschaftsinformatik, 46 (2004) 4, S. 251-264.

- [BeKK05] Becker, J.; Klose, K.; Knackstedt, R.: Process Modelling for Service Processes – Modelling Methods Extensions for Specifying and Analysing Customer Integration. In: Chen, C.-S. ; Filipe, J.; Seruca, I.; Cordeiro, J. (Hrsg.): Proceedings of the Seventh International Conference on Enterprise Information Systems (ICEIS 2005). Miami, USA 2005, S. 260-265.
- [BKHH01] Becker, J.; Knackstedt, R.; Holten, R.; Hansmann, H.; Neumann, S.: Konstruktion von Methodiken: Vorschläge für eine begriffliche Grundlegung und domänenspezifische Anwendungsbeispiele. In: Arbeitsberichte des Instituts für Wirtschaftsinformatik. Westfälische Wilhelms-Universität, Münster, 2001.
- [BoDö02] Bortz, J.; Döring, N.: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaften. Berlin u.a. 2002.
- [BWHW05] Braun, C.; Wortmann, F.; Hafner, M.; Winter, R.: Method Construction – A Core Approach to Organizational Engineering. In: Haddad, H. (Hrsg.): Applied Computing 2005, Proceedings of the 2005 ACM Symposium on Applied Computing, Santa Fe, USA, 13.03.2005. New York, USA, 2005, S. 1295-1299.
- [FeLo04] Fettke, P.; Loos, P.: Referenzmodellierungsforschung. In: Wirtschaftsinformatik, 46 (2004) 5, S. 331-340.
- [Gutz94] Gutzwiller, T.: Das CC RIM-Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen. Heidelberg, 1994.
- [Harm03] Harmon, P.: An Introduction to the Supply Chain Council's SCOR Methodology. http://www.businessprocesstrends.com/deliver_file.cfm?fileType=publication&fileName=Intro%20SCOR%20Method%20Whitepaper%2001-0311.pdf, Abruf am 2006-05-17.
- [Hein00] Heinrich, L. J.: Bedeutung von Evaluation und Evaluationsforschung in der Wirtschaftsinformatik. In: Heinrich, L. J.; Häntschel, I. (Hrsg.): Evaluation und Evaluationsforschung in der Wirtschaftsinformatik. München, Wien 2000, S. 7-22.

- [Heis02] Heisig, P.: GPO-WM: Methode und Werkzeuge zum geschäftsprozessorientierten Wissensmanagement. In: Abdecker, A. ; Hinkelmann, K.; Maus, H.; Müller, H. J. (Hrsg.): Geschäftsprozessorientiertes Wissensmanagement. Berlin u.a. 2002, S. 47-64.
- [HMPR04] Hevner, A. R.; March, S. T.; Park, J. and Ram, S.: Design Science in Information Systems Research. MIS Quarterly, 28 (1), 2004. 75-105.
- [Holt01] Holten, R.: Metamodell. In: Mertens, P. et al. (Hrsg.): Lexikon der Wirtschaftsinformatik, 4. Aufl., Springer, Berlin et al. 2001, S. 300 f.
- [King89] Kingman-Brundage, J.: The ABC's of Service System Blueprinting. In: Bitner, M. J.; Crosby, L. A. (Hrsg.): Designing a Winning Service Strategy. Chicago 1989, S. 551-565.
- [Klei99] Kleinaltenkamp, M.: Service-Blueprinting, Ein Instrument zur Steigerung der Effektivität und der Effizienz von Dienstleistungsprozessen. In: Technischer Vertrieb, 1. Jg. (1999), Nr. 2, S. 33-39.
- [Klei00] Kleinaltenkamp, M.: Customer Integration im Electronic Business. In: Weiber, R. (Hrsg.): Handbuch Electronic Business. Informationstechnologien, Electronic Commerce, Geschäftsprozesse. Wiesbaden 2000, S. 333-357.
- [KIMa95] Kleinaltenkamp, M.; Marra, A.: Institutionenökonomische Aspekte der 'Customer Integration'. ZfbF, (1995) Sonderheft 35, S. 101-117.
- [KnDa04] Knackstedt, R.; Dahlke, B.: Customer Process Management. In: Wirtschaftswissenschaftliches Studium 33 (2004) 1, S. 47-54.
- [Lore95] Lorenz, K.: Methode. In: Mittelstraß, J. (Hrsg.): Enzyklopädie Philosophie und Wissenschaftstheorie. Band 2. Stuttgart 1995, S. 876-879.
- [Luhm99] Luhmann, N.: Zweckbegriff und Systemrationalität. Über die Funktion von Zwecken in sozialen Systemen. 6. Aufl., Frankfurt am Main 1999.
- [RoSD05] Rosemann, M.; Schwegmann, A.; Delfmann, P.: Vorbereitung der Prozessmodellierung. In: Becker, J.; Kugeler, M.; Rosemann, M. (Hrsg.): Prozessmanage-

ment. Ein Leitfaden zur prozessorientierten Organisationsgestaltung. Berlin u.a. 2005, S. 45-103.

- [Saue05] Sauerbrei, E.: Analyse und Optimierung von Logistikprozessen unter besonderer Berücksichtigung der Kundenintegration am Beispiel von Hellmann Process Management. Diplomarbeit, Westfälische Wilhelms-Universität, Münster 2005.
- [SCC05] Supply-Chain Council: Supply Chain Operations Reference-model (SCOR) 7.0. http://www.supply-chain.org/galleries/default-file/SCOR%207.0_FINAL.pdf, 2005, Abruf am 2006-05-17.
- [Sche99] Scheer, A.-W.: ARIS – House of Business Engineering: Konzept zur Beschreibung und Ausführung von Referenzmodellen. In: Schütte, R.; Becker, J.; Rosemann, M. (Hrsg.): Referenzmodellierung: State-of-the Art und Entwicklungsperspektiven. Heidelberg, 1999.
- [Schü98] Schütte, R.: Grundsätze ordnungsmäßiger Referenzmodellierung. Konstruktion konfigurations- und anpassungsorientierter Modelle. Wiesbaden 1998.
- [ScLa05] Schwegmann, A.; Laske, M.: Istmodellierung und Istanalyse. In: Becker, J.; Kugeler, M.; Rosemann, M. (Hrsg.): Prozessmanagement. Ein Leitfaden zur prozessorientierten Organisationsgestaltung. Berlin u.a. 2005, S. 155-184.
- [Shos82] Shostack, G. L.: How to Design a Service. In: European Journal of Marketing, 16 (1982) 1, S. 49-63.
- [Stah95] Stahlknecht, P.: Einführung in die Wirtschaftsinformatik. 7. Aufl., Berlin 1995.
- [Teub97] Teubner, R. A.: Integrierte Organisations- und Informationssystemgestaltung: Methoden für das Organization and Information Engineering. Münster 1997.
- [Uthm01] von Uthmann, C.: Geschäftsprozesssimulation von Supply Chains. Ein Praxis-Leitfaden für die Konstruktion von Management-orientierten Modellen integrierter Material- und Informationsflüsse. Dissertation, Münster 2001.
- [ZeLe99] Zelewski, S.: Grundlagen. In: Corsten, H.; Reiß, M. (Hrsg.): Betriebswirtschaftslehre. 3. Aufl., München et al. 1999, S. 1-125.

Einführung in den Track

Collaborative Business

Dr. Oliver Frick

SAP Research

Prof. Dr. Gerhard Schwabe

Universität Zürich

PD Dr. Michael Koch

Technische Universität München

Prof. Dr. Volker Wulf

Universität Siegen

Der Track fokussiert die folgenden Themengebiete:

- Unterstützung von Kooperation im Unternehmen und in Business Webs
- Business Communities und Social Software
- Wissensmanagement über Communities of Practice
- Expertise Management und Expertise Recommendation
- Virtuelle Unternehmen und deren IT-Unterstützung
- Diffusion und Aneignung von Kooperationstechnologien
- Interoperabilität von Kooperationsplattformen
- Neue Benutzungsschnittstellen für Kooperationssysteme

Programmkomitee:

PD Dr. M. Koch, Technische Universität München

Prof. Dr. V. Wulf, Universität Siegen

Prof. Dr. G. Schwabe, Universität Zürich

Dr. O. Frick, SAP Research Business Development

Prof. Dr. Petra Schubert, FH Nordwestschweiz, Basel, Schweiz

Prof. Dr. Kathrin Möslein, Handelshochschule Leipzig

Dr. Marco Leimeister, Technische Universität München

Prof. Dr. Ulrike Lechner, Universität der Bundeswehr München

Prof. Dr. Ralf Reichwald, Technische Universität München

Prof. Dr. Johann Schlichter, Technische Universität München

Prof. Dr. Felix Hampe, Universität Koblenz-Landau

Dr. Stefan Uellner, T-Systems Enterprise Service GmbH, Darmstadt

Präsenzbasierte Echtzeitkommunikation

Eine prototypbasierte Untersuchung der Nutzbarkeit im Unternehmensberatungskontext

Kai Riemer

European Research Center for Information Systems (ERCIS)

Universität Münster

48149 Münster

kai.riemer@ercis.de

Abstract

Echtzeitkommunikationssysteme basieren auf der Integration von Kommunikationsmedien und der Bereitstellung von Präsenzinformationen und haben die Verbesserung der Erreichbarkeit in verteilten Teams zum Ziel. Die hier präsentierte Studie zeigt, dass diese Systemklasse ihren vollen Nutzen erst durch eine kontextspezifische Interpretation und Adaption entfalten kann. In Interviews mit Unternehmensberatern wurden Anforderungen für die Gestaltung einer Echtzeitkommunikationslösung für das Projektgeschäft in der Beratungsbranche abgeleitet. Auf der Plattform von Siemens Openscape wurde ein dedizierter Prototyp implementiert, der die Nutzbarkeit dieser Systemklasse für Unternehmensberatungen demonstriert. Die Ergebnisse werden reflektiert und in den Kontext einer weiteren Erforschung dieser Systemklasse eingeordnet.

1 Einleitung

Echtzeitkommunikationssysteme (Real-Time-Communication – RTC-Systeme) präsentieren sich als eine neue Klasse von Systemen, die aus der Konvergenz von Telekommunikations- und Informationstechnologie entstanden sind. RTC-Systeme integrieren Kommunikationsmedien und Endgeräte mit Groupwarefunktionen und Präsenzinformationen. Die Hersteller versprechen sich von RTC-Systemen eine Verbesserung des individuellen Kommunikationsmanagements, sowie der Erreichbarkeit in Teams und Unternehmensprozessen. Da sich die Systeme zurzeit noch im Entwicklungsstadium befinden, bieten sich zum jetzigen Zeitpunkt aus Sicht der For-

schung insbesondere die Beschäftigung mit Prototypen zur Untersuchung von Funktionsspektrum und Anwendbarkeit dieser Systemklasse an.

Der vorliegende Beitrag präsentiert eine prototypische Umsetzung einer RTC-Lösung für den Kontext der Unternehmensberatung basierend auf einer qualitativen empirischen Erhebung von Anforderungen eines solchen Nutzungsszenarios. Dabei verfolgt der Beitrag mehrere Ziele: Zunächst wird die Echtzeitkommunikation als neues Anwendungs- und Forschungsfeld eingeführt. Die Entwicklung des Prototyps dient der Illustration der Nutzbarkeit von RTC in dem ausgewählten Szenario. Eine wesentliche Erkenntnis der Studie ist die Notwendigkeit zur kontextspezifischen Umsetzung und Anpassung von RTC-Lösungen; hierzu werden typische Funktionen für das gewählte Szenario diskutiert; dabei steht die Konzeption des Prototyps, nicht die technische Umsetzung im Vordergrund.

Zunächst wird das Thema anhand einiger organisatorischer Entwicklungen motiviert (Kap. 2). Kap. 3 gibt eine kurze Einführung zu RTC-Systemen, bevor Kap. 4 den Forschungsansatz und das methodische Vorgehen erläutert. In Kap. 5 werden die Anforderungen des Szenarios herausgearbeitet, die sich aus der Analyse der Interviews ergeben. Kap. 6 erläutert dann die Konzepte und Funktionen des entwickelten Prototyps zur Behandlung der identifizierten Kommunikationsprobleme. Ein Ausblick auf zukünftige Forschungsaktivitäten schließt den Beitrag ab.

2 Hintergrund und Motivation

2.1 Herausforderungen organisationaler Virtualisierung und verteilter Gruppenarbeit

Veränderungen in der Wettbewerbssituation in nahezu allen Branchen führen zu einer zunehmenden Virtualisierung von Organisations- und Wertschöpfungsstrukturen. So flexibilisieren Unternehmen ihre internen Organisationsstrukturen, um auf externe Marktanforderungen besser und schneller reagieren zu können [LWCL01]. Dabei rücken virtuelle Teams in den Mittelpunkt, die unter Nutzung von IKT über zeitliche, räumliche und organisatorische Grenzen hinweg arbeiten und dabei die Arbeitsteilung und Spezialisierung bei gleichzeitig loser Kopplung anstreben [Scho02]. Die Idee ist, dass Teams nicht mehr nach ihrem räumlichen Standort sondern anhand der Qualifikation ihrer Mitglieder zusammengesetzt werden [KoHe02].

Virtuelle Kooperation jedoch ist besonders auf der Teamebene problematisch und mit zahlreichen Herausforderungen an die Koordination verbunden, die aus der räumlichen Verteilung und

Nutzung von Kommunikationstechnologie resultieren [RiKl05]. Bei virtuellen Teams kommt es häufiger als in traditionellen Gruppen zu Kommunikationsproblemen und Missverständnissen, sowie zu Problemen beim Herausbilden eines funktionierenden Gruppengefüges [Andr02]. Ein Ansatzpunkt zur Verbesserung der Situation liegt in der Verbesserung der Kommunikation, Erreichbarkeit und gegenseitiger Sichtbarkeit der Teammitglieder.

2.2 Neue Technologien erweitern das Kommunikationsspektrum

Auf der technischen Ebene hat sich die Zahl der verfügbaren Kommunikationsmedien und Geräte in den letzten Jahren stark erhöht, so dass der durchschnittliche Anwender heute mit einem heterogenen Geräte- und Kanalmix bei der Kommunikation konfrontiert ist [LyYo02]. Durch neuere Entwicklungen wie Voice-over-IP-Telefonie, Instant Messaging, und Desktop-Videokonferenzen haben sich die Kommunikationsoptionen weiter vermehrt [Laza06]. Die vorhandene Komplexität erhöht sich in einer konkreten Situation zudem dadurch, dass die zu erreichende Person oft mehrere Telefonnummern, E-Mail- oder andere Medien-Accounts besitzt. Als Folge daraus ergibt sich eine hohe Komplexität für den Kommunikationsempfänger, da er diverse Geräte und Accounts verwalten muss, wie auch für den Initiator, der oftmals mehrere Optionen probieren muss, um einen Empfänger zu erreichen. Als Resultate zeigen sich gescheiterte Kommunikationsversuche, Zeitverlust und Friktionen in der Zusammenarbeit.

2.3 Interaktionskomplexität als typisches Phänomen

Verteilte Arbeitskontexte und das Spektrum der verfügbaren Kommunikationsoptionen in Kombination führen zu einer oftmals unbefriedigenden Situation in der modernen Projektarbeit. Neben einer generell schlechten Erreichbarkeit der Mitarbeiter resultiert dies in häufigen Unterbrechungen des Arbeitsablaufs durch so genannte Kommunikationsasymmetrien; diese ergeben sich, wenn der Zeitpunkt eines Kontaktversuchs zwar für den Initiator, nicht aber für den Rezipienten passend ist [NaWB00]. Kommunikationsmittel wie Mobiltelefone, die eigentlich der Verbesserung der Erreichbarkeit dienen sollen, zeichnen sich durch eine hohe Spontaneität auf Seiten des Initiators aus. Allerdings führen sie gerade hierdurch leicht zu ungewollten Störungen auf Seiten des Rezipienten [ReGo05]. Dies ist insbesondere störend für die Konzentration bei komplexen, wissensintensiven Tätigkeiten [deIB05]. Mitarbeiter, die von den gestiegenen Kommunikationsanforderungen eines solchen Arbeitskontextes überfordert werden, erleben

einen so genannten „Interaction overload“ [SoMK02]. Typische Reaktionen sind das Ausschalten des Endgeräts und das Nichtbeantworten von Kommunikationsanfragen. Ein solches Schutzverhalten verschlechtert dann die ohnehin oft schlechte Erreichbarkeit.

3 Präsenzbasierte Echtzeitkommunikation

3.1 Überblick

Echtzeitkommunikationssysteme werden von ihren Herstellern zur Lösung der oben beschriebenen Probleme positioniert. Das Ziel ist eine Verbesserung des Medien- und Gerätemanagements, sowie der Erreichbarkeit in verteilten Arbeitssituationen. RTC-Systeme sind ausnahmslos komplexe Client-Server-Systeme, die sich von einfachen präsenzbasierten Tools wie Skype durch ihren Funktionsumfang und ihren Integrationsanspruch unterscheiden. RTC-Systeme sind als unternehmensweite Infrastruktur konzipiert, die den Aufbau komplexer Kommunikationslandschaften nebst Integration klassischer Telefoninfrastruktur ermöglichen [Laza06]. Als Anbieter solcher Systeme positionieren sich Unternehmen aus dem IKT- sowie aus dem Telekommunikationsmarkt¹. RTC-Systeme basieren auf zwei wesentlichen Bausteinen: der Medienintegration im Sinne eines Unified-Communication-Ansatzes (UC), sowie den Präsenzinformationen zur Verbesserung der Sichtbarkeit und Erreichbarkeit von Kommunikationspartnern.

3.2 Unified-Communications

Ziel des UC-Ansatzes ist es, insbesondere synchrone Medien und Geräteklassen in einem System zu vereinen, den Anwender bei deren Verwaltung zu unterstützen und die Auswahl der jeweils in einer Situation geeigneten Medien anhand von Regeln zu automatisieren. Basierend auf IP-Technologie sorgt eine logische Steuerungsschicht dafür, dass eingehende Kommunikationsvorgänge automatisch auf die vom Anwender situativ bevorzugten und gerade verfügbaren Endgeräte weitergeleitet werden. Hierfür müssen die Medien (Text, Audio, Video), Geräte (Mobiltelefon, IP-Telefon, etc.) und Softwareclients (Instant-Messenger, Video- und Audioclients) im RTC-System konfiguriert sein. Das System soll so die Komplexität reduzieren und die Kontrolle über die Medienwahl vom Initiator zum Rezipienten verlagern.

¹ Hierzu gehören u.a. IBM, Microsoft, Oracle, sowie Alcatel, Avaya, Cisco, Nortel und Siemens.

3.3 Präsenzinformation

Der zweite Baustein ist die Präsenzinformation (Presence Awareness Information), die aus Instant-Messaging-Tools bekannt ist und dort durch ein entsprechendes Icon die Erreichbarkeit eines Kontakts signalisiert. Im verteilten Kontext fehlt die Sichtbarkeit traditioneller Signale wie die physische Anwesenheit und die Körpersprache von Rezipienten, die die Verfügbarkeit für die Kommunikation signalisieren [deIB05]. Das Ziel von RTC-Systemen ist es, diese mangelnde Awareness durch technische vermittelte Signalisierung auszugleichen. Unter Awareness wird in diesem Zusammenhang das Verständnis für die Aktivitäten anderer Nutzer verstanden, das den notwendigen Kontext für die Abstimmung der eigenen Aktivitäten bildet [DoBe92]. Hierbei muss zwischen der Perspektive des Initiators und der des Rezipienten unterschieden werden: Aus Sicht des Rezipienten erlaubt das aktive Signalisieren mittels Präsenzstatus die Steuerung der eigenen Erreichbarkeit und die Vermeidung ungewollter Störungen. Der Initiator kann mittels Präsenzstatus die Erreichbarkeit des Rezipienten besser abschätzen und entscheiden, ob ein Kommunikationsversuch aktuell sinnvoll ist. Die geschaffene Awareness durch Präsenzinformationen soll so helfen, das Management der Erreichbarkeit in Gruppen zu verbessern.

4 Forschungsansatz und Methodik

4.1 Gestaltungsorientierte Forschung

RTC-Systeme befinden sich derzeit noch im Aufbau- und Entwicklungsstadium. Während Visionen, Funktionen und intendierter Nutzen den Produktbeschreibungen der Hersteller entnommen werden können, wirft die mögliche Anwendung der Systeme zum jetzigen Zeitpunkt zahlreiche Fragen auf [RiFr06]. Aufgrund des frühen Stadiums existieren noch keine empirischen Fälle der Umsetzung solcher Systeme in Organisationen. Somit kann zurzeit allenfalls die Nutzung von Teilfunktionen wie Instant-Messaging oder Voice-over-IP-Telefonie empirisch im Kontext untersucht werden. Um sich dem Thema RTC aus Forschungssicht zu nähern, bietet sich daher zum jetzigen Zeitpunkt die prototypische Umsetzung und Untersuchung solcher Systeme an. Die hier präsentierte Studie kann in Anlehnung an Hevner et al. als eine gestaltungsorientierte Studie beschrieben werden, die auf einer verhaltensorientierten Erhebung aufbaut.

Während der verhaltensorientierte Ansatz („behavioral science research“) den Einsatz und die Nutzung von IT-Artefakten in Organisationen untersucht, beschäftigt sich der gestaltungsorientierte Ansatz („design science research“) mit der Entwicklung und Evaluation von IT-Artefakten zur Lösung organisatorischer Probleme [HMPR04].

Für die RTC-Domäne bieten sich zu diesem frühen Zeitpunkt im Forschungsprozess besonders gestaltungsorientierte Ansätze an, um erste Erkenntnisse zum typischen Funktionsumfang und der Umsetzung von RTC-Systemen zu bekommen. Eine solche Studie kann jedoch nicht ohne ein verhaltensorientiertes Element auskommen; so wurde zu Beginn dieser Studie ein konkreter Nutzungskontext empirisch mittels qualitativer Erhebungsmethoden untersucht, um die Grundlage für die gestaltungsorientierte Phase zu bilden. Im Mittelpunkt der Studie stand dabei das Lernen über die kontexttypische Anwendung der RTC-Idee, sowie deren Limitationen.

4.2 Forschungsprozess

Im Rahmen dieser Studie wurde für ein konkretes organisatorisches Szenario ein RTC-Prototyp konzeptioniert und zu Demonstrationszwecken implementiert. Die Studie basiert auf der Grundannahme, dass RTC-Systeme einer kontextspezifischen Interpretation und Adaption bedürfen, um ihren vollen Nutzen entfalten zu können. Die Wahl eines Nutzungskontextes viel hierbei auf Unternehmensberatungen, da diese viele der oben geschilderten Merkmale moderner Arbeitskontexte, wie räumliche und zeitliche Verteilung, Flexibilität, Projektarbeit und mobiles Arbeiten aufweisen und somit einen geeigneten Anwendungsbereich für RTC abgeben sollten. Zur Identifikation der typischen Kommunikationsmuster in Unternehmensberatungen wurden Interviews mit Vertretern von vier Unternehmen unterschiedlicher Größe (von KMU bis multinationaler Beratung) durchgeführt. Hierbei bildete die Erhebung von typischen Friktionen im Kommunikationsprozess den Schwerpunkt zur Gewinnung von Ansatzpunkte für den RTC-Einsatz. Alle Interviews wurden transkribiert und ausgewertet. Dabei wurde von den Spezifika der einzelnen Fälle abstrahiert, um nur grundlegende Kommunikationsmuster zu identifizieren mit dem Ziel einer möglichst weit reichenden Übertragbarkeit. Die empirische Erhebung bildete den notwendigen Kontext für die gestaltungsorientierte Phase. Aufbauend auf den identifizierten Problembereichen und Anforderungen (siehe Kap. 5) folgte die Konzeption und Entwicklung eines dedizierten RTC-Prototyps für Unternehmensberatungen, der technisch auf Siemens Hipath Openscape aufsetzt (siehe Kap. 6). Die Demonstration des Prototyps und die Diskussion der Erkenntnisse schließen den Forschungsprozess ab.

5 Analyse des Anwendungsszenarios Unternehmensberatung

5.1 Organisationsprinzip und Kommunikationsmuster

Der Alltag in Unternehmensberatungen wird von der Projektarbeit dominiert. Unternehmensberater verbringen einen Großteil der Zeit mit Tätigkeiten in gemischten Teams an den Standorten der Klienten. Neben der Kommunikation innerhalb dieser Teams ist dabei die Kontaktaufnahme mit einer Reihe von außenstehenden Personen wichtig: Projektexterne Berater werden kontaktiert, um dringende Informationen zu projektrelevanten Themen und Hintergrundinformationen über abgeschlossene Projekte zu erfragen. Dabei enthalten archivierte Dokumente über abgeschlossene Projekte meist Angaben über die beteiligten Mitarbeiter, die für weitergehende Informationen aber meist persönlich kontaktiert werden müssen. Darüber hinaus stehen für Unterstützungsaufgaben wie Recherche und Präsentationserstellung oft spezielle Abteilungen („Backoffice“) zur Verfügung, die ebenfalls zeitnah kontaktiert werden müssen. Es wird deutlich, dass ein solches Szenario viele Charakteristika eines virtuellen Teams aufweist, wie räumliche Verteilung, gemischte Teams und den Einsatz von IKT für die Kommunikation.

5.2 Typische Kommunikationsprobleme und resultierende Anforderungen

Zur Analyse der Kommunikationssituation, sowie zur Identifikation typischer Kommunikationsprobleme wurden offene Interviews mit vier Beratern von Unternehmen unterschiedlicher Größe geführt. Von allen Interviewpartnern wurde die Kommunikation sowohl als kritischer Erfolgsfaktor wie auch als häufiges Problemfeld der täglichen Arbeit unmittelbar hervorgehoben. Die im Weiteren diskutierten Kommunikationsprobleme bilden die Schnittmenge, die in allen vier Interviews präsent war. Als wichtigster Bereich konnte dabei die mangelnde Erreichbarkeit, sowohl von verteilt arbeitenden Teammitgliedern, als auch von Ansprechpartnern zu bestimmten Themen und abgeschlossenen Projekten identifiziert werden. Hierbei steht zunächst einmal das Problem der erfolglosen Kommunikationsversuche im Mittelpunkt, das sich durch eine mangelnde Transparenz für den aktuellen Kontext der zu erreichenden Person ergibt. Hiermit einher geht die mangelnde Kenntnis des Mediums, über das ein Ansprechpartner erreichbar ist, so dass sich im Zweifel mehrere vergebliche Kontaktversuche aneinanderreihen. Des Weiteren wurde bemängelt, dass es häufig nicht möglich ist, die (oftmals unbekannt) Verfasser von Dokumenten oder Experten zu bestimmten Themen kurzfristig zu erreichen.

Der zweite Problembereich bezieht sich auf das Problem der Störung in Stresssituationen. Wenn Kollegen wichtige Informationen benötigen, können deren Kontaktversuche sehr ungelegen kommen, wenn z.B. die Vorbereitung eines Workshops die volle Aufmerksamkeit erfordert. Gleichzeitig kann sich der Berater aber nicht vollständig abschotten, da er für Teamkollegen immer noch erreichbar sein muss.

Zu guter Letzt wurden Probleme mit der Medienauswahl genannt, da Berater einerseits das Mobiltelefon als Standardmedium bevorzugen, obwohl dies hohe Kosten verursacht, und andererseits auch für komplexere Kommunikationsprobleme E-Mail als Medium der Wahl einsetzen, obwohl eigentlich reichhaltigere Medien (wie Telefon) zum Einsatz kommen könnten. Der letzte Aspekt kann wiederum eng mit der Nichterreichbarkeit über synchrone Medien zusammenhängen. Die aus den Interviewanalysen erarbeiteten Problembereiche (siehe Tab. 1) dienen als Input für die im Weiteren beschriebene Konzeption und Umsetzung des RTC-Prototyps.

Kommunikationsproblem	Resultierende Anforderung	RTC-Lösungsansatz
Erfolgreiche Kommunikationsversuche (besonders in Stresssituationen) durch schlechte Erreichbarkeit von Kollegen	Transparenz für Erreichbarkeit schaffen	Ausgefeiltes Statuskonzept zur Abschätzung der voraussichtlichen Abwesenheit eines Nutzers, Dringlichkeitssignalisierung, Notifikationsfunktion
Unkenntnis des Mediums, über das ein Ansprechpartner aktuell erreichbar ist	Transparenz schaffen, Hilfe bei der Medienauswahl	Präsenzstatus auf Medienebene, Anzeige verfügbarer Medienklassen
Fachexperten müssen zeitnah erreicht werden.	Unterstützung bei Auffinden und Auswahl von Experten	Integration mit (vorhandener) Fall- oder Kompetenzdatenbank
Ansprechpartner zu Dokumenten oder Fällen können nicht ermittelt werden (bei Bedarf an Hintergrundinformation)	Unterstützung bei Suche und Kontaktaufnahme (unbekannter) Gesprächspartner	Anhängen des Präsenzstatus von Mitarbeitern an (Projekt-) Dokumente in Datenbanken
Störende Unterbrechungen in Stresssituationen	Unterstützung bei der Priorisierung und Koordination	Präferenzgruppenkonzept und Filterfunktionen
Wahl eines ungeeigneten Mediums für die Kommunikation	Höhere Medienreichhaltigkeit gewährleisten	Integration mehrerer Medienklassen ins System (inkl. Video)
Mobiltelefone werden genutzt, obwohl günstigere Alternativen verfügbar sind	Unterstützung bei der Medienauswahl	Endgeräteauswahl je Medienklasse durch das System unterstützt

Tab. 1: Im Szenario identifizierte Probleme, resultierende Anforderungen und entsprechende Prototypfunktionen

6 Prototypische Umsetzung der Echtzeitkommunikationslösung

6.1 Eignung von RTC-Systemen im gewählten Anwendungsszenario

Die oben präsentierten Problembereiche bei der Kommunikation im Unternehmensberatungskontext bestätigten, dass sich RTC-Systeme grundsätzlich zum Einsatz in diesem Szenario verteilter Zusammenarbeit anbieten. Probleme wie erfolglose Kontaktversuche und die Unkenntnis, mit welchem Medium eine Person erreichbar ist, können durch die von RTC-Systemen bereitgestellten Funktionen potenziell gemildert werden. Gleichzeitig wird jedoch auch deutlich, dass hier kontextspezifisch ein komplexeres Verständnis von Präsenzsignalisierung und Erreichbarkeit notwendig ist. Ein einfaches Signalisieren von Erreichbarkeit von Seiten des Rezipienten ist nicht ausreichend, um den identifizierten Problem gerecht zu werden. Notwendig ist vielmehr ein komplexes, kontextspezifisches Erreichbarkeitsmanagement; hierzu müssen folgende Aspekte berücksichtigt werden:

- Eine Differenzierung der Erreichbarkeit nach verschiedenen Gruppen unterschiedlicher Priorität für den einzelnen Nutzer.
- Unterstützung bei der Abschätzbarkeit, wie lange ein Kontakt voraussichtlich nicht erreichbar ist, aufgrund der Dringlichkeit der Kommunikation.
- Eine Unterscheidung von Erreichbarkeit im Team und der Erreichbarkeit unbekannter Kontakte für Hintergrundinformationen zu Projekten.
- Die Behandlung des Trade-offs zwischen der Erreichbarkeit eines Nutzers und seinem eigenen Schutz vor ungewollten Störungen.

Es wird deutlich, dass RTC-Systeme in ihrer Standardausführung diesen Anforderungen nicht gerecht werden. Aus diesem Grund wurde eine eigene RTC-Lösung entwickelt. Hierzu wurde Siemens Openscape als Plattform ausgewählt, da das System seine gesamte Funktionalität über SDKs zur Verfügung stellt. Openscape dient dabei als Infrastruktur zur Umsetzung eines auf das Szenario abgestimmten Prototyps. Die wesentlichen Funktionen zum Management der Verfügbarkeit von Benutzern und zur Initiierung der Kommunikation über verschiedene Medien wurden innerhalb des Prototyps neu entwickelt, sie greifen aber auf Openscape-Dienste wie Endgeräte- und Statusverwaltung zu. Im Weiteren wird das als Prototyp entwickelte Consultant-Communication-System (COCOS) anhand der wesentlichen Konzepte und Funktionen beschrieben. In Tabelle 1 findet sich ein Überblick über die Zuordnung von Kommunikationsproblemen zu den im Weiteren beschriebenen Funktionen; Abbildung 1 zeigt einen Screenshot.

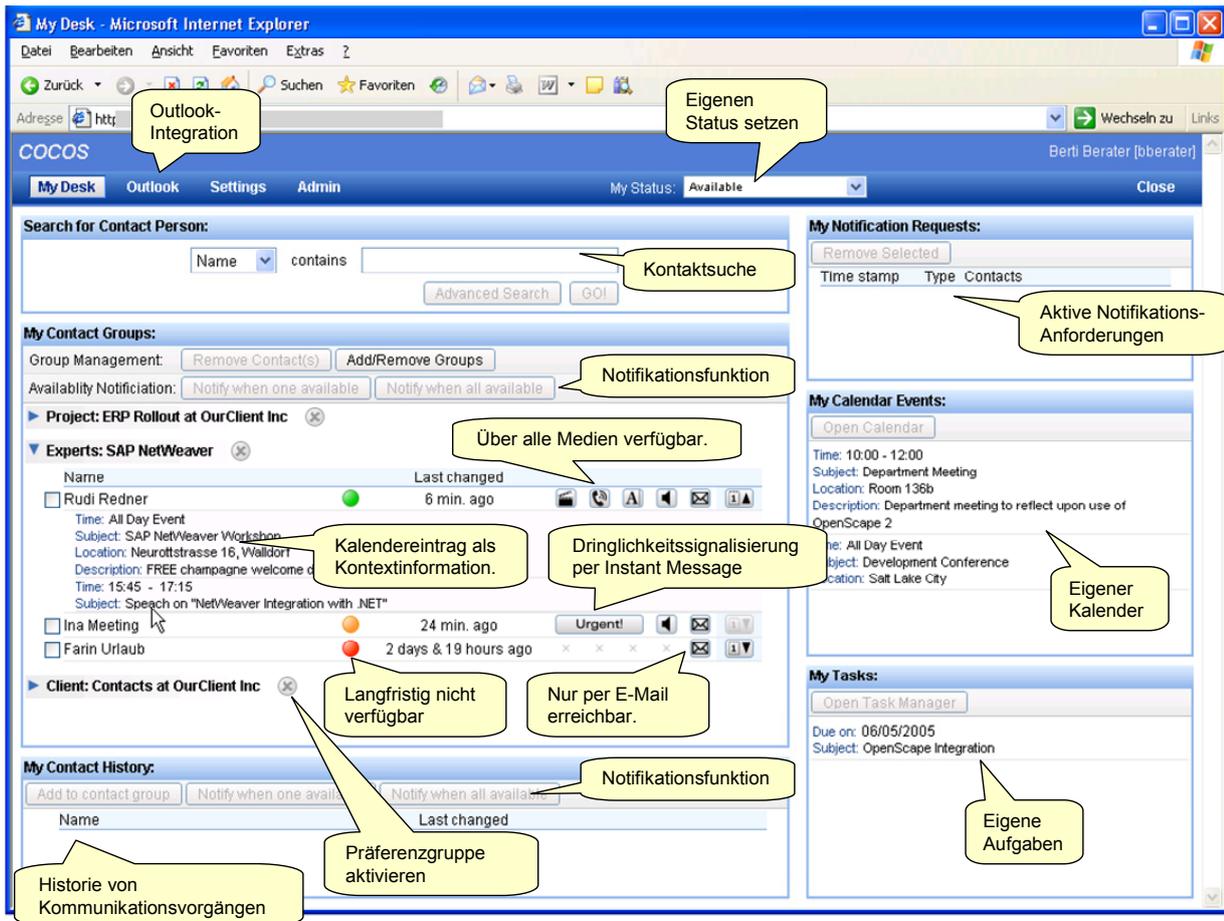


Abb. 1: Screenshot des COCOS-Prototyps²

6.2 Medienkonzept: Integration von Medientypen und Endgeräten

Ziel des Medienkonzeptes von COCOS ist es, mehrere Medien und Geräte in das System zu integrieren, um den Beratern die Kommunikation über Medien unterschiedlicher Reichhaltigkeit aus einem System heraus zu ermöglichen. Dies erstreckt sich sowohl auf synchrone Medien wie auch auf asynchrone. Einem Medientyp werden dabei mehrere Endgeräte zugeordnet, da es für den Initiator nicht von Interesse ist, welches Gerät der Empfänger gerade verwendet (z.B. Mobiltelefon), sondern lediglich, über welches Medium (z.B. Sprachkommunikation) er erreichbar ist (siehe Tab. 2). Der Benutzer hat die Möglichkeit, im System für jeden Medientyp ein präferiertes Gerät zu wählen oder dem System über Regeln die Auswahlentscheidung so zu übertragen, dass es stets für die Verbindung mit dem passenden Gerät sorgt. Ein solches Medienkonzept bietet darüber hinaus die Möglichkeit, einen differenzierten Verfügbarkeitsstatus des Benutzers auf der Basis verschiedener Medientypen zu ermitteln.

² Für eine Demonstration des Prototyps siehe: http://www-wi.uni-muenster.de/wi/forschen/COCOS_Demo.html

Synchrone Medien			Asynchrone Medien		
Video	Audio	Text	Video	Audio	Text
Video-Ausstattung (Webcam)	Festnetz-, Mobil-, IP-Telefon	Instant-Messaging, Chat	Video-Mail	Voice-Mail, Voicebox	E-Mail, SMS, Fax

Tab. 2: Kommunikationsmedien und deren zugeordnete Geräte

6.3 Kontaktgruppenkonzept: Abbildung von Teams und Expertengruppen im System

In der Unternehmensberatung wird ein Großteil der operativen Arbeit in Projektteams durchgeführt, innerhalb derer die Mitglieder eng zusammenarbeiten und daher häufig miteinander kommunizieren. Die Erreichbarkeit sowohl der Teammitglieder untereinander als auch externer Experten kann also kritisch für den Erfolg eines Projekts sein. Gleichzeitig ist es in zeitkritischen Phasen eines Projektes wichtig, die Kommunikation auf die Gruppe der Teammitglieder zu beschränken. Diesem Umstand wird innerhalb des COCOS-Prototyps durch das Kontaktgruppenkonzept Rechnung getragen. Sobald ein Berater einem neuen Projekt zugeteilt wird, werden alle Projektmitglieder standardmäßig als eine Kontaktgruppe für ihn angelegt. Der Nutzer kann zudem Kontaktgruppen nach seinen Wünschen gestalten, um zusätzlich Gruppen zu Themen und Kompetenzen zu unterhalten. So kann er Kollegen in eine Gruppe zusammenfassen, die Informationen zu wichtigen Arbeitsgebieten liefern können. Das Kontaktgruppenkonzept dient als Grundlage der unten beschriebenen Filter- und Notifikationsfunktionen.

6.4 Statusverwaltung und Signalisierung

Kern einer ERTC-Lösung ist das Signalisieren der Erreichbarkeit. Aus diesem Grund ist die Signalisierungsfunktion und Statusverwaltung durch den Benutzer ein wesentlicher Bestandteil des entwickelten Prototyps. Dabei wird im Weiteren synonym von Verfügbarkeit oder Erreichbarkeit, nicht jedoch von „Präsenz“ gesprochen, da der Begriff der „Präsenz“ die Anwesenheit eines Benutzers an einem Ort impliziert, was im vorliegenden Szenario jedoch nicht zielführend ist. So ist es für die Berater meist nicht relevant, wo sich ein Kollege aufhält, sondern nur, ob er für die Kommunikation tatsächlich zur Verfügung steht.

6.4.1 Statushierarchie

Die Statusverwaltung des COCOS-Prototyps basiert auf einer Statushierarchie; es wird zwischen einem technischen, einem gesetzten und dem angezeigten Verfügbarkeitsstatus unter-

schieden. Der technische Status bildet die Basis der Statusverwaltung: er zeigt an, ob ein Medium gerade verfügbar ist. Dabei wird dem Benutzer für jeden Medientyp ein eigener Status angezeigt, um dem Initiator die Entscheidung zu überlassen, auf welchem Medium ein Rezipient erreicht werden soll. Der technische Medienstatus hängt von der momentanen Verfügbarkeit der unter dem Medium zusammengefassten Geräte ab. Ist mindestens eines der Geräte eines Medientyps verfügbar, so zeigt der Status „Verfügbar“. Dabei überwacht das Openscape-System den Status der aller Geräte und ermittelt situativ den Status für die Medientypen. Dabei werden auch Kommunikationsaktionen des Benutzers einbezogen, so dass z.B. beim Telefonieren über das VoIP-Telefon der Verfügbarkeitsstatus der synchronen Medien „Video“ und „Audio“ für die Dauer des Gesprächs auf „Kurzfristig nicht verfügbar“ gesetzt wird. Der Benutzer selbst erhält die Möglichkeit, seinen persönlichen Verfügbarkeitsstatus zu setzen. Welcher Status einem anderen Benutzer schließlich angezeigt wird, ergibt sich dann aus einer Kombination der ersten beiden Status und einer zusätzlichen Filterfunktion (siehe Kap. 6.5.1).

6.4.2 Wahl der Statusausprägungen

Ziel bei der Entwicklung des Statuskonzeptes war es, die Anforderungen des Szenarios nach Verbesserung der kurzfristigen Verfügbarkeit und Abschätzbarkeit der voraussichtlichen Abwesenheit zu berücksichtigen. Als mögliche Status wurden daher die Ausprägungen „Verfügbar“, „Kurzfristig nicht verfügbar“, „Mittelfristig nicht verfügbar“ und „Langfristig nicht verfügbar“ gewählt, die bewusst auf ortsspezifische Angaben wie „Im Büro“ verzichten und sich auf den wesentlichen Aspekt der Verfügbarkeit und die zeitliche Dimension der voraussichtlichen Abwesenheitsdauer beschränken. Durch Verwendung unscharfer zeitlicher Begriffe wird die Notwendigkeit einer exakten Festlegung der Dauer der Nichtverfügbarkeit durch den Rezipienten vermieden. Stattdessen spiegeln die Status die subjektive Einschätzung des Nutzers wider, bei kurzer Nichtverfügbarkeit (z.B. während eine Meetings) signalisiert er „Kurzfristig nicht verfügbar“, bei Nichtverfügbarkeit für mehrere Stunden zeigt er „Mittelfristig nicht verfügbar“, ist der Benutzer den ganzen Tag oder mehrere Tage nicht verfügbar, so wählt er „Langfristig nicht verfügbar“. Ein solches Statuskonzept hat den weiteren Vorteil, dass es überschneidungsfrei ist. Openscape lässt standardmäßig durch Ausprägungen wie „Bin gleich zurück“, „In Besprechung“ und „Nicht im Büro“ keine eindeutige Interpretation der Verfügbarkeit zu.

In COCOS ist der angezeigte Verfügbarkeitsstatus eines Benutzers „Verfügbar“, wenn sein gesetzter Status aktuell auf „Verfügbar“ steht und mindestens eines seiner Geräte für synchrone Kommunikation den technischen Status „Verfügbar“ zeigt. Ist letzteres nicht der Fall, so wird

der Status „Mittelfristig nicht verfügbar“ angezeigt. Wird ein synchrones Medium gerade genutzt, so wird vom System „Kurzfristig nicht verfügbar“ signalisiert und die Icons der betroffenen Medientypen werden ausgeblendet. Ansonsten entspricht der angezeigte dem gesetzten Verfügbarkeitsstatus. Hinzu kommen die Auswirkungen der Filterfunktion.

6.4.3 Zusätzliche Statusinformationen

Um die Abschätzbarkeit der Nichterreichbarkeit eines Benutzers weiter zu verbessern, wurden dem Statuskonzept zwei wichtige Informationen hinzugefügt: solche zur Alterung von Statusinformation im Sinne von „Letzte Statusänderung erfolgt vor X min/std/tagen“, sowie die aktuellen Kalenderdaten des Rezipienten. Der Statusabrufer bekommt somit vor der Kontaktaufnahme nicht nur die Information, dass die gewünschte Person nicht verfügbar ist, sondern auch die Kontextinformation, wann mit einem erfolgreichen Kontaktversuch (kurz-/mittel-/langfristig) gerechnet werden kann, sowie ob die Präsenzinformation noch aktuell ist und was die Person zurzeit tatsächlich tut. Dadurch wird die Unsicherheit auf Seiten des Beraters, wann er die Hilfe des gewünschten Kollegen in Anspruch nehmen kann, deutlich gesenkt. Die alleinige Information, dass der Kollege nicht verfügbar ist, kann die Unsicherheit des Beraters nicht reduzieren.

6.5 Weitere kontextspezifische Funktionen zum Management der Erreichbarkeit

6.5.1 Filterfunktion

In zeitkritischen Phasen des Projekts muss sich ein Berater vollständig auf die Projektarbeit konzentrieren und möchte nicht durch Kontaktversuche unnötig unterbrochen werden. Eine traditionelle Verhaltensweise ist das Abschalten der Endgeräte, oder das Signalisieren von „Nicht verfügbar“. Da der Berater jedoch für unmittelbare Teammitglieder weiterhin kontaktfähig sein muss, ist eine Unterscheidung zwischen erwünschten und unerwünschten Kontaktversuchen notwendig. Aufbauend auf dem Kontaktgruppenkonzept kann der Nutzer in COCOS eine oder mehrere Gruppen als Präferenzgruppen definieren. Mittels eines Filters kann er dann bei Bedarf den angezeigten Verfügbarkeitsstatus differenzieren, so dass für alle Benutzer außerhalb der Präferenzgruppe der angezeigte Status „Mittelfristig nicht verfügbar“ erscheint, während die Präferenzgruppe den tatsächlichen Verfügbarkeitsstatus angezeigt bekommt.

6.5.2 *Notifikationsfunktion*

Mittels der Verfügbarkeitsnotifikation kann sich ein Nutzer auf Wunsch automatisch benachrichtigen lassen, sobald ein gewünschter Ansprechpartner wieder verfügbar ist („tell-me-when“). Die Verfügbarkeitsnotifikation kann auch auf mehrere Personen bzw. eine Gruppe bezogen werden; der Initiator erhält dann eine Information, wenn entweder eine Person der Gruppe oder alle Personen der Gruppe zur Verfügung stehen. Ersteres ist wichtig, wenn der Berater eine Information von einem Experten eines bestimmten Themenbereichs benötigt, letzteres kann z.B. die Initiierung einer Adhoc-Telefon- oder Videokonferenz unterstützen.

6.5.3 *Dringlichkeitssignalisierung*

In bestimmten Situationen reicht einem Initiator die Notifikationsfunktion nicht aus, da er nicht auf die Verfügbarkeit des Rezipienten warten kann. Wenn der Nutzer seinen Status auf „Nicht verfügbar“ gesetzt, trotzdem aber technisch im System verfügbar ist, kann die Nichterreichbarkeit durch Klick auf den Button „Urgent“ umgangen und ein direkter Kontakt ermöglicht werden. Es liegt auf der Hand, dass diese Funktion nur in Notfällen, genutzt werden sollte, da durch Missbrauch das gesamte Statusprinzip gefährdet wäre. Aus diesem Grund wurde diese Funktion in COCOS mittels einer vom Sender verfassten Kurznachricht (Instand Message oder SMS) dergestalt implementiert, dass der Empfänger entscheiden kann, ob er die Nachricht annimmt. Ein Dringlichkeitssignal führt in dem Fall nur zum Erfolg, wenn das Problem sowohl aus Sicht des Initiators als auch aus der Sicht des Rezipienten eine hohe Priorität besitzt.

6.5.4 *Integration mit Dokumenten- und Kompetenzdatenbank³*

Als ein weiteres Problem wurde das Kontaktieren von Ansprechpartnern zu bereits abgeschlossenen Projekten angeführt. Hier wäre denkbar, die Verfügbarkeitsinformationen von Benutzern direkt an jegliche Namensnennung innerhalb von Kompetenz- oder Falldatenbanken anzuhängen. Auf diese Weise könnte ein Berater direkt sehen, ob ein ehemaliger Projektteilnehmer gerade über synchrone Medien erreichbar ist. Hier ist eine Aggregation von Verfügbarkeitsstatus auf der Dokumentenebene denkbar. Zeigt der Verfügbarkeitsstatus für ein Dokument „Verfügbar“, dann ist mindestens ein Verfasser des Dokuments erreichbar. Eine solche Integration des Systems mit vorhandenen Datenbanken könnte die Informationssuche der Berater noch einmal verbessern, da sie das Kontaktieren bis dahin unbekannter Experten unterstützt.

³ Diese Funktion wurde lediglich konzeptioniert, aber nicht softwaretechnisch umgesetzt. Eine Integration auf Basis der gewählten Architektur wird jedoch durch den Einsatz von Webservices ermöglicht (siehe Kap. 6.6).

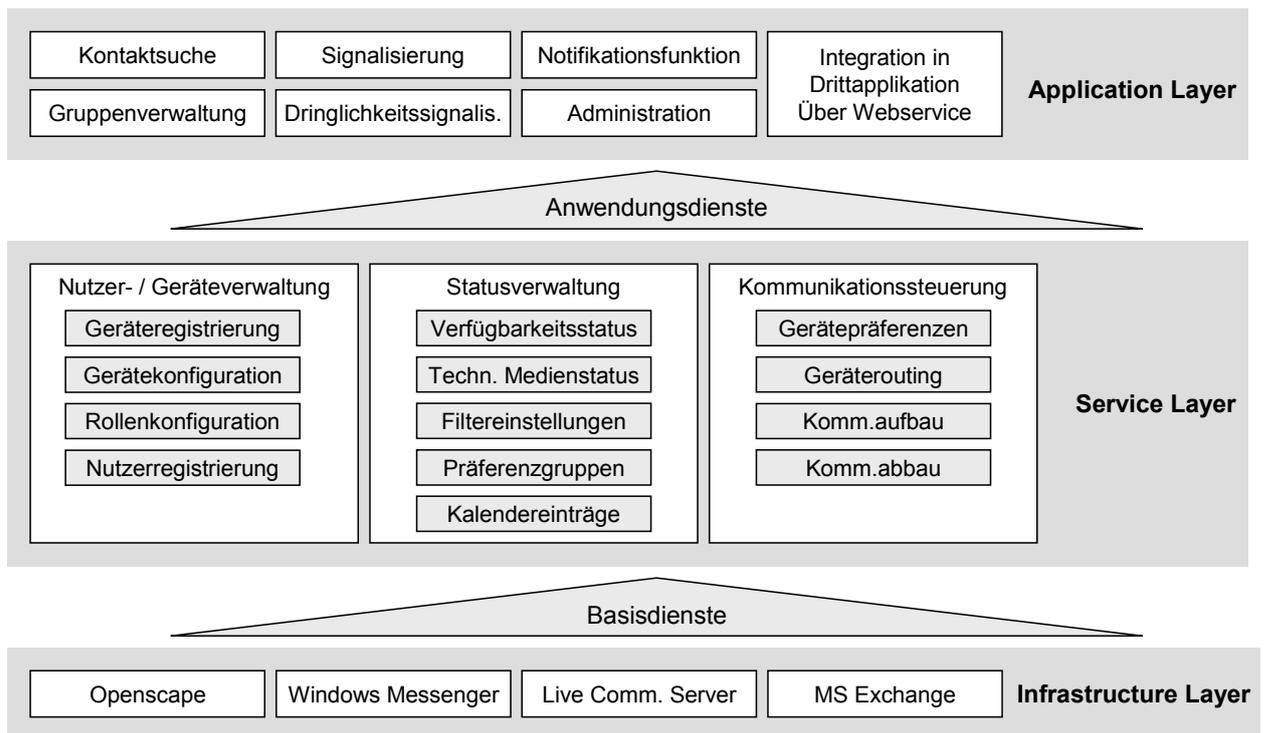


Abb. 2: Architektur des COCOS-Prototyps als Schichtenmodell

6.6 Architektur des Prototyps

Der COCOS-Prototyp basiert auf einer Drei-Schichten-Architektur (siehe Abb. 2). Der Infrastructure-Layer besteht aus Systemen, die Basisdienste für den Service-Layer bereitstellen. Siemens OpenScape stellt hierbei Webservices zum Auslesen und Setzen der Präsenz- und Gerätestatus, sowie zum Aufbau der Kommunikationsverbindungen zur Verfügung (mittels IP-Technik). Hierfür greift Openscape auf den Microsoft Live Communication Server und den Windows Messenger für das Instant-Messaging zu. Zu guter letzt wird Microsoft Exchange für Kalender- und E-Mail-Funktionalität genutzt. Der Service-Layer bildet den Kern des Prototyps; hier wurden die oben vorgestellten Konzepte und Funktionen implementiert, die dem Application-Layer in Form einer einheitlichen Webservice-Schnittstelle zur Verfügung stehen. Die Dienste sind in Nutzer- und Geräteverwaltung, Statusverwaltung und Kommunikationssteuerung unterteilt. Auf dem Application-Layer wurde ein browserbasiertes Demosystem entworfen, das dem Nutzer die beschriebenen Funktionen bereitstellt (siehe Screenshot). Da die Dienste des Service-Layers universell nutzbar sind, könnten sie auch in Dritt-Applikationen wie Groupwaresysteme und virtuelle Teamräume integriert werden [RiAW05].

7 Fazit und weiterer Forschungsbedarf

Die beschriebenen Konzepte und Funktionen des COCOS-Prototyps belegen, wie die in den Interviews identifizierten Kommunikationsprobleme des Szenarios mit Hilfe von RTC-Systemen technisch aufgegriffen werden können. Dabei wird deutlich, dass RTC-Systeme ihren Nutzen erst durch eine weitgehende Kontextualisierung auf ein spezifisches Anwendungsszenario werden entfalten können. Erst durch eine kontextspezifische Auslegung des Statuskonzepts, eine geeignete Anpassung der Medienklassen, sowie insbesondere die Umsetzung spezifischer Funktionen des Erreichbarkeitsmanagements kann der hier vorgestellte Prototyp dem Szenario funktional gerecht werden. Es wird deutlich, dass RTC-Systeme keine unmittelbar einsetzbaren Systeme „von der Stange“ sind. Sie benötigen vielmehr eine kontextspezifische Interpretation, sowie ggf. eine Integration mit bestehenden Anwendungen und Gerätelandschaften.

Die Umsetzung der hier skizzierten Vision eines Consultant-Communication-Systems mit Siemens Openscape kann den Anforderungen eines verteilten Unternehmensberatungsszenarios technisch zurzeit nur bedingt gerecht werden. Konkret ist zu bemängeln, dass OpenScape bisher nur sehr eingeschränkt portable Endgeräte unterstützt. Da die Berater jedoch eine klare Präferenz für die Nutzung mobiler Endgeräte (z.B. Mobiltelefone) geäußert haben, wird die Akzeptanz eines solchen Systems von der Verfügbarkeit der Kernfunktionen über mobile Geräte abhängen. Eine weitere Herausforderung besteht darin, die Nutzer bei der Pflege des Verfügbarkeitsstatus technisch zu unterstützen, damit die Akzeptanz nicht an einem unzumutbaren Pflegeaufwand scheitert. Hier ist z.B. die Nutzung von RFID-Chips technisch denkbar, die eine Statusänderung veranlassen, wenn der Nutzer ein Gebäude verlässt. Solche Techniken sind andererseits natürlich mit Privacy- und Datenschutzproblemen behaftet.

Auf der methodischen Seite macht die Studie deutlich, dass die von Hevner et al. beschriebenen Ansätze nicht als Dichotomie aufgefasst werden sollten. Gestaltungsorientierte Forschung zu anwendungsnahen IKT-Aspekten sollte nicht im „luftleeren Raum“ stattfinden, sondern eingebettet in verhaltensorientierte Erkenntnisse sein. Der hier vorgestellte Prototyp kann als kontextspezifische Interpretation der RTC-Idee zur Lösung von empirisch herausgearbeiteten Problemen verstanden werden. Dabei ist die durchgeführte Studie nicht im Sinne einer vollständigen Anforderungsanalyse für einen konkreten Fall zu sehen, dafür wäre die empirische Basis zu klein. Vielmehr ging es darum, einige typische Problembereiche und mögliche RTC-Lösungsansätze herauszuarbeiten. Während einige der identifizierten Aspekte dabei eher spezifisch für die Unternehmensberatungsbranche sind, lassen sich viele der Erkenntnisse vermutlich

auch auf andere Formen der Projektarbeit übertragen. Der empirische Kontext liefert für diese Art von Studie einerseits die notwendige Grundlage für die Gestaltung von IT-Artefakten durch das gewonnene Verständnis für die Sicht potenzieller Nutzer. Andererseits ist er das Ökosystem für die Forschung zu Nutzung und Effekten des Einsatzes von IT-Artefakten. Diese nächste Phase im Forschungsprozess zu RTC-Systemen setzt das Vorhandensein erster Implementierungen in Organisationen voraus. Hier stehen Fragen zur sozialen und organisatorischen Dimension im Mittelpunkt, z.B., ob die neu gewonnene Awareness von allen Mitarbeitern gleichermaßen positiv aufgenommen wird, welche Auswirkungen sich für die Kommunikations-, wie auch die Unternehmenskultur ergeben und ob die Verfügbarkeitsignalisierung mit Privacy-Bedenken der Beteiligten kollidiert [RiFr06]. Die Tatsache, dass sich die Systeme in einem frühen Stadium befinden, erlaubt die Erforschung ihres kompletten Lebenszyklus, angefangen von gestaltungsorientierten Vorhaben, wie in dieser Studie gezeigt, bis hin zu Nutzerexperimenten, Fallstudien und Langzeitstudien in Organisationen.

Literaturverzeichnis

- [Andr02] Andres, H. P.: A comparison of face-to-face and virtual software development teams. In: Team Performance Management: An International Journal 8 (2002) 1/2, S. 39-48.
- [deIB05] de Poot, H.; Mulder, I.; Kijl, B.: How do Knowledge Workers cope with their Everyday Job? In: eJOV - The Electronic Journal for Virtual Organizations and Networks 9 (2005), S. 70-88.
- [DoBe92] Dourish, P.; Belotti, V.: Awareness and Coordination in Shared Workspaces. In: Turner, J.; Kraut, R. (Hrsg.) Proc. of CSCW '92 - Sharing Perspectives. Assn for Computing Machinery, Toronto 1992, S. 107-114.
- [HMPR04] Hevner, A. R.; March, S. T.; Park, J.; Ram, S.: Design Science in Information Systems Research. In: MIS Quarterly 28 (2004) 1, S. 75-105.
- [KoHe02] Konradt, U.; Hertel, G.: Management Virtueller Teams - Von der Telearbeit zum Virtuellen Unternehmen. Beltz Verlag, Weinheim, Basel 2002.

- [Laza06] Lazar, I.: Integrating Telephony, IM, Video and Mobility with Presence. In: Business Communications Review (2006) June, S. 28-31.
- [LWCL01] Lau, T.; Wong, Y. H.; Chan, K. F.; Law, M.: Information technology and the work environment - does IT change the way people interact at work? In: Human Systems Management 20 (2001) 3, S. 267-279.
- [LyYo02] Lyytinen, K.; Yoo, Y.: Research Commentary: Next Wave of Nomadic Computing. In: Information Systems Research 13 (2002) 4, S. 377-388.
- [NaWB00] Nardi, B. A.; Whittaker, S.; Bradner, E.: Interaction and Outeraction: Instant Messaging in Action. In: Proceedings of the Computer Supported Cooperative Work. Philadelphia, 2000, S. 79-88.
- [ReGo05] Rennecker, J.; Godwin, L.: Delays and Interruptions: A Self-Perpetuating Paradox of Communication Technology Use. In: Information and Organization 15 (2005) 3, S. 247-266.
- [RiAW05] Riemer, K.; Arendt, P.; Wulf, A.: Marktstudie Kooperationssysteme - Von E-Mail über Groupware zur Echtzeitkooperation. Cuvillier, Göttingen 2005.
- [RiFr06] Riemer, K.; Fröbner, F.: Presence-based, Context-sensitive Real-Time Collaboration (RTC) - research directions for a new type of eCollaboration system. In: Proceedings of the 19th Bled eConference. Bled, Slovenia, 2006.
- [RiKl05] Riemer, K.; Klein, S.: Herausforderungen virtueller Organisation - Sozialkapital als Voraussetzung für das Funktionieren virtueller Organisationen. In: HMD - Praxis der Wirtschaftsinformatik (2005) 242, S. 21-31.
- [Scho02] Scholz, C.: Virtuelle Teams - Neuer Wein in neue Schläuche! In: zfo 71 (2002) 1, S. 26-33.
- [SoMK02] Soerensen, C.; Mathiassen, L.; Kakihara, M.: Mobile Services: Functional Diversity and Overload. In: Proceedings of the New Perspectives On 21st-Century Communications. Budapest, Hungary, 2002, S. 1-12.

Audio vs. Chat bei Aufgaben mit Unsicherheit: Die Produktivität folgt anderen Regeln als bei mehrdeutigen Aufgaben

Andreas Löber, Gerhard Schwabe

Institut für Informatik
Universität Zürich
8050 Zürich
{loeber, schwabe}@ifi.unizh.ch

Abstract

Chatkommunikation ist inzwischen weit verbreitet im privaten Umfeld und wird zunehmend populär auch im geschäftlichen Bereich. Audiokommunikation bietet dank VoIP neue Möglichkeiten der Gruppenkommunikation. Doch die Auswirkung der Medienwahl zwischen beiden Medien ist weitgehend unbekannt. Auch die Effekte der Gruppenkommunikation über diese Medien sind ungeklärt. Ausgehend von den etablierten Theorien stellt dieser Text die Ergebnisse eines Experiments zur Untersuchung von Produktivität und Zufriedenheit vor. In diesem mussten mit Audio und Chat und unterschiedlichen Gruppengrößen kritische Hinweise ausgetauscht werden, um einen Mordfall zu lösen. Aufgrund der Resultate wird ein Brückenschlag zwischen den Theorien vorgenommen und neue Faktoren für die Produktivität identifiziert.

1 Einleitung

Chatkommunikation findet immer mehr Verwendung sowohl im privaten [GrPa02, NaWB00], als auch betrieblichen Umfeld [Mull03]. Hier trifft sie auf die seit langem etablierte Sprachkommunikation. Dort bietet Voice over IP neue Möglichkeiten für günstige Konferenzschaltungen mit einer beliebig hohen Anzahl Nutzer ohne die Notwendigkeit von Spezialausrüstung. Dabei nimmt sowohl die Nutzung von Computerprogrammen wie Skype oder SIP-kompatiblen Standards als auch die Verbreitung von transparenten Anschlüssen, die sich wie normale Telefonverbindungen verhalten, stetig zu. Beide Medien bieten den Nutzern die Möglichkeit einfach und günstig in Gruppen zu kommunizieren, was vorher nur mit hohem Aufwand und hohen technischen Anforderungen möglich war.

Aber diese neuen Möglichkeiten erfordern vom Nutzer auch neue Entscheidungen. Welches der beiden Medien soll wann genutzt werden, um eine möglichst produktive und zufrieden stellende Zusammenarbeit zu erreichen? Für diese Fragestellung gibt es erstaunlicherweise fast keine Untersuchungen und damit keine belastbaren Antworten. Für ähnliche Medienwahlentscheidungen (z.B. Video vs. Face-to-Face) wird üblicherweise auf die Media Richness Theorie [DaLe86, DaLT87] zurückgegriffen, für die Messung der Zufriedenheit auf das Social-Influence-Model [FuSS90]. Für die Auswirkungen der Veränderung der Gruppengröße ist die Media Synchronicity Theorie [DeVa99] einschlägig. Die wenigen existierenden Experimente sind jedoch durch die technische Weiterentwicklung überholt. Im Rahmen zweier Experimente [Müry05, Grim06] haben wir die Auswirkung der Medienwahl auf die Produktivität (Leistung pro Zeit) und Zufriedenheit von Vierer-Gruppen untersucht und dabei eine Überlegenheit von Audio festgestellt. In einem Folgeexperiment [LöSc07], haben wir gezeigt, dass eine Vergrößerung der Gruppe auf sieben Personen für Designaufgaben Chatgruppen produktiver und Audiogruppen weniger produktiv werden lässt. Es erhebt sich nun die Frage, ob diese Produktivitätsveränderung spezifisch für mehrdeutige Aufgaben ist oder ob sie auch bei Aufgaben unter Unsicherheit beobachtbar sind. Anschließend an dieses Kapitel werden kurz die verwendeten Theorien zur Medienwahl erläutert. Darauf folgt eine Zusammenfassung des Experimentaldesigns und vorangegangener Ergebnisse. Das vierte Kapitel präsentiert neue Ergebnisse zu Aufgaben unter Unsicherheit für Vierer- und Siebener-Gruppen. Im fünften Kapitel werden die Gründe für die Unterschiede in den Beobachtungen untersucht. Der Text schließt mit einem Fazit und den Empfehlungen für weitere Arbeiten.

2 Theoretische Grundlagen

2.1 Medienwahl in kleinen Gruppen

Für die Wahl der Medien bei vorgegebenen Aufgaben schlägt die Media Richness Theorie von Daft und Lengel [DaLe86] vor, das richtige Maß an Reichtum zu wählen. Je mehrdeutiger eine Aufgabe ist, desto reicher muss das Medium sein. Dabei wird jedem Medium ein Reichtum zugeordnet, der sich an vier Faktoren orientiert [DaLT87]: Vielfältigkeit der vermittelten Sprache, Anzahl der hinweisgebenden Kanäle, Maß an Persönlichkeit und Unmittelbarkeit des Feedbacks. Die Media Richness Theorie ist geeignet, um die Auswirkungen der Aufgabe auf die Produktivität von Kleingruppen zu beschreiben.

Das Social-Influence-Model von Fulk et al. [FuSS90] beschreibt die Medienwahl hingegen als Entscheidung die nicht nur aufgrund von Mediencharakteristiken getroffen wird, sondern auch auf sozialen und subjektiven Faktoren beruht. Dabei spielen neben den eigenen Erfahrungen auch die Ansichten, Trends und Verhaltensmuster anderer Nutzer und vor allem Gruppenmitglieder eine Rolle. Das Social-Influence-Modell ist geeignet, um die subjektive Zufriedenheit der Nutzer zu beschreiben.

Bisher gibt es relativ wenige Studien über den Vergleich von Audio und Chatkommunikation und diese fokussieren sich vor allem auf Zweiergruppen. Die Studien haben sich auf die Media Richness Theorie als Erklärungsansatz gestützt, widersprachen sich jedoch zum Teil in den Ergebnissen. Kinney und Watson [KiWa92], sowie Suh [Suh99] zeigten in ihren Studien, dass Chatgruppen unabhängig von der Art der Aufgabe längere Zeit zur Lösung der Aufgabe brauchten als Audiogruppen. Es gab keinen Unterschied in der Zufriedenheit. Die Studie von Valacich et al. [Vala94] widerspricht dieser Erkenntnis in sofern, dass die Chatgruppen bei seinem Experiment schneller waren als die Audiogruppen. Dabei zeigten sie jedoch deutlich niedrigere Zufriedenheit als die Audiogruppen. Bos et al [Bos02] untersuchten Gruppen mit drei Mitgliedern und zeigten dort, dass Audiogruppen ein Problem mit hoher Mehrdeutigkeit besser lösen konnten als Chatgruppen. Im Experiment von Graetz et al. [Grae98] brauchten Chatgruppen mit vier Teilnehmern für eine Aufgabe, die Unsicherheit und Mehrdeutigkeit kombiniert, länger als die Audiogruppen. Jedoch ist die Aufgabe nicht für alle Produktionseffekte verantwortlich. Zusammenfassend lassen sich aus den Studien bisher keine Konsens über die Medienwahl zwischen Audio und Chat herleiten. Die existierenden Studien

sind zudem recht alt und alle vor der Zeit entstanden, in der sich durch die Verbreitung von Skype sowohl Chatkommunikation als auch Audiokonferenzen stark verbreitet haben.

2.2 Medienwahl in größeren Gruppen

Im Rahmen von Untersuchungen traditioneller Brainstorming-Sitzungen wurde von Diehl und Ströbel [DiSt87, DiSt91] und Mullen et al. [MuJS91] herausgefunden, dass die Erweiterung von Gruppen um zusätzliche Mitglieder sich negativ auf die Produktivität auswirken kann. Dies ist bedingt durch die Produktionsblockade, die entsteht, wenn Gruppenmitglieder durch die Äußerung anderer gezwungen werden auf ihren Beitrag zu warten. Nunamaker et al. [NDVV91] und Gallupe et al. [Gall92] zeigten, dass durch elektronische Mittel dieser Effekt behoben werden kann, so dass zusätzliche Mitglieder einen Produktivitätsgewinn bringen. Wesentliche Faktoren für diese Produktivitätsgewinne sind in der Media Richness Theorie und das Social-Influence Modell nicht enthalten. Deshalb können sie zwar für Erklärungen für die Medienwahleffekte bei kleineren Gruppen verwendet werden, sind jedoch mit steigender Gruppengröße nur eingeschränkt gültig. Die Media Synchronicity Theorie erweitert deshalb die Media Richness Theorie um Faktoren der Gruppenarbeit. Medien werden anhand von 5 Faktoren charakterisiert: Geschwindigkeit des Feedbacks, Symbolvarietät, Parallelität, Überarbeitbarkeit und Wiederverwendbarkeit. Einige der Faktoren, wie z. B. Geschwindigkeit des Feedbacks und Parallelität schließen einander aus. Dahingehend gilt es, die für die Aufgabe wichtigsten Medieneigenschaften zu identifizieren, und Medien gezielt aufgrund dieser Bedürfnisse zu wählen. Wir verwenden die Faktoren der Media Synchronicity Theorie, um die Auswirkung der verschiedenen Gruppengrößen zu erklären. Faktoren wie die Parallelität und Geschwindigkeit des Feedbacks sind alleine kaum zu quantifizieren. Letztendlich kommt es auf den Einzelfaktor auch nicht so sehr an, sondern eher auf die gesamte Mediengeschwindigkeit, d.h. die Frage, wie viele Informationen pro Zeit insgesamt übertragen werden können. Dies beinhaltet dann auch Einzelfaktoren wie Parallelität und Feedbackgeschwindigkeit. Dabei ist die Mediengeschwindigkeit vergleichsweise einfach zu messen, ist aber in der erstaunlicherweise in der internationalen Literatur für die Medienwahl bisher nicht diskutiert. Lediglich in der deutschsprachigen Forschung führten Picot und Reichwald [PiRe85] diesen Faktor ein. In [LöSG06] konnte gezeigt werden, dass sich Audio und Chat wesentlich voneinander in Hinblick auf die Mediengeschwindigkeit unterscheiden.

3 Bisherige Resultate und Hypothesen

Um die Auswirkung der Medienwahl von Audio und Chat und den Einfluss der Gruppengröße auf den Kommunikationsprozess von kooperativen Gruppen zu untersuchen, haben wir im November 2004 (Vierergruppen) und 2005 (Siebenergruppen) zwei Versuche [Müry05, Grim06] mit insgesamt 440 Teilnehmern durchgeführt. Dabei griffen wir auf zwei bewährte Experimentalaufgaben zurück, um die Ergebnisse mit anderen Studien vergleichbar zu machen. Als mehrdeutige Aufgabe wurde eine Designaufgabe gewählt, und zwar das Design eines automatischen Postamts der Zukunft von Olson et al. [OOSC93]. Als Aufgabe unter Unsicherheit wurde der Murder Mystery Kriminalfall von Stasser und Steward [StSt92] verwendet.

3.1 Bisherige Resultate

3.1.1 Designaufgabe

Die Designaufgabe forderte die Konzipierung eines automatischen Postamts. Durch die offene Fragestellung war sie eine Aufgabe mit hoher Mehrdeutigkeit, denn es hängt von den Präferenzen der Teilnehmer ab, wie das Design aussieht. Die Ergebnisse sind im Detail in [LöSc07] besprochen und sollen hier nur kurz zusammengefasst werden.

Bei den Gruppen mit 4 Teilnehmern schnitten die Audiogruppen signifikant besser ab als die Chatgruppen. Gleichzeitig waren auch die Audiogruppen signifikant zufriedener mit ihrem Medium als die Chatgruppen. Mit der auf sieben steigenden Teilnehmerzahl kehrte sich das Bild um. Während die Audiogruppen an Leistung verloren, gewannen die Chatgruppen mit den zusätzlichen Teilnehmern an Produktivität und erreichten faktisch das gleiche Niveau der Audiogruppen. Die Untersuchungen konnten klar zeigen, dass durch die Wahl des Mediums in Abhängigkeit zur Gruppengröße die Produktivität stark beeinflusst werden kann. Die Gruppen mit sieben Mitgliedern zeigten eine signifikant niedrigere Zufriedenheit als die Gruppen mit vier Teilnehmern. Bei der Verwendung von Chat trat dieser Effekt seltener auf, so dass bei Gruppen mit sieben Mitgliedern Chatgruppen mit ihrem Medium eine signifikant höhere Zufriedenheit zeigten als Audio nutzende Gruppen.

3.1.2 *Kriminalfall*

Anderen Vierer-Gruppen wurde der Kriminalfall als Aufgabe gestellt. Bei diesem geht es darum, Unsicherheit durch den Austausch von Informationen aufzulösen. Die Aufgabe hat aber eine eindeutig richtige Antwort die nicht von den Präferenzen der Teilnehmer abhängig ist. Die Aufgabe erfordert die Identifikation des Mörders aus einer Gruppe von 3 Verdächtigen. Dazu erhalten die Versuchsteilnehmer Hinweise in einer Dokumentensammlung, die aus Verhörprotokollen, einem Brief und einer Karte besteht. Jeder Teilnehmer erhält diese allgemeinen Informationen. Darüber hinaus gibt es 9 kritische Hinweise die zur eindeutigen Identifikation des richtigen Täters dienen. Diese kritischen Hinweise werden auf 3 Teilnehmer aufgeteilt, so dass jeder 3 kritische Hinweise hat. Alle anderen Teilnehmer erhalten keine kritischen Hinweise, sondern nur die allgemeinen Tatsachen. Klar formulierte Aufgabe der Teilnehmer war es also, durch Austausch des Wissens alle kritischen Hinweise zu einem gesamtheitlichen Bild zu konkretisieren und damit den Täter bestimmen zu können. Die Teilnehmer wurden angehalten, primär den richtigen Täter zu ermitteln und erst sekundär möglichst schnell zu einem Ergebnis zu kommen.

Bei den Experimenten mit Viergruppen konnte festgestellt werden [Müry05], dass Chat und Audiogruppen die gleiche Produktivität aufzeigten; sie fassten die gleiche Anzahl an Mördern. Gleichzeitig zeigten die Audiogruppen eine signifikant höhere Zufriedenheit mit dem Kommunikationsmedium als die Chatgruppen. Ferner konnte bei für die Gruppen mit vier Teilnehmern in [LöSG06] für beide Aufgaben festgestellt werden, dass die Audiogruppen mehr Informationsfragmente pro Zeiteinheit übermittelten. Außerdem übertrugen sie signifikant mehr kritische Informationen als die Chat-Gruppen. Diese zeigten jedoch ein höheres Verhältnis von kritischen Informationen bezogen auf alle Informationsfragmente. Letztere Erkenntnisse trafen auch für den Designfall zu. Jetzt stellt sich die Frage, ob eine Erhöhung der Gruppengröße die gleichen Auswirkungen auf Produktivität der Gruppen wie bei der Designaufgabe hat.

3.2 **Hypothesen**

Ausgehend von den Erkenntnissen über die Vierergruppen und die bekannten Effekte der Gruppengröße aus dem Vergleich von Vierer- und Siebnergruppen bei der Designaufgabe formulieren wir im Folgenden die Hypothesen.

Lange Wartezeiten auf die Möglichkeit eigene, vielleicht einzigartige Informationen, an die Gruppe zu geben führen dazu, dass die Zufriedenheit mit dem Medium sinkt. Audio wird durch die Verfügbarkeit des einen Sprachkanal beschränkt. Und obwohl Chatkommunikation mühsamer ist als die Nutzung von Audio, dürfte die jederzeitige Verfügbarkeit der Sendefunktion deutlich höhere Zufriedenheit erzeugen. Auch wird sich, gemäß dem Social-Influence-Modell die Gruppe als ganze beeinflussen in ihrer Sichtweise des Mediums. Dies führt bei negativen Erfahrungen mit dem Medium zu deutlicheren Abneigungen.

H7.1: Chatgruppen zeigen mehr Zufriedenheit als Audiogruppen bei Gruppen mit sieben Mitgliedern.

Die Gruppen mit vier Mitgliedern zeigten identische Erfolgsquoten bei der Identifikation der Täter. Bei der Hinzufügung von weiteren Gruppenmitgliedern verhalten sich die Medien jedoch unterschiedlich. Bei Audio können die weiteren Teilnehmer kaum beitragen und ihre Beiträge blockieren automatisch alle anderen Gruppenmitglieder. Dies ist nicht so beim parallelen Medium Chat. Hier kann jeder Teilnehmer schreiben wann und wie er will. Dementsprechend postulieren wir:

H7.2: Chatgruppen zeigen mehr Produktivität als Audiogruppen bei Gruppen mit sieben Mitgliedern

Zusätzliche Gruppenmitglieder bedeuten mehr Aufwand, was sich grundsätzlich negativ auf die Zufriedenheit auswirkt. Bei den Gruppen mit vier Mitgliedern waren die Audiogruppen deutlich zufriedener als die Chatgruppen. Jedoch wird mit jedem zusätzlichen Mitglied die Redezeit kürzer und der Frust damit größer. Die Hindernisse des Chat Mediums sind sehr viel weicher. Falls zuviel Text auf einmal kommt, kann dieser überflogen werden, um die Kernpassagen zu entnehmen, was deutlich weniger stört als zur Untätigkeit verdammt zu sein. Dahingehend postulieren wir:

H4vs7.1: Audiogruppen verlieren mehr Zufriedenheit aufgrund der Erweiterung der Gruppe als Chatgruppen

Da das Audiomedium eine maximale Sprechzeit aufweist, gibt es eine Obergrenze an möglichen Informationen, die in dieser Zeit gesendet werden können. Diese bleibt konstant, so dass mit jedem weiteren Mitglied die Konkurrenz um die Zeit ansteigt. Dies führt zu mehr Aufwand für Absprachen wer wann etwas sagen darf. Chat hat jedoch diese Einschränkung beim Senden nicht. Alle Teilnehmer können gleichzeitig schreiben. Dies kann jedoch dazu führen, dass die Mitglieder nicht mehr in der Lage sind mit dem Lesen hinterherzukommen,

was ihre Produktivität einschränkt. Jedoch kann der Mensch deutlich schneller lesen als er hören kann [Will79]. Dementsprechend postulieren wir:

H4vs7.2: Chatgruppen gewinnen mehr Produktivität aufgrund der Erweiterung der Gruppe als Audiogruppen

3.3 Methode und Datenerhebung

Um die Auswirkung der Aufgabe auf die Medienwahl zu untersuchen wurden Experimente mit 10 Gruppen à 4 Teilnehmern mit jeweils einem der beiden Medien und einer der beiden Aufgaben durchgeführt (160 Teilnehmer). Um die Auswirkung der Veränderung der Gruppengröße zu untersuchen wurden die gleichen Versuche anschließend mit jeweils 7 Teilnehmern durchgeführt (280 Teilnehmer). Die beiden Experimente fanden mit insgesamt 440 bezahlten Freiwilligen statt, die aus dem Studentenkreis rekrutiert wurden, ohne dass es größere Besonderheiten in Hinblick auf Studienrichtung oder Geschlecht gab. Die Teilnahme an beiden Durchgängen war nicht möglich. Alle Teilnehmer konnten Eingangs die Fallbeschreibung des Kriminalfalls für 25 Minuten lesen, durften während dieser Zeit jedoch nicht sprechen. Anschließend wurden sie in separate Räume geführt, wo jeweils ein X30 Notebook, ein hochwertiges Headset und eine Maus genutzt werden konnte. Die Medienzweisung wurde ausgelost. Für die Chatgruppen stand Microsoft Netmeeting und die Audiogruppen Skype bzw. Teamspeak zur Verfügung. Die Aufgabe musste innerhalb von 30 Minuten gelöst werden, wobei die Teilnehmer darauf hingewiesen wurden, dass die Identifikation des richtigen Mörders die Hauptpriorität war.

Zur Erfassung der Zufriedenheit der Nutzer wurde der SUS Fragebogen [Broo96] benutzt, bei dem mittels 10 Fragen auf einer Lickertskala und einer anschließenden Berechnung die Gesamtzufriedenheit auf einer Skala von 0 bis 100 ermittelt wird. Die Produktivität wurde berechnet, indem Leistung pro Zeiteinheit gemessen wurde. Dabei ist die Leistung beim Kriminalfall als die richtige Identifikation des Mörders definiert, wobei die Beschuldigung eines Unschuldigen als keine Produktivität gemessen wurde. Die Zeit wurde vom Experimentalleiter gestoppt und später nochmals mit Analysen der Kommunikationsprotokolle überprüft. Einheit der Produktivität ist somit ermittelte Mörder pro Zeiteinheit (hier Minute). Statistisch verändert dies nichts gegenüber der benötigten Zeit von Gruppen, die den Mörder erfolgreich identifiziert haben. Statistische Untersuchungen finden als zweiseitige T-Tests für die konstanten Gruppengrößen. Zum Vergleich von Medium und Gruppe wurden, wie im Originalexperiment

von Stasser und Steward auch, 2x2 faktorielle Varianzanalyse-Tests (ANOVA) gewählt, um die Interaktionseffekte zwischen Medium und Gruppengröße zu berechnen.

3.4 Einschränkungen

Die Bedeutung der Gruppengröße für den Kommunikationsprozess und die Produktivität wurde erst im Rahmen der Nachuntersuchung des ersten Experiments mit Vierergruppen offensichtlich. Deshalb wurde das Experiment mit den Siebenergruppen erst im Folgejahr durchgeführt, was aufgrund der hohen Anforderungen an Teilnehmerzahlen auch praktisch nicht anders umsetzbar gewesen wäre. Dies könnte zu besseren Vorkenntnissen der Versuchsteilnehmer führen. Die im SUS Fragebogen abgefragte Benutzerfreundlichkeit zeigte jedoch keine großen Unterschiede auf (4,4 von 5 Punkten in 2004, 4,34 in 2005). Ferner musste aufgrund der Gruppengröße in den Tests mit sieben Teilnehmern Skype gegen Teamspeak getauscht werden, da Skype keine Unterstützung von Gruppen mit mehr als 5 Mitgliedern anbietet. Die Unterschiede in Funktionalität und Qualität sind jedoch in einem lokalen Netzwerk vernachlässigbar.

Eine weitere bewusste Einschränkung war die Begrenzung auf die beiden dargelegten Aufgabentypen. Obwohl solch klar getrennte Aufgabentypen in der Praxis selten vorkommen mögen und es sicherlich auch noch Gruppenaufgaben gibt, die jenseits der Unsicherheit und Mehrdeutigkeit liegen, glauben wir doch, dass die meisten in Gruppen bearbeiteten Aufgaben Ähnlichkeit mit einer der untersuchten Experimentalaufgaben aufweisen.

Ferner wurde die Anzahl von kritischen Hinweisen im Kriminalfall konstant gehalten bei der Erhöhung der Gruppengröße. Neben der Aufrechterhaltung der Vergleichbarkeit als Experimentalanforderung verfügen oftmals auch in echten Kommunikationsfällen nicht alle Mitglieder über eigenständige Informationen.

4 Ergebnisse der Versuche über die Auswirkung der Veränderung der Gruppengröße

4.1 Gruppen mit sieben Mitgliedern

Die Zufriedenheit mit dem Medium ist nicht signifikant unterschiedlich ($t(18)=1,324$; $p=0,202$) bei Gruppen mit sieben Mitgliedern. Die Chatgruppen haben einen Mittelwert von 82,1 (Standardabweichung 4,02), während die Audiogruppen einen Mittelwert von 79,14

(Standardabweichung 5,83) haben. Damit ist auch Hypothese H7.1, die postuliert, dass die Chatgruppen zufriedener mit ihrem Medium als die Audiogruppen sind, widerlegt.

Sowohl die Chat als auch Audiogruppen identifizierten jeweils 6 Verdächtige richtig und lagen in 4 Fällen daneben. Dahingehend gibt es auch keinen signifikanten Unterschied in der Produktivität ($t(18)=-0,084$, $p=0,934$). Damit ist die Hypothese H7.2, die besagt, dass Chatgruppen mehr Produktivität als die Audiogruppen zeigen, widerlegt.

4.2 Vergleich der Gruppen mit vier und sieben Mitgliedern

Es gibt einen großen Effekt bei der Zufriedenheit, ausgelöst von der Veränderung der Gruppengröße. Während das Medium alleine keinen signifikanten Einfluss hat ($F(1,36)=0,124$; $p=0,727$), ist die Gruppenerweiterung alleine ($F(1,36)=18,997$; $p<0,01$) für 34,5 % der Varianz verantwortlich. Zudem ist die Kombination aus Gruppenerweiterung und Medium für weitere 12,4 % verantwortlich ($F(1,36)=5,087$; $p=0,03$). Wie in Abbildung 1 zu sehen ist, kann also von der Gruppengröße ausgehend die Medienwahl sehr wichtig für die Zufriedenheit sein. Chatgruppen verlieren dabei signifikant weniger an Zufriedenheit als Audiogruppen. Damit ist Hypothese H4vs7.1 bestätigt.

Bei den Gruppen mit vier Mitgliedern waren jeweils 7 Fälle richtig und 3 falsch gelöst, während bei den Gruppen mit sieben Mitgliedern es jeweils 6 richtige

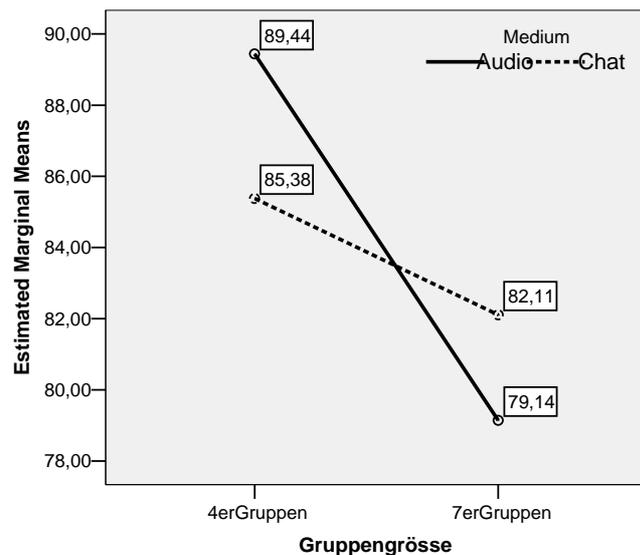


Abbildung 1: Abhängigkeit der Zufriedenheit von Medium und Gruppengröße

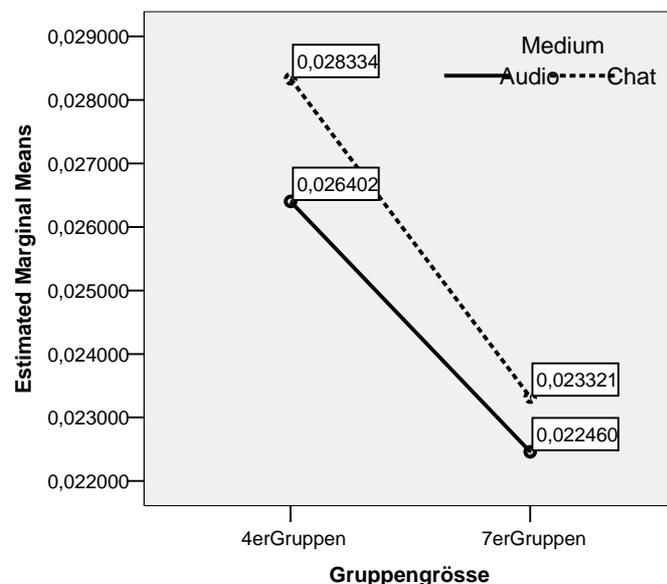


Abbildung 2: Abhängigkeit der Produktivität von Medium und Gruppengröße

und 4 falsch gelöste Fälle waren. Dieser Unterschied ist nicht signifikant, auch wenn man die Zeit mit einbezieht, die für die richtigen Lösungen gebraucht wurde. Sowohl die Analyse der Auswirkung der Gruppengröße ($F(1,36)=0,517$; $p=0,477$), als auch des Mediums ($F(1,36)=0,05$; $p=0,824$) oder der Kombination aus Gruppengröße und Medium ($F(1,36)=0,07$; $p=0,932$) ergaben keinerlei signifikante Unterschiede. Damit ist Hypothese H4vs7.2 widerlegt, die besagt, dass die Chatgruppen stärker als die Audiogruppen an Produktivität gewinnen mit zusätzlichen Gruppenmitgliedern. Zusammenfassend kann man sagen, wie in Abbildung 2 dargestellt, dass beide Medien jeweils ähnliche Produktivität zeigten und der Abfall der Produktivität von vier auf sieben Gruppenmitgliedern nicht signifikant war.

5 Diskussion

5.1 Zufriedenheit

Die Chatgruppen mit sieben Mitgliedern waren nur geringfügig zufriedener als die Audiogruppen. Damit haben die Chatgruppen jedoch im Schnitt weit weniger Zufriedenheit verloren als die Audiogruppen, die bei einer Gruppengröße von vier Personen eine deutliche höhere Zufriedenheit zeigten. Die Untersuchung der Effekte der Gruppenerweiterung hat klar gezeigt, dass dies hauptsächlich mit der Gruppengröße zu tun hat. Je mehr Teilnehmer eine Gruppe hatte, desto unzufriedener wurden die Mitglieder. Jedoch lässt sich durch die richtige Auswahl des Mediums dieser Effekt zumindest abschwächen. Allgemein lässt sich sagen, dass die Aufgabe keinen Einfluss auf die Zufriedenheit hat. Dahingehend sind die Ergebnisse nicht überraschend. Mitglieder von Gruppen mit mehr Teilnehmern sind deutlich unzufriedener als Teilnehmer kleiner Gruppen. Durch eine geschickte Auswahl des Mediums kann jedoch die Zufriedenheit signifikant beeinflusst und gesteigert werden. Bei wenigen Teilnehmern sollte Audio als Medium gewählt werden. Hier bietet das Medium mit seiner gewohnten Nutzung die beste Unterstützung für die Zufriedenheit der Teilnehmer. Sollte die Kommunikation mehr als fünf Teilnehmer erfordern, dann ist Chatkommunikation in jedem Fall das Medium, welches die Nutzer mehr zufrieden stellt.

5.2 Produktivität

Entgegen unserer Vermutungen gab es keinerlei Produktivitätsunterschiede bei den Gruppen mit sieben Mitgliedern. Auch die Erweiterung um zusätzliche Gruppenmitglieder hat zu keiner signifikanten Verschlechterung der Leistung geführt. Eine mögliche Erklärung wäre, dass die

Kriminalaufgabe schlecht gewählt wurde und schlicht und einfach keine Medienwahl- und Gruppengrößeneffekte aufzeigt. Dies erscheint jedoch auf den ersten Blick unwahrscheinlich, wurde der Kriminalfall doch gerade für die Untersuchung der Auswirkung von Veränderungen in der Gruppengröße geschaffen [StSt92]. In anderen Untersuchungen zeigten die Gruppen mit 6 Teilnehmern wesentlich mehr Kommunikation, auch über die kritischen Informationen, als diese mit 3 Teilnehmern. Auch gibt es im Rahmen der Mediengeschwindigkeit einige Veränderungen in einigen wesentlichen Merkmalen der Kommunikation [LöSG06], welche zurückzuführen sind auf die Medienwahl. Dahingehend sind auch Experimentalfehler einer gleichgroßen, jedoch gegengesetzten Größenordnung auszuschließen, zumal die gleichzeitig stattfindenden Experimente mit der Designaufgabe entsprechende Produktivitätsunterschiede aufgezeigt haben. Auch von Kodierungsfehlern ist nicht auszugehen, da die Erkennung des richtigen Tatverdächtigen eine binäre Entscheidung ist.

Dementsprechend muss es Größen geben, die gegeneinander gerichtet arbeiten und sich damit aufheben. Im Folgenden wird eine mögliche Erklärung diskutiert.

5.2.1 Medienreichtum und Mediengeschwindigkeit

Die Media Richness Theorie sagt aus, dass für die Kriminalaufgabe Chatkommunikation besser als Audio geeignet ist, da diese den passenden Medienreichtum bietet. Durch die Reduktion auf den Textkanal sind die Teilnehmer gezwungen sich auf die Informationsübertragung zu konzentrieren. Audiokommunikation lenkt durch seine zusätzlichen Möglichkeiten dabei eher ab. Dieser Effekt wird aber aufgehoben von der höheren Geschwindigkeit, mit der Einzelne mit Audiokommunikation umgehen können. Mühsames Tippen fällt weg und Aussagen können schnell in die Gruppe geworfen werden, oftmals erst im Moment der Aussprache überlegt. In [LöSG06] konnte gezeigt werden, dass für Gruppen mit 4 Mitgliedern Audiokommunikation schneller als Chat ist. Deutlich mehr Informationen und kritische Hinweise konnten von den Audioteilnehmern gesendet werden als bei Chat. Jedoch waren Chatgruppen nachweislich effizienter in der Übertragung der kritischen Daten, wiesen sie doch ein besseres Verhältnis zwischen den wichtigen Informationen und der allgemeinen Kommunikation auf.

Offensichtlich gibt es somit zwei Kräfte die sich hier aufheben. Einerseits die höhere Geschwindigkeit von Audio (ersichtlich in der höheren Anzahl kritischer Hinweise pro Zeiteinheit, die übertragen werden). Auf der anderen Seite steht der besser passenden Medienreichtum von Chat, der nichts überkompliziert (ersichtlich im besseren Verhältnis kritische Hinweise pro Gesamtkommunikation). Dies alleine kann jedoch nur die fehlenden

Unterschiede in der Produktivität der Gruppen mit vier Teilnehmern erklären. Mit der Erweiterung der Gruppe sollte eigentlich zusätzliche Produktivität bei den Chatgruppen auftreten, wie dies bei der Designaufgabe der Fall war. Offensichtlich jedoch verhindert ein weiterer Effekt beim Kriminalfall diesen Produktivitätsgewinn. Dieser Effekt soll im Folgenden diskutiert werden.

5.2.2 *Gruppengröße und Informationsvolumen*

Die Designaufgabe hat klar gezeigt, dass mit dem richtigen Medium zusätzliche Gruppenmitglieder deutliche Produktivitätsgewinne erzeugen können. Gerade Chatkommunikation mit seiner parallelen Sendemöglichkeit bietet die Gelegenheit, gleichzeitig an einem Thema zu diskutieren. Damit wird die Blockade der anderen Gruppenmitglieder bei Audio aufgehoben. Beim Kriminalfall jedoch wird offensichtlich die Produktivität zusätzlicher Gruppenmitglieder durch die Aufgabenstellung gehindert. Aufgaben mit hoher Unsicherheit sind laut Media Richness Theorie dadurch charakterisiert, dass nur Informationen ausgetauscht werden müssen, um deterministisch zu einem Ergebnis (oder der Erkenntnis, dass es kein Ergebnis geben kann) zu führen. Damit ist in den meisten Fällen das zu kommunizierende Informationsvolumen begrenzt. Deshalb sind Personen, die über keine benötigten Informationen verfügen oder nur Duplikate besitzen, unnötig bei Aufgaben mit hoher Unsicherheit und niedriger Mehrdeutigkeit. Sie können sogar zu einer Verschlechterung der Produktivität führen, da sie keinen Zusatznutzen einbringen, jedoch durchaus durch Rückfragen und Verständnisprobleme Fragen aufwerfen können, die zu einem Produktivitätsverlust führen, da sie erst die ihnen ggf. fehlenden Informationen erhalten müssen. Damit ist die maximale Produktivität dann erreicht, wenn exakt nur die Personen in der Gruppe sind, deren Wissen unbedingt benötigt wird, jedes weiteres Gruppenmitglied sorgt für eine Senkung der Produktivität. Aufgaben mit hoher Mehrdeutigkeit jedoch haben ein nicht vorhersehbares Volumen an Informationen, da durch den Verständnisfindungsprozess die Ansichten aller Beteiligten in die Lösungsmenge einfließen und somit den Informationsraum erweitern. Ferner findet durch Gruppeneffekte (z.B. im Rahmen eines Brainstormings) auch gegenseitige Beeinflussung statt, die zu neuen Ideen und Ansichten führen kann, die so keiner der Teilnehmer vorher alleine hatte. Bei diesem Aufgabentyp wird durch die Steigerung der Gruppengröße die Produktivität der Gruppe ebenfalls gesteigert. Jedes neue Mitglied kann Input geben, der direkt der Gruppe hilft und sie zusätzlich auch zu neuen Ideen anspricht.

Dementsprechend wird die zusätzliche Leistung, die von einer Gruppenerweiterung erhofft wird, durch den Aufgabentyp beschränkt. Deshalb bringt eine Erweiterung der Gruppe bei Aufgaben mit hoher Unsicherheit über die benötigte Minimalbesetzung hinaus keinen Produktivitätsgewinn. Stasser und Steward [Stst92] haben sich in ihrer Untersuchung primär auf den Kommunikationsprozess konzentriert und dabei die Menge an kommunizierten Informationen untersucht. Dabei nahm die Kommunikation, auch über die kritischen Details, mit steigender Gruppengröße zu. Dies führt jedoch offensichtlich nicht automatisch zu höherer Produktivität. Die inhärente Natur der Aufgabe mit ihrer Einschränkung des Informationsvolumens führt dazu, dass die Produktivität mit zusätzlichen Mitgliedern ohne kritische Informationen abnimmt. Dies trifft auf die meisten Aufgaben mit hoher Unsicherheit zu, da sich diese durch den Informationsaustausch charakterisieren und dieser, zumindest im Rahmen von synchroner Kommunikation mit Audio oder Chat, zeitlich begrenzt ist. Nur Aufgaben mit hoher Mehrdeutigkeit profitieren uneingeschränkt von zusätzlichen Gruppenmitgliedern, da diese immer zusätzliche Ansichten und Informationen einbringen können. Damit ist klar zu sagen: Ausgehend von der zu bearbeitenden Aufgabe gilt es die Anzahl der Gruppenmitglieder festzulegen. Aufgaben mit hoher Mehrdeutigkeit profitieren von weiteren Mitgliedern. Aufgaben die von hoher Unsicherheit gekennzeichnet sind sollten jedoch ausschließlich von den unbedingt notwendigen Teilnehmern bearbeitet werden.

5.3 Gesamtbetrachtung

Die Zufriedenheit hat sich im Rahmen der Erwartungen verhalten. Das Experiment konnte zeigen, dass mit zunehmender Gruppengröße Gruppen mit Chatkommunikation zufriedener sind als mit Audiokommunikation. Noch größere Gruppen werden wahrscheinlich noch deutlichere Unterschiede in der Zufriedenheit zugunsten der Chatkommunikation aufzeigen. Jedoch hat sich die Produktivität entgegen der Erwartungen entwickelt. Die Nutzung des laut Media Richness Theorie besser passenden Mediums brachte keine Produktivitätssteigerung. Die Erweiterung der Gruppen um zusätzliche Mitglieder hatte gar einen leicht negativen Effekt. Wir konnten zwei Stellen identifizieren, an denen Faktoren der Medienwahl sich gegenseitig beeinflussen. Medienreichtum und Mediengeschwindigkeit können unterschiedliche Effekte haben, die sich gegenseitig aufheben. Dadurch kann es vorkommen, dass die Auswahl des Mediums keinen signifikanten Effekt hat. Zudem wirkt sich das Informationsvolumen der Aufgabe auf den Effekt der Gruppengröße aus. Bei mehrdeutigen Aufgaben profitieren Gruppen mit dem richtigen Medium von zusätzlichen Mitgliedern, die neue Ansichten

einbringen können. Bei unsicheren Aufgaben führt aber das limitierte Informationsvolumen dazu, dass zusätzliche Mitglieder ohne neue Informationen die Produktivität der Gruppe einschränken. Dieser Effekt ist auch jenseits des Experimentalsettings bei langwierigen Gruppensitzungen zu beobachten, wo redundante Informationen immer wieder wiederholt werden, ohne dass neue Ansichten eintreten können.

Zusammenfassend lässt sich sagen, dass die beiden besprochen Effekte maßgeblich für die Produktivität verantwortlich sind. Detailauswertungen der Kommunikationsprotokolle werden weitere Einblicke geben und auch erlauben, die Größenordnung der Effekte zu bestimmen.

6 Ausblick

Die Forschungsergebnisse rund um die Medienwahl zwischen Audio und Chat sind sehr widersprüchlich. Keine der Theorien konnte bisher alleine erklären, warum signifikante Ergebnisse auftraten oder auch nicht. In diesem Text haben wir Faktoren identifiziert, die sich gegenseitig beeinflussen. Zum Teil heben sie sich sogar gegenseitig auf. Mit diesen Faktoren ist es möglich, die Auswirkungen der Medienwahl und der Gruppengröße für Audio und Chatkommunikation zu beschreiben. Dabei steht noch aus, die Effekte in ihrer Wirkung und Größe genauer zu untersuchen. Dazu wird es notwendig sein die Kommunikation der Gruppen genauer zu analysieren. Dabei sollen Hinweise auf die Mediengeschwindigkeit in Gruppen mit sieben Mitgliedern gefunden werden. Ferner ist es notwendig, den Effekt der Aufgabe auf die benötigte Medienreichhaltigkeit und die Produktivität zusätzlicher Gruppenmitglieder zu analysieren. Weitere Erkenntnisse können gewonnen werden, wenn man Gruppen mit Zugriff auf beide Medien beobachtet. Dabei können Erfahrungen gewonnen werden, was erfolgreiche Gruppen im Umgang und Wechsel der Medien auszeichnet. Zusätzlich erkennt man Fehler, die von weniger geschickt agierenden Gruppen gemacht werden. Damit ist es dann möglich, mit höherer Sicherheit eine Medienwahlempfehlung zwischen Audio und Chat abzugeben.

Danksagung

Wir bedanken uns bei Mathis Müry und Sibylle Grimm für das Durchführen der Experimente. Ferner gilt unser Dank Christoph Göth, Marco Prestipino und Peter Vorburger, die uns bei der Auswertung der Designaufgaben halfen.

Literatur

- [Bos02] *Bos, Nathan; Olson, Judy; Gergle, Darren; Olson, Gary; Wright, Zach*: Effects of four computer-mediated communications channels on trust development. Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves. Minneapolis, Minnesota, USA, ACM Press: (2002), 135-140.
- [Broo96] *Brooke, J.*: SUS—A quick and dirty usability scale. Usability Evaluation in Industry. Jordan, P.W.; Thomas, B.; Weerdemeester, B.A.; McClelland, I.L. London, Taylor&Francis: (1996), 189-194.
- [DaLe86] *Daft, Richard L.; Lengel, Robert H.*: "Organizational information requirements, media richness and structural design." *Manage. Sci.* 32(5): (1986), 554-571.
- [DaLT87] *Daft, Richard L.; Lengel, Robert H.; Trevino, Linda Klebe*: "Message equivocality, media selection and manager performance: implications for information systems." *MIS Q.* 11(3): (1987), 355-366.
- [DeVa99] *Dennis, Alan R.; Valacich, J.S.*: Rethinking media richness: towards a theory of media synchronicity. System Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on, (1999).
- [DiSt87] *Diehl, M.; Stroebel, W.*: "Productivity Loss in Brainstorming Groups: Towards the Solution of a Riddle." *Journal of Personality and Social Psychology* 53: (1987), 497-509.
- [DiSt91] *Diehl, M.; Stroebel, W.*: "Productivity Loss in Idea-Generating Groups: Tracking Down the Blocking Effect." *Journal of Personality Assessment* 61: (1991), 392-403.
- [FuSS90] *Fulk, Janet; Schmitz, J.; Steinfeld, Charles W.*: A social influence model of technology use. *Organizations and Communication Technology*. Fulk, Janet; Steinfeld, Charles W. Newbury, Sage Publications: (1990), 117-140.

- [Gall92] Gallupe, R.B.; Dennis, Alan R.; Cooper, W.H.; Valacich, J.S.; Bastianutti, L.M.; Nunamaker, J.F.jr.: "Electronic Brainstorming and Group Size." *Academy of Management Journal* 35(2): (1992), 350-369.
- [Grae98] *Graetz, K. A.; Boyle, E. S.; Kimble, C. E.; Thompson, P.; Garloch, J. L.:* "Information sharing face-to-face teleconferencing, and electronic chat groups." *Small Group Research* 29(6): (1998), 714-743.
- [Grim06] *Grimm, S:* Empirische Untersuchung des Kommunikationsverhaltens virtuell verteilter Gruppen anhand von Audio- und Chatkonferenzen Department of Informatics, University of Zurich, Zurich, (2006).
- [GrPa02] *Grinter, Rebecca E.; Palen, Leysia:* Instant messaging in teen life. Proceedings of the 2002 ACM conference on Computer supported cooperative work. New Orleans, Louisiana, USA, ACM Press: (2002), 21-30.
- [KiWa92] *Kinney, Susan T.; Watson, Richard T.:* The effect of medium and task on dyadic communication. Proceedings of the thirteenth international conference on Information systems. Dallas, Texas, United States, University of Minnesota: (1992), 107-117.
- [LöSc07] *Löber, A.; Schwabe, G.:* Audio vs. Chat: The Effects of Group Size on Media Choice. accepted at HICCS 2007, Hawaii, (2007).
- [LöSG06] *Löber, A., Grimm, S. and Schwabe, G., Audio vs Chat:* Can media speed explain the differences in productivity ? in ECIS 2006, Göteborg, (2006).
- [Müry05] *Müry, M.:* Kommunikationsverhalten virtuell verteilter Gruppen. Department of Informatics, University of Zurich, Zurich, (2005)
- [Mull03] *Muller, Michael J.; Raven, Mary Elizabeth; Kogan, Sandra; Millen, David R.; Carey, Kenneth:* Introducing chat into business organizations: toward an instant messaging maturity model. Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work. Sanibel Island, Florida, USA, ACM Press: (2003), 50-57.
- [MuJS91] Mullen, B.; Johnson, C.; Salas, E.: "Productivity Loss in Brainstorming Groups." *Basic and Applied Social Psychology* 12: (1991), 3-23.

- [NaWB00] *Nardi, Bonnie A.; Whittaker, Steve; Bradner, Erin*: Interaction and outeraction: instant messaging in action. Proceedings of the 2000 ACM conference on Computer supported cooperative work. Philadelphia, Pennsylvania, United States, ACM Press: (2000), 79-88.
- [NDVV91] *Nunamaker, J.; Dennis, A.R.; Valacich, J. S.; Vogel, D.*: "Information Technology for Negotiating Groups: Generating Options for Mutual Gain." *Management Science* 37: (1991).
- [Nuna91] *Nunamaker, J. F.; Dennis, Alan R.; Valacich, Joseph S.; Vogel, Douglas; George, Joey F.*: "Electronic meeting systems." *Commun. ACM* 34(7): (1991), 40-61.
- [OOSC93] *Olson, Judith S.; Olson, Gary M.; Storrosten, Marianne; Carter, Mark*: "Groupwork close up: a comparison of the group design process with and without a simple group editor." *ACM Trans. Inf. Syst.* 11(4): (1993), 321-348.
- [PiRe85] *Picot, A.; Reichwald, R.*: Bürokommunikation. Leitsätze für den Anwender. München, CW-Publikationen, (1985).
- [StSt92] *Stasser, G. Stewart, D.*: "Discovery of Hidden Profiles by Decision-Making Groups: Solving a Problem versus Making a Judgment." *Journal of Personality and Social Psychology* 63(4): (1992), 426-434.
- [Suh99] *Suh, Kil Soo*: "Impact of communication medium on task performance and satisfaction: an examination of media-richness theory." *Information & Management* 35(5): (1999), 295-312.
- [Vala94] *Valacich, Joseph S.; Mennecke, Brian E.; Wachter, Renee M.; Wheeler, Bradley C.*: Extensions to Media Richness Theory: A Test of the Task-Media Fit Hypothesis. HICSS, Maui, (1994).
- [VPGN93] *Valacich, Joseph S.; Paranka, D.; George, J. F.; Nunamaker, J.F.jr.*: "Communication Concurrency and the New Media. A New Dimension of Media Richness." *Communication Research* 20(2): (1993), 249-276.
- [Will79] *Williams, E.*: „Experimental Comparison of Face-to-Face and Mediated Communiation: A Review“, *Psychological Bulletin*, 85, 5, (1979), 963-976

Diskontinuierliche Innovation fördern

Die Rolle von Idea Mirrors zur Unterstützung von Innovation und Kooperation im Unternehmen

Michael Koch

Universität der Bundeswehr München
Fakultät für Wirtschafts- und Organisationswissenschaften
Werner-Heisenberg-Weg 39, 85577 Neubiberg
michael.koch@unibw.de

Kathrin M. Möslein

HHL – Leipzig Graduate School of Management
Center for Leading Innovation & Cooperation (CLIC)
Jahnallee 59, 04109 Leipzig
kathrin.moeslein@hhl.de

Abstract

In diesem Beitrag diskutieren wir die Idee, Konzepte aus den Bereichen der Community-Unterstützung und ubiquitären Benutzungsschnittstellen auf die Förderung der Fähigkeit zu diskontinuierlicher Innovation in Unternehmen anzuwenden. Unser Fokus liegt dabei auf der frühen Phase des Innovationsprozesses – der Ideengenerierung und kooperativen Ideenvernetzung. Kernidee des Ansatzes ist es, das kreative und insbesondere disruptive Ideenpotential des Unternehmens sichtbar zu machen und so die Motivation zur Ideengenerierung und kooperativen Ideenvernetzung zu fördern. „Idea Mirrors“ – große Wandbildschirme – liefern dabei die Basis für *Gewahrsein* („awareness“) und *Wertschätzung* („appreciation“) als Treiber der Innovationsfähigkeit. Das Innovation Lab Germany, ein Gemeinschaftsprojekt von Forschern des Advanced Institute of Management Research (AIM) an der London Business School, des Center for Leading Innovation & Cooperation (CLIC) der HHL – Leipzig Graduate School of Management, der FH Ingolstadt, des Imperial College London sowie der TU Berlin und der

TU München sowie zwanzig Partnerunternehmen bildet die empirische-explorative Basis der Konzeptentwicklung.

1 Motivation

Die Unterstützung der Kommunikation in Communities gewinnt immer mehr Bedeutung in Unternehmen. Besonders hervorzuheben ist dabei der Bereich des Wissensmanagements, wo Community-Plattformen Mitgliedern in so genannten „Communities of Practice“ dabei helfen, effizient zu kommunizieren und Information auszutauschen oder Experten zu finden (siehe z.B. Wenger 1998). Dabei bleibt die Unterstützung aber meist auf relativ strukturierte oder Dokumenten-basierte Aufgabenbereiche beschränkt. Der Grad der IT-Unterstützung für die Bereiche Kreativität, Motivation und Kooperation als Basis von Innovation ist demgegenüber deutlich unterentwickelt und steht im Zentrum unserer Forschungsarbeiten. In diesem Beitrag diskutieren wir die Idee, Konzepte aus den Bereichen der Community-Unterstützung und ubiquitären Benutzungsschnittstellen auf die Förderung der Innovationsfähigkeit in Unternehmen anzuwenden. Besonderes Anliegen ist es dabei die Fähigkeit zur diskontinuierlichen Ideengenerierung und kooperativen Ideenvernetzung gezielt zu fördern. Der Fokus liegt dementsprechend auf der frühen Phase des Innovationsprozesses. Kernidee des Ansatzes ist es, von Kunden, Lieferanten, Partnern und Mitarbeitern eingebrachten „Out-of-the-box“-Ideen und Vorschlägen durch öffentliche Präsentation *Gewahrsein* („awareness“) und *Wertschätzung* („appreciation“) zukommen zu lassen - das kreative Potential des Unternehmens also sichtbar zu machen.

Abschnitt 2 des Beitrags liefert eine kurze Einführung in Community-Unterstützung und Community-Awareness. Abschnitt 3 stellt dann den Anwendungsbereich der Ideengenerierung und kooperativen Ideenvernetzung im organisatorischen Innovationsmanagement und die Rolle der Motivation im Rahmen dieser zentralen Phase des Innovationsmanagements dar. *Gewahrsein* und *Wertschätzung* werden dabei als wichtigste Faktoren zur besseren Unterstützung von Innovation identifiziert. Abschnitt 4 geht anschließend darauf ein, wie die Idee der Visualisierung von Information aus Community-Unterstützungssystemen im organisatorischen Kontext eingesetzt werden kann, um die Motivation zur Kommunikation und kooperativen Vernetzung von „Out-of-the-box“-Ideen zu stärken. In Abschnitt 5 stellen wir dann vor, wie eine auf öffentlichen Wandbildschirmen basierende Lösung für die Förderung der Ideengenerierung im Bereich

diskontinuierlicher Innovation konkret aussehen kann (der „Idea Mirror“). Abschnitt 6 liefert noch erste Schlussfolgerungen und gibt einen Ausblick auf den nächsten Schritt im Forschungsprozess.

2 Community-Unterstützung und Community Awareness

2.1 Communities und Community-Unterstützung

Eine Community ist eine Gruppierung von Personen, die ein gemeinsames Interesse haben, sich mit einer gemeinsamen Idee identifizieren oder ähnliche Aufgaben zu erledigen haben, d.h. Gemeinsamkeiten aufweisen, die sie von anderen abgrenzt und die Basis von Interaktion und Kollaboration sein kann (siehe z.B. Hillery 1955, Mynatt et al. 1997). Neuere Charakterisierungen heben besonders die Notwendigkeit von gegenseitiger Hilfe in der Community hervor, z.B. die Bereitschaft, Wissen auszutauschen (Ishida 1998, Wenger 1998). Dieser Diskussion entsprechend, wird eine Community nicht nur als Menge von Personen mit Gemeinsamkeiten, betrachtet, sondern als Menge von Personen, die bereit sind einander zu unterstützen, die sich zum Vorteil aller austauschen und zusammenarbeiten. Neben der konkreten Zusammenarbeit zwischen den Mitgliedern sind die Hauptaktivitäten in Communities die Kommunikation und das Finden von Personen zum Kommunizieren. Community-Unterstützung kann deshalb auch als „Kommunikations- und Matchmaking-Unterstützung“ betrachtet werden. In der Funktionalität von Community-Unterstützungs-Werkzeugen finden sich folglich hauptsächlich folgende beiden Grundkonzepte (siehe hierzu auch Koch und Prinz 2005, Koch 2003 S31f):

- Bereitstellung eines Mediums für direkte Kommunikation und den indirekten Austausch von Information und Kommentaren zwischen den Community-Mitgliedern.
- Bereitstellung von Gewährsein über andere Community-Mitgliedern und Unterstützung bei dem Finden von Kommunikationspartnern (Medium zum Matchmaking).

Die Aufgabe, Kunden, Lieferanten, Partner und Mitarbeiter eines Unternehmens beim Austausch von Innovationsideen, -vorschlägen und -anregungen zu unterstützen, wird von uns daher konsequent als Community-Unterstützungs-Aufgabe betrachtet. Ideengeber und Ideenneh-

mer bzw. Ideenunterstützer tauschen in einer durch das Unternehmen definierten Community kooperativ Information aus.

2.2 Common Ground and Community Awareness

Ein wichtiger Aspekt in Community-Unterstützung ist die Unterstützung der informellen Kommunikation zwischen den Mitgliedern als Basis zum Aufbau eines gemeinsamen Kontexts („common ground“) der notwendig ist für erfolgreiche Kommunikation und Beziehungen. Clark definiert Common Ground“ in seinem Buch „Using Language“ (Clark 1996) als *„information that two parties share and are aware that they share“*. Clark zufolge ist

“Everything we do is rooted in information we have about our surroundings, activities, perceptions, emotions, plans, interests. Everything we do jointly with others is also rooted in this information, but only in that part we think they share with us.” (Clark 1996).

Mit dem Konzept des gemeinsamen Kontexts („common ground“) ist eng verwandt das Konzept des Gewährseins („awareness“), das bereits intensiv im Bereich der Team-Unterstützung erforscht worden ist. Dourish und Belotti definieren Awareness als *“an understanding of the activities of others, which provides a context for your own activities”* (Dourish and Belotti 1992). Zu Awareness kann dabei sehr unterschiedliche Information beitragen, von Information über die Erreichbarkeit von Kollegen bis zu Hinweisen auf Personen oder Informationen, die für die eigene Arbeit oder Freizeitaktivitäten hilfreich sein können. Schlichter et al. betrachten die Bereitstellung von Awareness als die größte Gemeinsamkeit in verschiedenen Arten der Zusammenarbeitsunterstützung (Schlichter et al. 1998). Während sich Groupware häufig auf so genannte Workspace-Awareness konzentriert, liegt der Fokus bei Community-Unterstützung auf Awareness über Personen und ihre Beziehungen untereinander bzw. zu Inhalten. Grund dafür ist das Fehlen eines gemeinsamen Arbeitsbereichs in der Community und der Umstand, dass sich Community-Mitglieder nicht zwangsweise untereinander kennen.

Die theoretischen und empirischen Erkenntnisse zu gemeinsamem Kontext und Awareness (siehe z.B. Dourish und Belotti 1992, Gutwin et al. 1996) regen an, dass die Bereitstellung einer detaillierten und zusammengefassten Sicht auf eine Community (die Mitglieder, die verfügbare Expertise und die von den Mitgliedern erzeugte Information) den Community-Mitgliedern bei ihren Aktivitäten helfen kann. Der Wert des Gewährseins liegt dabei vor allem in der Senkung

von Koordinationskosten (durch Ermöglichung impliziter Koordination), im Finden von Kommunikationspartnern und Information sowie insbesondere in der Unterstützung von intrinsischer Motivation durch die Ermöglichung des Aufbaus und der Wahrnehmung einer Wertschätzung. Denn die Motivation verschiedener Personen, etwas zu einer Community beizutragen, hängt häufig gerade von der Möglichkeit ab, dass sie oder ihr Beitrag wahrgenommen werden und dies, auch transparent wird. Wie wir später noch weiter ausführen, kann diese Idee der Unterstützung von Gewahrsein und Wertschätzung durch Idea Mirrors das Innovationsmanagement in Unternehmen, insbesondere in der frühen Phase der Generierung und Vernetzung von Ideen mit Potential zur Diskontinuität, maßgeblich unterstützen.

3 Innovationsfähigkeit von Unternehmen

Die Innovationsfähigkeit eines Unternehmens hängt ab von der Kreativität, Motivation und Bereitschaft ihrer Kunden, Lieferanten, Mitarbeiter und Partner, Ideen zur Verbesserung von Produkten, Prozessen und Strukturen zu generieren, beizutragen und zu vernetzen (Tidd, Bessant, Pavitt 2005). Das Innovationsmanagement eines Unternehmens zielt auf die Verbesserung des Prozesses der Umsetzung von Ideen in vermarktbar Produkte und Dienste oder Prozessverbesserungen (siehe Goffin und Pfeiffer 1999; Oke und Goffin 2001).

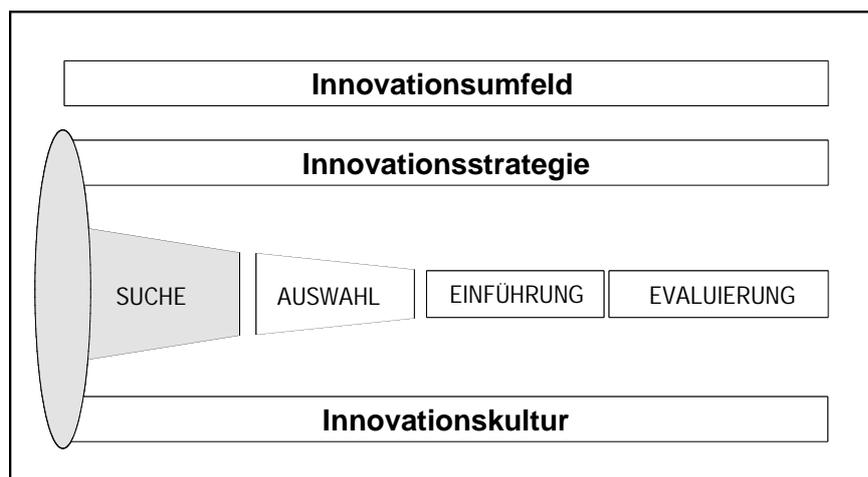


Abbildung 1: Der Innovationsprozess im Unternehmen

In vielen Organisationen steht ein breites Spektrum von Instrumenten, Mechanismen und Systemen zur Verfügung um diesen fundamentalen Prozess zu unterstützen. Klassisch ist der Blick

dabei stark nach innen gerichtet: Hauptlieferant im Bereich der Ideenproduktion sind die Mitarbeiter des Unternehmens, insbesondere im Bereich der Forschung und Entwicklung. Darüber hinaus liegt der Fokus auf inkrementellen Innovationsschritten - der so genannten kontinuierlichen Innovation. Um dem Innovationsprozess jedoch auch verstärkt „Out-of-the-box“-Ideen und Vorschläge von außen zuzuführen und insbesondere die Fähigkeit zu größeren Innovations-sprüngen – der so genannten diskontinuierlicher Innovation - zu fördern, ist es ein wichtiges Ziel, die Ideen eines möglichst breiten und diversen Kreises von Mitarbeitern, Kunden, Lieferanten oder Partnern in die Innovationspipeline des Unternehmens einzuspeisen (Oke und Goffin 2001; Munshi et al. 2005). Man spricht in diesem Zusammenhang von der Öffnung der frühen Phasen der Innovationspipeline im Unternehmen (siehe Abbildung 1).

3.1 Explorative Erkenntnisse zur Innovationsfähigkeit

Um der Fähigkeit zur Generierung und dem Management diskontinuierlicher Innovation in Unternehmen besser auf die Spur zu kommen, haben sich zu Beginn des Jahres 2006 Innovationsforscher aus Universitäten und Innovationsmanager aus Industrie- und Dienstleistungsunternehmen in Dänemark, Deutschland und UK zusammengefunden und jeweils ein nationales Innovation Lab ins Leben gerufen. Im Innovation Lab Germany sind Vertreter der folgenden Unternehmen beteiligt: Allianz, Arvato mobile, BMW Group, Bühler Motor, Deutsche Bank, E.ON Energie, EOS, Flughafen München, Henkel, ITQ, Microsoft, O2, SAP, Schreiner Group, Segma, Siemens, SZ, Vaillant Group, Voith, Webasto. Das Spektrum der Unternehmen im Untersuchungspanel ist bewusst über unterschiedliche Unternehmensgrößen, Branchen und Innovationskontexte gestreut. In regelmäßigen Workshops und Arbeitstreffen sowie durch die Erhebung unternehmensspezifischer Fallstudien werden im Rahmen der Innovation Labs die aktuellen Gegebenheiten des Innovationsmanagements erhoben, geplante und in Umsetzung befindliche Ansätze des Innovationsmanagements exploriert und neue Instrumente und Maßnahmen des Innovationsmanagements interaktiv entwickelt (vgl. Möslein und Neyer 2006).

Die aktuelle Bestandsaufnahme nach einem Jahr der Exploration macht deutlich, dass die Qualität des Innovationsergebnisse zwar stark von der Kreativität und der Qualität der in die Innovationspipeline eingespeisten Ideen abhängt, dass sich andererseits aber viele Unterstützungskonzepte noch immer primär auf die späten Phasen im Prozess, die Auswahl von Ideen, ihre Implementierung sowie deren Evaluierung fokussieren. Die frühe Phase der Suche, Kreativität

und Ideengenerierung wird heute meist nur wenig unterstützt, insbesondere durch Systeme des innerbetrieblichen Vorschlagswesens. Die Hauptherausforderung ist dabei, dass der Ideengeber meist vom Prozess der Ideenauswahl, -implementierung und -evaluierung isoliert wird. Oft erhalten Mitarbeiter, die Vorschläge liefern, kein Feedback zur Nutzung ihrer Ideen und verlieren so über die Zeit die Motivation dem System Input zur Verfügung zu stellen. Software-Hersteller wie Imaginatik (www.imaginatik.com) haben Prozesse und Werkzeuge zum Ideenmanagement entwickelt. Das Produkt „Idea Central“ von Imaginatik unterstützt zum Beispiel das Sammeln von Ideen von Mitarbeiter und stellt Funktionalität für den gesamten Innovationsprozess zur Verfügung. Nicht adressiert wird aber die Frage, wie man die notwendige Awareness über generierte Ideen und deren Nutzung schafft.

Viele formale Lösungen zu diesem Problem sind bereits diskutiert und implementiert worden. Institutionalisierte Ideenjäger, Gatekeeper oder Beobachter, hypothesengetriebene Suchstrategien und bewusste Diversitätsstrategien sind zentrale, aber bislang wenig verbreitete Ansätze bei der Suche nach diskontinuierlichen Ideen (Bessant und Francis, 2004; Philipps et al. 2004; Tidd, Bessant und Pavitt 2005; Bessant 2006). Robinson und Stern führen dazu beispielsweise aus:

“Much creativity is the result of informal poking around, experimenting and exploiting the unexpected. In the race for the reward, not only is creativity sacrificed, but opportunities for what cognitive psychologists call ‘incidental learning’, the important knowledge and insight gained from such exploration, are greatly reduced. We do not mean to suggest that those involved in creative acts should receive no rewards whatsoever. It is of course important for people to be recognized and treated fairly” (Robinson und Stern 1997, p. 55).

3.2 Motivatorische Grundlagen der Innovationsfähigkeit

Die Motivationsforschung ist sich der zugrunde liegenden psychologischen und sozialen Prozesse wohl bewusst: Sie unterscheidet (1) intrinsische, (2) extrinsische und (3) soziale Formen menschlicher Motivation, basierend darauf ob die Motivation zurückgeht auf (1) die Aufgabe selbst, (2) externe monetäre und nicht-monetäre Belohnung, oder (3) den sozialen Kontext der persönlichen Netzwerke und sozialen Beziehungen (Deci 1971, 1975; Deci und Ryan 1985; Frey und Oberholzer-Gee 1997; Amabile 1993). Ergebnisse der Motivationsforschung erklären wie extrinsische Motivation und speziell monetäre Belohnungen die intrinsische Motivation

durch das Hervorrufen eines rein Belohnungs-orientierten Verhaltens zerstören kann (der so genannte “crowding-out effect”: siehe Deci, 1971; Amabile 1987; Robinson und Stern 1997; Frey und Jegen 2000) und argumentiert, dass ein Mangel an sozialem Kontext und persönlicher Wertschätzung ein Hauptgrund für geringe Motivation und persönliches Engagement ist (Osterloh und Frost 2000; Gottschalg 2004). Häufig liegt der Fokus auf der extrinsischen Motivation. Die Gefahr der Zerstörung intrinsischer Motivation liegt auf der Hand. Und auch wenn es gelingt, den “crowding-out effect” zu adressieren, wird dennoch meist der „soziale Leim“, der persönliche Motivation in Unternehmenskontexten unterstützt, ignoriert.

In diesem Beitrag konzentrieren wir uns darauf, wie intrinsische und soziale Motivation in Innovationsprozessen durch öffentliche Präsentation und Feedback unterstützt werden können. Wir untersuchen wie Community Mirrors das Gewahrsein und die Wertschätzung von Beiträgen steigern können.

4 Community Mirrors

Im vorhergehenden Abschnitt haben wir die Bedeutung von Gewahrsein und Wertschätzung bei der Suche nach Ideen für diskontinuierliche Innovation hervorgehoben. Dies führt uns Kern unseres Projektvorhabens: der Unterstützung von Gewahrsein und Wertschätzung zu bereitgestellten Ideen und den Bereitstellern, um die Motivation zur Ideengenerierung zu erhöhen und Ideen kommunikativ zu vernetzen. Im Zentrum unseres Lösungsansatzes steht die Visualisierung des kreativen Potentials in Form von individuellen Ideen, insbesondere auch an Orten der Entscheidungsfindung. Wir schlagen vor, Ideen mittels ubiquitärer Benutzungsschnittstellen in den Aufmerksamkeitsbereich von Entscheidungsträgern zu bringen. Unter ubiquitären Benutzungsschnittstellen verstehen wir dabei Möglichkeiten zur Interaktion mit IT-Systemen, die in die (Arbeits-)Umgebung von Menschen integriert sind. Durch die einfache Verfügbarkeit in allen (Arbeits-)Kontexten entsteht so eine „Allgegenwärtigkeit“ der Benutzungsschnittstellen und damit der damit zugreifbaren IT-Systeme und der in ihnen bereitgestellten Information. Konkret könnte dies so aussehen, dass Kunstwerke, die üblicherweise die Wände in Konzernzentralen schmücken, durch dynamische Awareness-Displays ersetzt oder ergänzt werden. Die Visualisierung von Ideen in einer ansprechenden Form mittels halb-öffentlichen großen Wandbildschirmen ermöglicht zugleich Gewahrsein und Wertschätzung für das kreative Potential des Unternehmens.

4.1 Große Wandbildschirme zur Realisierung von Community Mirrors

Die Verfügbarkeit des Zugriffs auf Community-Unterstützungs-Anwendungen ist eines der wichtigsten Erfolgskriterien für solche Systeme. Ubiquitous Computing und Mobile Computing, d.h. neue Benutzungsschnittstellen, die in die reale Welt integriert sind, können die bisherigen Grenzen von Community-Unterstützung aufbrechen und neue Möglichkeiten für die Vergrößerung der Reichweite von Community-Unterstützungs-Anwendungen eröffnen. Dieser Ansatz kann besonders im Bereich Awareness-Unterstützung von Nutzen sein, d.h. bei der Visualisierung von Aktivitäten aus der Community, von Beziehungen und Interaktionen zwischen den Community-Mitgliedern, an Orten und in Situationen, an denen die Community-Mitglieder für solche Information empfänglich sind. Grund dafür ist, dass Awareness-Information meist nicht absichtlich nachgefragt wird, aber eine besondere Wirkung bei peripherer Wahrnehmung und Konsumierung entfaltet. Solche Awareness-Anwendungen für Communities („*Community Mirrors*“) stellen Information über die Community und die Aktivitäten ihrer Mitglieder zur Unterstützung von Interaktion und Matchmaking in der Community zur Verfügung. Community Mirrors können mit großen (interaktiven) Wandbildschirmen, die an halb-öffentlichen Orten aufgestellt werden, an denen die potentiellen Konsumenten der Information arbeiten oder sich sonst aufhalten, realisiert werden.

4.2 Community Mirror Prototypen und verwandte Arbeiten

Öffentliche gemeinsame Wandbildschirme sind kein völlig neues Konzept für Benutzungsschnittstellen. Erste Arbeiten in diesem Bereich können bei Myron Krueger in den 1970ern gefunden werden (Krueger 1991). Aktuelle Arbeiten konzentrieren sich häufig auf die Unterstützung von Teams (z.B. DynaWall, Geissler 1998). Daneben gibt es aber auch Ansätze, die Wandbildschirme zur Unterstützung des Informationsflusses in Communities vorschlagen, z.B. die Plasma Poster am Fuji Xerox Palo Alto Laboratory (Churchill et al. 2003, 2004) oder die CWall am Xerox Research Lab Europe (Snowdon und Grasso 2002) sowie weitere Anwendungen interaktiver Wandbildschirme im Medien- und Werbeumfeld (Scanlon 2003). Das Hauptproblem bei fast allen existierenden Wandbildschirm-Anwendungen für Community-Unterstützung ist dabei, dass sie meist geschlossene Systeme darstellen, also nicht mit anderen (Community-Unterstützungs-)Anwendungen interagieren. Dadurch ist es meist sehr schwierig, geeignete Daten zur Anzeige zu bekommen. In unserer Gruppe haben wir deshalb Community Mirrors entwickelt, die sich speziell um die Integration mit anderen Anwendungen kümmern.

Erste Prototypen sind der *Library Mirror*, der *Meeting Mirror* und der *Announcement Mirror* (Koch 2004, 2005). Die Erfahrung bei der Entwicklung und dem Einsatz dieser Prototypen hat uns motiviert, das Konzept des „Idea Mirror“ als Anwendung zur Unterstützung der Ideengenerierung zu entwickeln.

5 Der „Idea Mirror“

Ergebnis der Diskussion in den vorhergehenden Abschnitten ist das Konzept eines „Idea Mirrors“, eines Community Mirrors mit großen Wandbildschirmen, die den Innovationsprozess in Unternehmen durch Steigerung der Gewährsein und Wertschätzung für diverse Ideen von Mitarbeitern, Kunden, Lieferanten und Partnern im Bereich der Suche nach diskontinuierlicher Innovation unterstützen. Die Motivation für Ideengeber wird dabei hauptsächlich durch die breite Visualisierung und Kommunikation als notwendige Voraussetzung der Wahrnehmung und Nutzung sichergestellt. In diesem Abschnitt beschreiben wir das Grobdesign dieser Anwendung. Das Design stützt sich dabei auf die Charakteristika der Anwendungsdomäne und den Erfahrungen, die wir mit anderen Community Mirror Prototypen gemacht haben (siehe auch Koch und Möslein 2006).

Als erstes ist zu bemerken, dass der Idea Mirror auf die frühe Phase der Innovationssuche im Innovationsprozess und hierbei insbesondere auf die Ideengenerierung im Bereich der diskontinuierlichen Innovation fokussiert. Im Sinne der Integration nehmen wir an, dass es Desktop- oder Web-basierte Anwendungen zur Eingabe von Ideen und zur Begutachtung und Zugriff auf detaillierte Informationen gibt (siehe auch Abbildung 3). Dies können kommerzielle Anwendungen mit integrierter Workflow-Unterstützung wie Idea Central von Imaginatik oder einfache Web-basierte Datenbanklösungen sein. Der Idea Mirror nutzt die Information aus diesen Systemen, um

- Gewährsein („*awareness*“) zu den eingegebenen Ideen (und allgemein dem kreativem Potential des Unternehmens) zur Verfügung zu stellen und einfachen Zugriff auf Kurzbeschreibungen der Ideen zu bieten und

- Wertschätzung („*appreciation*“) für die Ideengeber sicherzustellen (durch die öffentliche bzw. halb-öffentliche Anzeige) und ggf. einfache Möglichkeiten zu Kontaktaufnahme oder Weiterverfolgung von Ideen zu bieten.

Wie in Abschnitt 2.1 schon angesprochen gehen wir dabei davon aus, dass Ideengeber und Ideennehmer bzw. Ideenunterstützer in einer durch das Unternehmen und die gemeinsamen Interessen definierten Community kooperativ zusammenarbeiten.

Von den Prototypen anderer Community Mirror Anwendungen haben wir gelernt, dass für das Funktionieren der Anwendung wichtig ist, dass die Anzeige Aufmerksamkeit auf sich lenkt und ansprechend gestaltet ist. Der Prototyp des Idea Mirrors löst dies mit folgenden Maßnahmen, die teilweise in Interaktion mit Künstlern und Designern entwickelt worden sind:

- Ideen werden nur mit einem Portraitbild des Ideengebers und einem griffigem Kurztitel (Slogan) als Karteikarten dargestellt.
- Der Titel (Slogan) ist in einer großen Schriftgröße ausgeführt – was nur Platz für vier bis fünf Ideen zur selben Zeit lässt, dafür aber periphere Wahrnehmung erlaubt.
- Die Darstellung ist animiert – Ideen erscheinen am Bildschirmrand, verweilen einige Zeit auf dem Schirm und verschwinden dann wieder.
- Die Hintergrundfarbe der Darstellung ändert sich laufend (nahtloser Übergang zwischen verschiedenen Farben).

Abbildung 2 zeigt die links den Prototypen des in Abschnitt 4.2 angesprochenen Meeting Mirrors im Einsatz und rechts einen Entwurf dafür, wie die Darstellung des Idea Mirrors aussehen könnte.



Abbildung 2: Meeting Mirror im Einsatz und mögliche Darstellung im Idea Mirror

Die Interaktion mit dem Idea Mirror sollte dabei so einfach wie möglich sein: Durch Berührung der Karteikarte zu einer Idee wird die Karte erweitert und mehr Information zur Idee angezeigt – z.B. ein Abstract, Meta-Information zur Klassifizierung der Idee und weitere Kontaktinformation. Zusätzlich können Benutzer ein Menü aufrufen, um Ideen nach den zugeordneten Meta-Informationen abzufragen. Dieses Menü wird angezeigt, wenn ein spezieller Bereich des Bildschirms mit einem großen Fragezeichen berührt wird. Das Menü zeigt die verschiedenen nutzbaren Meta-Informationen (z.B. Kategorien) und alle möglichen Werte, so dass die Auswahl durch einfaches Anklicken (Berühren) durchgeführt werden kann.

Um das Ziel zu erreichen, die Entscheidungsträger auf das kreative Potential des Unternehmens aufmerksam zu machen und um die notwendige Anerkennung für die Ideengeber zu realisieren, schlagen wir vor, die Idea Mirrors an halb-öffentlichen Orten zu installieren, wo sich Entscheidungsträger regelmäßig aufhalten (z.B. Gänge, Eingangsbereiche, Pausenräume). Aufbauend auf unseren Erfahrungen mit anderen Community Mirror Anwendungen nehmen wir an, dass die Anwendung hauptsächlich so genutzt wird, dass peripher einige Schlüsselworte wahrgenommen werden und dann bewusst weitere Information nachgefragt wird (entweder am Idea Mirror oder über andere Wege). Durch Anbringung der Bildschirme in halb-öffentlichen Berei-

chen wo man häufig auf Gruppen von Personen trifft, ermöglichen wir weiterhin die sofortige Diskussion von Ideen in den Gruppen (wir haben solches Verhalten bei unseren anderen Community Mirror Prototypen beobachtet). Die Installation sollte als interaktive Kunst betrachtet werden – weshalb wir auch Designer und Künstler in den Entstehungsprozess mit einbezogen haben. Ein positiver Seiteneffekt davon ist, dass für die notwendigen Investitionen alternative Budgets in Unternehmen offen stehen.

Wie schon angesprochen stellt der Idea Mirror technisch nur eine zusätzliche Möglichkeit zur Anzeige der Informationen aus einem (Web-basierten) Ideen-Management System dar (siehe Abbildung 3). Das bedeutet, dass die schon etablierten Möglichkeiten zur Eingabe von Ideen und zum Koordinieren des Innovationsprozesses genutzt werden können. Die eventuell schon vorhandenen Systeme müssen nur um eine Web Service Schnittstelle für die Idea Mirror Anwendung ergänzt werden.

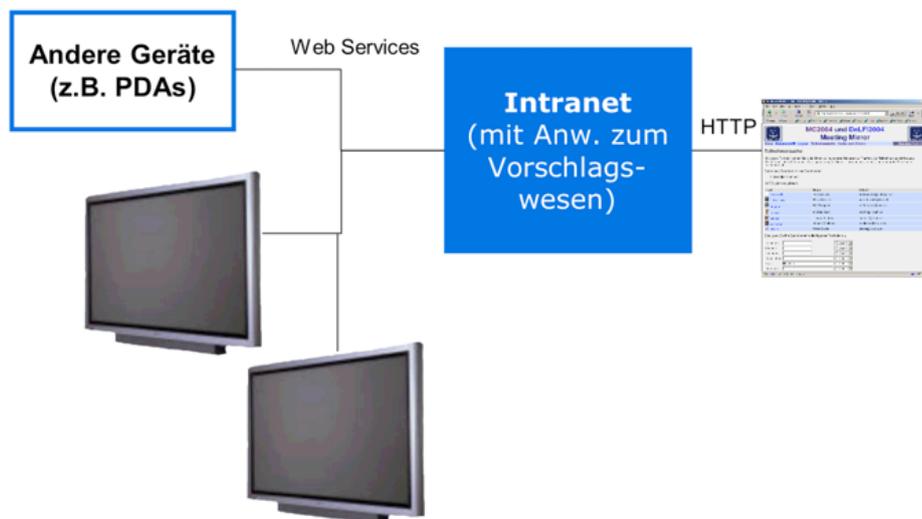


Abbildung 3: Idea Mirror als zusätzliches Fenster zum Intranet

6 Zusammenfassung und Ausblick

In diesem Beitrag haben wir die Idee untersucht, Community-Unterstützungs-Konzepte und große Wandbildschirme zur Unterstützung von Motivation der Ideengeber einzusetzen. Unser spezieller Fokus war dabei auf die frühe Phase der Ideensuche im Innovationsprozess gerichtet und hat sich dabei insbesondere auf die Unterstützung der Generierung von diversen Ideen für

diskontinuierliche Innovation fokussiert. Die spezifischen Erfahrungen, Anforderungen und Herausforderungen dieser Phase der Ideensuche konnten wir mit den zwanzig Partnerunternehmen des Innovation Lab Germany als empirisch-explorativer Basis der Konzeptentwicklung diskutieren und mit den beteiligten nationalen und internationalen Wissenschaftlern im Kontext jüngster Erkenntnisse der Innovationsforschung spiegeln. Dabei zeigte sich, welche zentrale Rolle Gewahrsein („awareness“) und Wertschätzung („appreciation“) für die Förderung und Motivation des kreativen Potentials eines Unternehmens spielen. Hierauf zielt die Anzeige der über eine (Web-basierte) Innovationsplattform zur Verfügung stehenden Ideen auf großen Wandbildschirmen in halb-öffentlichen Bereichen.

Die Herausforderungen bei der Einführung neuer Benutzungsschnittstellen in betrieblichen Informationssystemen liegen sowohl in der Integration der Technik (Verknüpfung der neuen Benutzungsschnittstellen mit den Web-basierten Informationssystemen) als auch in der soziopolitischen Akzeptanz und dem resultierenden Verhalten. Letzteres steht im Zentrum weiterführender Forschungsarbeiten zur Umsetzung und Implementierung der Idea Mirror-Konzeption im Unternehmen. Unsere bisherigen Erfahrungen mit Community Mirrors in anderen Anwendungskontexten haben uns bereits gezeigt, dass die Bereitstellung eines „Fensters“ im natürlichen Umfeld potentieller Nutzer (als „Kunst“ an den Wänden halb-öffentlicher Räume) die Reichweite eines solchen Community-Unterstützungs-Systems entscheidend erweitern kann. Dies motiviert die weiterführende Forschung im beschriebenen Anwendungsfeld.

Auch im Kontext soziopolitischer Akzeptanz zu sehen ist die berechtigte Frage nach der Einhaltung von Vertraulichkeit bei der Kommunikation von Ideen. Wir sind in unserer Motivation immer davon ausgegangen, dass es vorteilhaft ist, Ideen innerhalb des Unternehmens möglichst frei zugänglich zu machen. In einigen Anwendungsszenarien lässt sich diese Forderung aber voraussichtlich nicht durchsetzen. Hier ergeben sich also Einschränkungen für den Einsatzbereich von Idea Mirrors.

Literatur

- Amabile, T.M. (1987): The Motivation to be Creative. In: Isaksen, S.G. (Ed.): *Frontiers of Creativity Research*, Bearly Press: Buffalo NY 1987, S. 229-230.
- Amabile, T.M. (1993): Motivational Synergy: Toward New Conceptualizations of Intrinsic and Extrinsic Motivation in the Workplace. *Human Resource Management Review*, 3(3): S. 185ff.
- Bessant, J. und Francis, D. (2004): *Developing Parallel Routines for Radical Product Innovation*, AIM Working Paper No. 10, Advanced Institute of Management Research, London.
- Bessant, J. (2006): *Dealing with Discontinuous Innovation*, Unpublished Working Paper, HHL – Leipzig Graduate School of Management, Center for Leading Innovation & Cooperation (CLIC), Leipzig.
- Churchill, E.; Nelson, L. und Denoue, L. (2003): *Multimedia Flyers – Informal Information Sharing with Digital Community Bulletin Boards*. In: *Proc. Communities and Technologies*, Amsterdam, Kluwer Publishers.
- Churchill, E.; Girgensohn, A.; Nelson, L. und Lee, A. (2004): *Blending Digital and Physical Spaces for Ubiquitous Community Participation*. *Communication of the ACM*, Feb. 2004, 47 (2), S. 39-44.
- Clark, H. H. (1996): *Using Language*, Cambridge University Press.
- Deci, E.L. (1971): Effects of Externally Mediated Rewards on Intrinsic Motivation, in: *Journal of Personality and Social Psychology*, Vol. 18, 1971, S. 114ff.
- Deci, E.L. (1975): *Intrinsic motivation*. New York and London: Plenum Press.
- Deci, E.L. und Ryan, R.M. (1985): *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.

- Dourish, P. und Belotti, V. (1992): Awareness and Coordination in Shared Workspaces. In: Proc. of the Conf. on Computer-Supported Cooperative Work, S. 107-114.
- Frey, B.S. und Jegen, R. (2000): Motivation Crowding Theory: A Survey of Empirical Evidence. CESifo Working Paper Series CESifo Working Paper No. 26.
- Frey, B.S. und Oberholzer-Gee, F. (1997): The cost of price incentives: An empirical analysis of Motivation Crowding-Out. *American Economic Review*, 87(4): S. 746ff.
- Geissler, J. (1998): Shiffle, throw or take it! Working efficiently with an interactive wall. In: Proc. CHI'98, Los Angeles, LA.
- Goffin, K. und Pfeiffer, R. (1999). *Innovation Management in UK and German Manufacturing Companies*, London: Anglo-German Foundation.
- Gottschalg, O. (2004): Towards a Motivation-Based Theory of the Firm: Integrating Governance and Competence-Based Approaches. INSEAD Working Paper.
- Gutwin, C., Greenberg, S. und Roseman, M. (1996): Workspace Awareness in Real-Time Distributed Groupware: Framework, Widgets, and Evaluation. In: Proc. Conf. on Human-Computer Interaction: People and Computers (HCI), Springer, S. 281-298.
- Hillery, G. A. (1955): Definitions of Community: Areas of Agreement, *Rural Sociology*, 20, S. 111 – 123.
- Ishida, T. (1998): *Community Computing*. John Wiley and Sons.
- Koch, M. und Möslein, K. (2006): Community Mirrors for Supporting Corporate Innovation and Motivation. Proc. Europ. Conf. on Information Systems (ECIS), Göteborg, Sweden, Jun. 2006.
- Koch, M. und Prinz, W. (2005): Communities und Community-Unterstützung. *i-com Zeitschrift für interaktive und kooperative Medien*, 4(2), 2005, S. 4-7.
- Koch, M. (2005): Supporting Community Awareness with Public Shared Displays. Proc. Bled Intl. Conf. on Electronic Commerce, Bled, Slowenien, Jun. 2005.

- Koch, M. (2004): Building Community Mirrors with Public Shared Displays. Proc. eChallenges e-2004 Conference, Vienna, Austria, Okt. 2004.
- Koch, M. (2003): Community-Unterstützungssysteme – Architektur und Interoperabilität, Habilitationsschrift, Fakultät für Informatik, Technische Universität München, Feb. 2003
- Krueger, M. W. (1991): Artificial Reality III, Addison-Wesley.
- Möslein, K. und Neyer, A.-K. (2006): Leading Innovation: The role of leadership systems for the leadership of boundary spanning innovation, 26th Annual International Conference "Strategy and Governance in a World of Institutional Change", Strategic Management Society (SMS), Wien, 29. Oktober - 1. November 2006.
- Munshi, N.; Oke, A.; Puranam, P.; Stafylarakis, M.; Towells, S.; Möslein, K.; Neely, A. (2005): Leadership for Innovation. Summary Report from an AIM Management Research Forum in cooperation with the Chartered Management Institute, AIM Research, London, Feb. 2005.
- Mynatt, E. D.; Adler, A.; Ito, M.; Oday, V. L. (1997): Design for Network Communities. In: Proc. ACM SIGCHI Conf. on Human Factors in Computing Systems.
- Oke, A. und Goffin, K. (2001): Innovation Management in the Service Sector. Management Focus, Cranfield School of Management, UK, Summer Issue.
- Osterloh, M. und Frost, J. (2000): Motivation in a Knowledge-Based Theory of the Firm, Working Paper, Universität Zürich.
- Phillips, W.; Noke, H.; Bessant, J. und Lamming, R. (2004): Beyond The Steady State: Managing Discontinuous Product and Process Innovation, AIM Working Paper No. 10, Advanced Institute of Management Research, London.
- Robinson, S. und Stern, B (1997): Corporate Creativity: How Innovation and Improvement Actually Happen. Koehler Publishers.
- Scanlon, J. (2003): If walls could talk, streets might join in. New York Times, 18.9.2003.

- Schlichter, J.; Koch, M. und Xu, C. (1998): Awareness - The Common Link Between Groupware and Community Support Systems. Community Computing and Support Systems (Toru Ishida eds.), Springer Verlag, S. 77-93, Jun. 1998.
- Snowdon, D. und Grasso, A. (2002): Diffusing information in organizational settings: learning from experience. Proc. ACM CHI'02 Conf. on Human Factors in Computing Systems, S. 331-338.
- Tidd, J.; Bessant, J. und Pavitt, K. (2005): Managing Innovation: Integrating Technological, Market and Organizational Change, 3rd ed., Wiley & Sons: Chichester, UK.
- Wenger, E. (1998): Communities of Practice: Learning, Meaning and Identity. Cambridge University Press.

Conceiving an environment for managing the lifecycle of collaborative business processes

Dirk Werth, Philipp Walter, Peter Loos

Institut für Wirtschaftsinformatik
im Deutschen Forschungszentrum für Künstliche Intelligenz
66123 Saarbrücken
{dirk.werth,philipp.walter,peter.loos}@iwi.dfki.de

Abstract

Business process lifecycle management is established for the continuous improvement of business processes within a single enterprise. However, the concept could also be applied to enhance collaborative business processes spanning over multiple enterprises. In contrast to the intra-organizational case, lifecycle management of cross-organizational collaborative processes imposes several organizational and technological challenges that results from the multiple-independent-actors-environment of collaborations. This paper focuses on these challenges and presents a conceptual solution for the different phases of this lifecycle. This concept has already been integrated into a prototypical software environment, that supports networked enterprises in the lifecycle management of collaborative business processes and that is presented in this paper.

1 Introduction

Classical industrial enterprises incorporated a high amount of the value generation within themselves. Due to the increasing complexity in management and manufacturing, enterprises nowadays tend to transform into smaller value units which are strongly specialized in their core competence [PrHa90]. Thus the added value is successively generated in networked structures [ÖsFl00], i.e. the companies intensively interact along the added value chain in order to produce the intended output conjointly. This intensification of exchanges leads to strong collaborative relationships (also called collaborative business, cf. [RöSc01]). So the ground is prepared for enterprise networks and virtual organizations [DaMa92]. Such collaborations are mainly driven

by the intention to generate added value, which is achieved through synchronized execution of associated business tasks. This activity sequence constitutes a collaborative business process.

Collaborative business processes [Wert06] are a special kind of conventional (intra-organizational) business processes. Like the conventional ones, they consist of a sequence of business activities. However, they possess special properties that strongly differ from the regular case. First, they are spanning over multiple organizations, because generation of added value is performed through cross-organizational division of labour. Second, each of the individual business activities that compose the process clearly belongs to a unique organization. Thus the collaborative business process can be partitioned into several parts, each of which contains one or more activities distinctly associated with an organization that fully controls this part in the sense that it independently executes, administrates and manages it. In this way those parts of cross-organizational business processes can be characterized as autonomous fragments.

Therefore collaborative business processes strongly differ from intra-organizational ones. Consequently concepts and solutions that are developed for the intra-organizational case are in most cases not suitable for cross-organizational purposes. This article investigates the aptitude of the business process lifecycle concept for such collaborative environments. After showing the gaps within the 'classic' lifecycle concept, we propose a platform which is apt to support the lifecycle for cross-organizational business processes. In the following sections the conceptual and technical basics of this platform are presented. In contrast to other approaches, e.g. [GrAb01], we do not focus bilateral processing of business processes only, but complete end-to-end processes. Therefore we will step through the three phases of the cross-organizational Business Process Lifecycle and show the concepts we developed for every phase. Afterwards we will show how the concept is realized so far and finish with a short outlook

2 The Business Process Lifecycle in Collaborations

For a continuous and successful business strategy it is insufficient to cover the design of business processes only, since the design solely results in static models of the considered processes which do not allow for process changes. However, execution of these static models usually yields improvement potential over time, e.g., because the execution context changed or certain execution aspects were not reflected in the model. To realize and quantify these improvement

potentials, it is necessary to measure execution of the models, i.e., perform controlling of them, which allows for identifying weaknesses and changing the models accordingly.

These three steps are integrated in the Business Process Lifecycle shown in Figure 1: business process design, business process implementation and business process controlling [ScJos02]. The basic lifecycle concept can be found in the House of Business Engineering [ScNü95;Sche96]. Business process design refers to modelling of existing as-is or intended to-be processes. This can be accomplished using modelling languages (e.g., EPC [KeNü92], BPML [Arki02]) and the respective modelling tools. Business process implementation summarizes all operative steps that are necessary to execute a process which was modelled before, including IT systems for execution as well as human interaction. Among the technical means for process execution are for example ERP systems and workflow engines. Research effort is currently put into the exploration of mechanisms to minimize the need for human interaction in business process implementation. Business process controlling denotes all actions that aim towards measurement and examination of running and finished processes with the goal of discovering optimization potentials. Once found, such a potential can be realized by changing the process model in the modelling phase of the next cycle pass.

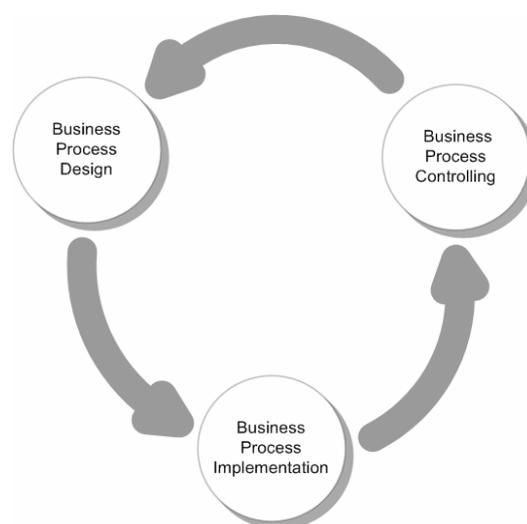


Fig. 1: Business Process Lifecycle

This lifecycle is conceived for a single organization. In the design phase, each process model is changed by a single modeller at a time. During execution, the process is handled by a single execution system within a single organization. Consequentially all controlling information can be gathered “indoor”, i.e., within the organization. However, in environments with multiple organizations acting cooperatively, collaborative processes cannot be regarded as monolithic

anymore, since the different parts of them are designed, executed and controlled by different organizations [LuBu99]. Consequently the lifecycle gets very complex and difficult to handle:

The *design* (respectively *modelling*) task comprises multiple autonomous modellers that act independently and follow different goals. This results in self-contained parts of the collaborative business process. Therefore the process design can rather be characterized as an assembly task of autonomous process parts.

The *execution* is distributed over different enterprises. Consequently there is no central processing engine. Instead each autonomous process part has its own independent processing engine, so classic workflow concepts and technologies have to be extended to match the new cross-organizational requirements [Schu02].

Controlling means monitoring of running and finished processes and comparing them with set values. However, monitoring in the sense of determining unique process states is impossible for collaborative workflows, because their state is hidden in the autonomous workflow engines. They only disclose virtual state information that clouds the real procedures. Moreover, the controlling comprises the aggregation and calculation of valuation functions. However, these functions contain information on business structures (esp. cost factors). Such information is considered business-critical and inaccessible to third parties, even if they are partners.

Having revealed these gaps, we will step through these three phases and show the concepts for collaborative business processes in the next section.

3 Conceiving a Cross-Organizational Business Process Lifecycle

Transferring the concept of lifecycle-based business process management to cross-organizational environments requires a shift from a centralized paradigm to a support for distributed environments, because cross-organizational business processes are characterized by the involvement of multiple actors in the different life-phases. For these actors a collective behaviour cannot be supposed. Thus each phase requires new techniques that are different to those of the classical business process management and that incorporate the split activities. Therefore we do not focus on bilateral processing of business processes, but on end-to-end processes with potentially a huge number of contributors.

3.1 Distributed Business Process Modelling

The design of business processes is considered a fundamental management task. In order to document the design, a specification medium is needed. On the conceptual level models have raised as the primary medium for business process specifications (e.g. EPC, BPML, BPEL, etc.). Thus the design task can be summarized as the creation of business process models. With regard to cross-organizational business processes, this actually comprises the model generation for an original that spans over multiple organizations. In principle this can be performed in a centralized and a decentralized way:

Supposing a centralized model creation, a single actor (that may also be incorporated by a group of collectively acting individuals) is responsible for the whole process model. This implies detailed knowledge of and unrestricted access to all aspects of the process. Due to the individual demand of secrecy, real-world organizations usually do not agree to fully expose their knowledge and processes to a third party. So this case can be considered implausible.

Assuming a decentralized model creation, this implies the existence of different modelling individuals, each of which generates only parts of the process. Within this procedure they may follow different modelling paradigms, methods and languages. Therefore this approach requires both a technique for assuring the consistent individual model creation and a technique for the integration of the partial models.

Another dimension is the direction of the model creation procedure. Here we distinguish between creating a model by more and more detailing an abstract description of the model object, and building a model by adding more and more aspects to it and aggregate it afterwards. So we can distinguish a top-down and a bottom-up approach in the model creation procedure:

Many approaches follow a *top-down modus operandi* for modelling cross-organizational business processes [AdCh05]. The foundation for this procedure is a blueprint model (a.k.a. reference model) of the cross-organizational business process which is to be implemented by the different participants. In a second step, each of them must adapt his process parts and refine them according to the blueprint. However, if we postulate independent organizations, i.e. they are legally independent and acting on their own behalf exclusively, this forced adaptation mechanism contradicts with the autonomy property. The presupposition of independent organizations fits most real-world collaboration scenarios, so the modelling procedure must not interfere with the autonomy of the individual organizations.

A *bottom-up approach* follows the reverse direction. It finds on as-is models that reflect the status quo of the participants. These detail models are reduced by removing information that is confidential and not adequate for third party usage. These “alienated” models are composed in a second step and integrated into an overall model. The main advantage of this approach is the elimination of the need to adapt. It satisfies the autonomy postulation, because no participant is obliged to adapt their models and consequently no modifications in the processes have to be performed. In this context, the detail model that constitutes the base of this procedure reflects existing process capabilities, i.e. the ability of a participant to perform a business process. These capabilities are encapsulated in modules from which a new collaborative business process is composed. In order to include flexibility mechanisms in the assembly procedure, the individual actor can incorporate its readiness to adaptation within its module design.

Because these two dimensions in model creation are independent of each other, they can be aligned orthogonally in a 2x2 solution matrix, as shown in Table 1. Reflecting the different characteristics of the two dimensions, neither a centralized modelling nor a top-down approach seems to be appropriate for the special organizational environment of collaborations. In our concept we therefore follow a bottom-up approach using a decentralized modelling procedure (Option 4).

	Centralized Modelling Single actor in charge of whole process model (overall knowledge and unrestricted access)	Decentralized Modelling Multiple modellers generate partial models (for their specific area)
Top-Down From a master plan to the operational realization	1	2
Bottom-Up From the capabilities to the overall inter-working plan	3	4

Tab. 1: Procedure options for model creation in multi-actor-environments

Although from a theoretical perspective, such an approach has to cope with all potential permutations of modelling techniques, our approach is a homogeneous one based on a single modelling language. Even in this scenario there are sufficient degrees of freedom for the modelling

subject. In our implementation the event-driven process chain (EPC) language is used. Since EPC is one of the most common process model languages (at least in Europe), this seems to be a suitable assumption. More precisely, our design procedure comprises four steps:

1. *Definition of process modules:* In contrast to the top-down approach described above, we start with the assessment of the status quo of the different organizations involved by specifying their capabilities. In our case they have to express their ability to produce output using process models that describe their possible processing sequences. The results are component-like models that can be assembled together and that incorporate process interface descriptions specifying interaction points.
2. *Definition of process intentions:* The composition of process modules has to follow certain business objectives. In order to construct an objective-adequate process model, the intention of the underlying process must be defined. In particular this addresses the output the process has to deliver as well as the organizational constraints (e.g., the whole process has to be performed within the EU).
3. *Process module composition:* The composition itself is performed by analyzing compatibility of process interface pairs. That yields pairs of matching interfaces through which process modules can be connected. Based on those modules which are able to produce the intended outcome, a network of modules is successively constructed. This finally results in a set of modules that generate the final product. Thus the composition is directed by the matching assignments of the process interfaces. The set is filtered by the organizational constraints of step 2 and rated by a common target function. The best rated result is the final one and describes a common cross-organizational business process model for all participants.
4. *Process model consistency analysis:* To avoid contradictions within the overall process model, the composition phase closes with a consistency analysis during which the model is analyzed with respect to flow logic consistency. Such a test is described for example in [SaOr99]. Having passed this test, the cross-organizational business process model can be realized within all involved organizations.

3.2 Distributed Business Process Execution

The distributed execution of a business process starts with a common process model that all participants share and that is business oriented, i.e., its content is mainly conceptual and its purpose is organizational management. From this model every participant extracts those parts that he has to execute and augments them with arbitrary information he needs for execution, e.g., refinements of process sub-parts or execution context parameters (cf. Figure 2). Thus the business model is transformed into an IT-oriented workflow model, the main purpose of which is the execution of the contained process. The following section introduces the steps from the common process model to execution of the workflow model:

1. *Splitting Up the Common Process Model:* All activities in the common process model are annotated with the executing organization unit (“Company X”), or with an organization unit role (“Customer”) that can be mapped onto a concrete actor within the execution context. So the common model disaggregates in disjoint process model fragments that are executed by exactly one actor each. Because the process modules, which were composed to the common process model during the modelling phase, have interface descriptions, it is possible to define exactly which goods and which information must be transferred from one actor to another.
2. Apart from goods and information, the execution of the whole process devolves from one actor to another at an interface. Therefore it is necessary to define how the control of the process is transferred. At process junctions it may be even possible to split up process control or join multiple execution threads again.
3. *Augmenting the Process Fragments:* Execution of a process fragment usually requires considerable prearrangements on the part of the executing actor. Therefore the process fragment is first transformed from the modelling language into an executable language. Since the business process model is business oriented, it usually does not contain information about execution parameters, e.g., an IP address of an interface or authentication credentials for an ERP system. So it must be augmented with these missing execution parameters during or after transformation to the executable language. After transformation and augmentation, the process fragment is contained in an executable workflow model.

4. Usually the common business process model disaggregates into multiple process fragments, each of which is transformed into a single workflow model. These workflow models are deployed to the respective IT systems then, which are finally configured with the contained information.
5. *Executing the Process*: Figure 2 shows how the whole top-level process is implemented by executing the workflow models of the process fragments which it consists of. After configuration of all involved systems this happens automatically, i.e., without interaction with individual process instances.

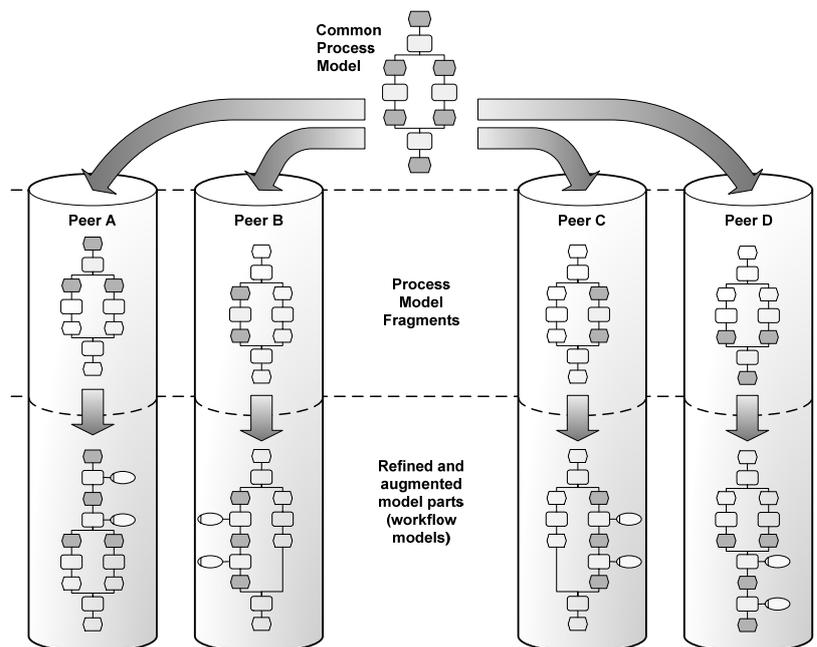


Fig. 2: Distributed business process execution

Since the whole process is executed fragment-wise by multiple separate systems, there must be transition points from one system to another where execution is finished or suspended at the source system and perpetuated at the target system. This transition has two different aspects: data flow and control flow. Data transfer between separate IT systems is widely used already, e.g., between departments within a single organization. However, the transfer of process execution control and context via push and pull mechanisms is not common. Especially in split and join situations, e.g., when a simultaneous execution of multiple process parts on multiple systems begins or finishes, the process context must be duplicated and merged accordingly. During execution, performance data is gathered as a means for the next lifecycle step: the controlling phase.

3.3 Distributed Business Process Controlling

From the management perspective, the ability to execute a business process is not sufficient. In order to improve the design and the way of execution it is essential to measure the target object, i.e., to reveal performance indicators of the cross-organizational business process. In the intra-organizational case, this means to extract historical execution information from a single process execution system (mostly a workflow management system) and to calculate the performance indicators from them. In contrast to that, the cross-organizational case is rather complicated. On the one hand there are multiple execution systems, each of which holds only partial information about the execution of a single cross-organizational business process. Thus the challenge is not only to compose performance data from multiple sources, but also to identify linked process chunks and to reconstruct the complete structures of historical cross-organizational business processes under the side condition of heterogeneous keeping of data and system ownership. On the other hand this information on the reconstructed process not necessarily leads to performance indicators for the whole process, because the calculation of these indicators requires the valuation of process execution data. However this valuation (e.g., the cost function) is usually considered a business secret, so an overall indicator processing cannot be performed without exposing individual business knowledge. Therefore we propose to calculate distributed performance indicators in a way equivalent to the execution data processing: each organization transforms the process information gathered from the execution systems into its individual (partial) performance indicators. These figures will then be used to compute the overall indicators. Following this procedure, the organizations are not obliged to publish their calculation scheme and only communicate the resulting values.

4 Technical Realization

In this section the realization of the concepts described above will be presented. Within the research project P2E2 (Peer-to-Peer Enterprise Environment¹), a platform has been developed that prototypically implements the distributed Business Process Lifecycle management principles. The basic idea is to form a network of actors (“peers”) which are all equal with respect to rights and what they are able to do [ScFi02]. The network is dynamic, i.e., peers may enter and leave

¹ P2E2 is funded under the SE2006 initiative by the German ministry of education and research (BMBF).

the network at any time. The peer-to-peer principle guarantees equal opportunities for all participating parties. Every party distributes models of the processes that it offers to perform. A customer peer can reassemble these process fragments to the model of a complete process and buy the execution of it (or parts of it) from other peers. Thus the P2E2 network structurally corresponds to the organizational network of the collaborating organizations and therefore provides a wide set of advantages as a technological base for enterprise networks [KuWe04].

4.1 Distributed Business Process Modelling

First, the processes offered in the network must be modelled, aggregated, assembled and so on. The top-level modelling language used in the P2E2 prototype is the event-driven process chain (EPC). Modelling is performed using the ARIS Toolset by IDS Scheer AG. However, the P2E2 meta-model explicitly supports other modelling languages, too.

In the first step, every peer designs his own processes in any desired detail, thus obtaining a “private” model which can contain arbitrary (even secret) information about the process and therefore is not shared with other peers. Then he generates a “public” view to the model by reducing the contents of the private model to the minimum that is necessary for other peers to comprehend the modelled process and its interfaces.

In the next step, all public models by all actors are distributed among the network. For this purpose we developed the Process Distribution and Discovery Tool (PDDT), a peer-to-peer software which is based on the JXTA peer-to-peer framework and supports distributing, versioning, searching and transferring models (see Figure 3). With the shared information about the available process fragments, any peer can construct a complete process from the fragments. Using the PDDT again, this common process model is shared with all peers that participate in its execution.

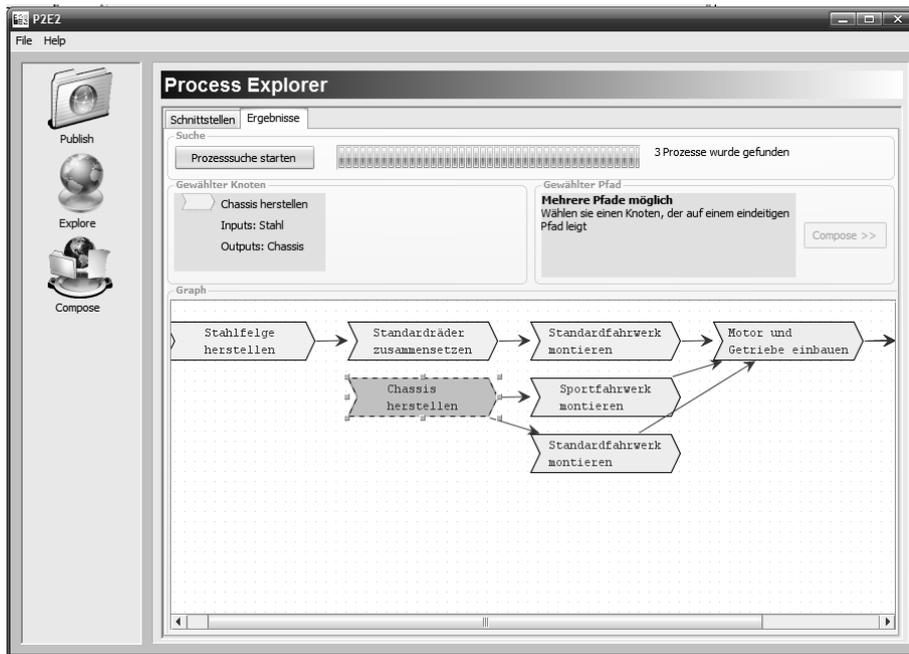


Fig. 3: Screenshot of PDDT

4.2 Distributed Business Process Execution

Figure 4 shows the architecture of a P2E2 peer along with the controlling and configuration applications which are not an integral part of the peer itself. This subsection about execution starts with the output of the modelling tool in the lower left corner of the figure.

In P2E2, the execution part of the lifecycle is simplified compared to the scenario outlined in Section 2, because the common process model is composed from several process fragments. So the responsibilities for the execution of the process parts are ex ante established and partitioning the common process can be omitted, because the fragments already exist. The augmentation of the process fragments with execution information also benefits from the fact that the private model with all execution details already exists. So it is sufficient that every peer augments its process fragments once and reuses this information in every execution.

Another part of the augmentation phase is the conversion of all models into a common execution model language, i.e., XPDL in our case: finally, all P2E2 process fragments exist as executable XPDL models. To obtain the final XPDL models, a multi-stage conversion and augmentation is performed. First, the EPC models are automatically converted into XPDL format using the modelling tool. Then the attributes of all XPDL model elements are filled in with data necessary for execution using another tool developed within the project, which is named “augmentation tool” in Figure 4.

Execution in P2E2 is finally performed using workflow engines by Carnot AG and abaXX Technology AG (“WFMS” in Figure 4). Whenever necessary, communication between executing peers is performed by calling BAPI methods using Wf-XML.

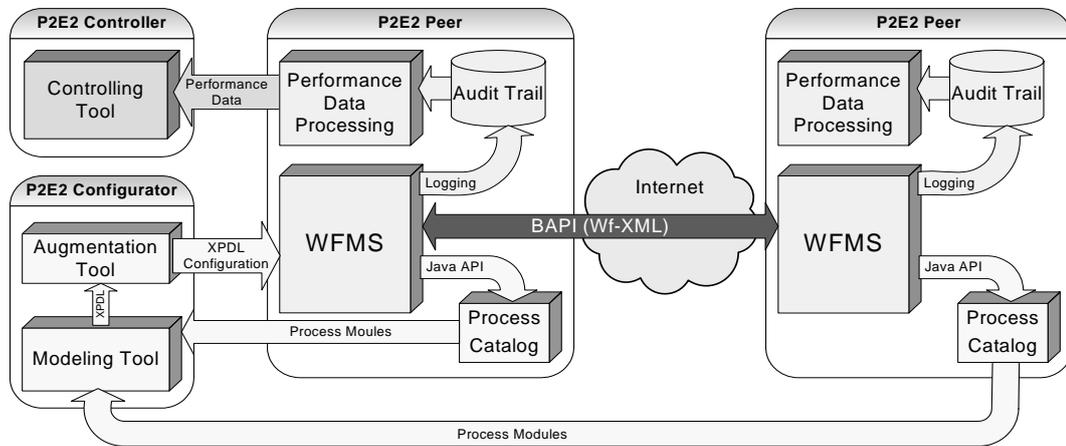


Fig. 4: P2E2 Technical Architecture

4.3 Distributed Business Process Controlling

During execution, every engine records performance data and stores it for the third lifecycle phase: controlling. The most basic performance data gathered during execution is stored in the audit trails of the workflow management systems (see Figure 4). However, mainly due to business secrecy, their content is not exposed directly. Instead, every peer processes its performance data to its liking and exposes the results or parts of the results over a specific web service interface exclusively. Of course, this information only refers to the execution of a process fragment, not the process as a whole.

The reassembly from fragments to the whole process is achieved using a specific controlling tool (see Figure 4). It first fetches performance information about process fragments from all participating peers using the web service described above. Then the information how the whole process is composed from process fragments is used to aggregate per-process information from per-fragment data.

5 Related Work

The approach presented tries to bring together several research areas that originally are addressed isolated. The concept of distributed business processes has raised ten years ago (e.g. [GrGr96], [AaWi98]). It was mainly driven by distributed system research and tried to archive the cross-system execution of workflows (e.g. [BaDa97], [ScSt94]). Such attempts also resulted in the definition of various standards (e.g. WF-XML) to simplify the interoperability of workflow management systems (cf. [WFMC96]). But they assume the existence of a single, atomic workflow specification model (e.g. [MuWo98]). On the other hand exists various approaches of distributed business process resp. workflow modelling (e.g. [GrGr95]). They describe the creation of singular models by multiple actors. But they mainly miss either the link to the distributed execution or the interconnection to the controlling task. Especially this task is neglected in other management approaches to cross-organizational business processes [DaHs01; LeRo02; PeK199].

6 Conclusion and Outlook

In this paper, we have presented a concept for the cross-organizational business process lifecycle management, including distributed modelling, execution and controlling, that is already implemented in most parts. In particular we addressed and ensured the continuous IT support of all three lifecycle phases, the decision autonomy and secrecy demand of the participating organizations during all three lifecycle phases, and the technical and conceptual feasibility of our approach (which will be finally verified when the entire prototype is completed).

Currently, two business scenarios are evaluated with our concept. One of them is taken from the financial services sector and deals with factoring, the other one deals with supply chain management in international and national product distribution.

This concept was developed at the Competence Centre Business Integration (CCBI), Institute for Information Systems (IWi) at the German Research Centre for Artificial Intelligence (DFKI), Saarbruecken. The work is performed by clustering national and international funded research projects (esp. ArKoS, ATHENA, INTEROP, P2E2), intending the development of solutions for a better interoperability in business networks.

References

- [AaWi98] van der Aalst, Willibrordus M. P.: Interorganizational Workflows. In: Proceedings of the PROLAMAT 98., 1998, pp. 21-43
- [AdCh05] Adam, O.; Chikova, P.; Hofer, A.; Vanderhaeghen, D.: Customer-Driven Process Management in Value-Added Networks Using an Architecture for Collaborative Business. In: Karduck, A. P. (ed.): Proceedings of COLLECTeR (Collaborative Electronic Commerce Technology and Research) Europe 2005 - Collaborative Business, Session 2: Customer-Supported Activity Coordination, 2005, pp. 1-9.
- [Arki02] Arkin, A.: Business Process Modelling Language-Business Process Management Initiative. 2002.
- [BaDa97] Bauer, T.; Dadam, P.: A Distributed Execution Environment for Large Scale-Workflow Management Systems with Subnets and Server Migration. In: Proceedings of the 2nd IFCIS Conference on Cooperative Information Systems., 1997
- [Cama02] Camarinha-Matos, L. M.: Collaborative Business Ecosystems and Virtual Enterprises. Kluwer Academic, Norwell 2002.
- [DaHs01] Dayal, U.; Hsu, M.; Ladin, R.: Business process coordination: State of the art, trends, and open issues. In: Proceedings of 27th International Conference on Very Large Data Bases., 2001
- [DaMa92] Davidow, W. H. & Malone, M. S.: The Virtual Corporation - Structuring and Revitalizing the Corporation for the 21st Century. Harper Collins, New York 1992.
- [GrAb01] Grefen, P., Aberer, K., Hoffner, Y., & Ludwig, H.: CrossFlow: Cross-organizational Workflow Management in Dynamic Virtual Enterprises. In: International Journal of Computer Systems, Science and Engineering 15 (2001) 5, pp. 277-290.

- [GrGr95] Graw, G.; Gruhn, V.: Distributed Modeling and Distributed Enaction of Business Processes. In: Proceedings of the 5th European Software Engineering Conference. Berlin et al. Springer, 1995, pp. 8-27
- [GrGr96] Graw, G.; Gruhn, V.; Krumm, H.: Support of cooperating and distributed business processes. In: Parallel and Distributed Systems., 1996, pp. 22-31
- [KeNü92] Keller, G., Nüttgens, M., & Scheer, A.-W.: Semantische Prozessmodellierung auf der Grundlage "Ereignißgesteuerter Prozeßketten (EPK). In: Veröffentlichungen des Instituts für Wirtschaftsinformatik, A.-W. Scheer 1992.
- [KuWe04] Kupsch, F.; Werth, D.: Process-Driven Business Integration Using Peer-to-Peer Technology. In: Mehdi Khosrow-Pour (ed.): Innovations Through Information Technology., 2004, pp. 1294-1300.
- [LeRo02] Leymann, F.; Roller, D.; Schmidt, M.-T.: Web services and business process management. In: IBM Systems Journal 41 (2002), No. 2, pp. 198-211
- [LuBu99] Ludwig, H., Bussler, C., Shan, M., & Grefen, P.: Cross-Organizational Workflow Management and Co-ordination - WACC99 Workshop Report. In: ACM ASIGGROUP Bulletin 20 (1999) 1.
- [MuWo98] Muth, Peter; Wodtke, Dirk; Weissenfels, Jeanine; Kotz-Dittrich, Angelika; Weikum, Gerhard: From Centralized Workflow Specification to Distributed Workflow Execution. In: Journal of Intelligent Information Systems 10 (1998), No. 2, pp. 159-184
- [ÖsFl00] Österle, H., Fleisch, E., & Alt, R.: Business Networking - Shaping Enterprise Relationships on the Internet. Springer, Berlin et al 2000.
- [PeK199] Pereira Klen, A. A.: Distributed Business Process Management. In: Proceedings of the IFIP / PRODNET Working Conference on Infrastructures for Virtual Enterprises: Networking Industrial Enterprises. Kluwer, 1999, pp. 241-258
- [PrHa90] Prahalad, C. K. & Hamel, G.: The core competence of the corporation. Springer, Berlin et al. In: Harvard Business Review 63 (1990) 3, pp. 79-91.

- [RöSc01] Röhrich, J. & Schlögel, C.: cBusiness. Erfolgreiche Internetstrategien durch Collaborative Business. Addison-Wesley 2001.
- [SaOr99] Sadiq, W.; Orłowska, M.E: Applying Graph Reduction Techniques for Identifying Structural Conflicts in Process Models. In: Proceedings of the 11th International Conference on Advanced Information Systems Engineering (CAiSE '99). Springer, 1999, pp. 195-209.
- [Sand99] Sandler, T.: Collective Action - Theory and Applications. University of Michigan Press 1999.
- [Sche96] Scheer, A.-W.: ARIS-House of Business Engineering: Von der Geschäftsprozeßmodellierung zur Workflow-gesteuerten Anwendung ; vom Business Process Reengineering zum Continuous Process Improvement. In: Scheer, A.-W. (ed.): Veröffentlichungen des Instituts für Wirtschaftsinformatik, No. 133, Saarbrücken : Universität des Saarlandes, 1996.
- [ScNü95] Scheer, A.-W.; Nüttgens, M.; Zimmermann, V.: Rahmenkonzept für ein integriertes Geschäftsprozeßmanagement. In: Wirtschaftsinformatik 37 (1995), No. 5, pp. 426-434.
- [ScJo02] Scheer, A.-W. & Jost, W.: ARIS in der Praxis - Gestaltung, Optimierung und Implementierung von Geschäftsprozessen. Springer, Berlin et al 2002.
- [ScFi02] Schoder, D.; Fischbach, K.: Peer-to-Peer : Anwendungsbereiche und Herausforderungen. In: Schoder, D.; Fischbach, K.; Teichmann, R. (eds.): Peer-to-peer: ökonomische, technische und juristische Perspektiven. Springer, 2002, pp. 3-21.
- [Schu02] Schulz, K.: Modelling and Architecting of Cross-Organizational Workflows, PhD Thesis. University of Queensland 2002.
- [ScSt94] Schuster, Hans; Jablonski, Stefan; Kirsche, Thomas; Bussler, Christoph: A client/server architecture for distributed workflow management systems. In: Proceedings of the 3rd International Conference on Parallel and Distributed Information Systems., 1994, pp. 253-256

- [Wert06] Werth, D.: Kollaborative Geschäftsprozesse - Integrative Methoden zur modellbasierten Deskription und Konstruktion. PhD Thesis, Saarland University 2006.
- [WFMC96] Workflow Management Coalition: Workflow Standard - Interoperability, Abstract Specification, TC-1012, Version 1.0., 1996

Infrastrukturen zur Aneignungsunterstützung

Ein Konzept zur Integration von produkt- und prozessorientierter Flexibilisierung

Gunnar Stevens¹⁾; Volker Wulf^{1,2)}; Volkmar Pipek²⁾

¹⁾Fraunhofer-Institut für Angewandte
Informationstechnik FIT
53754 Sankt Augustin
{gunnar.stevens@fit.fraunhofer.de}

²⁾Wirtschaftsinformatik
Universität Siegen
57068 Siegen
{volker.wulf, volkmar.pipek}@uni-siegen.de

Abstract

Die Aneignung und Einbettung von Computersystemen in den Nutzungskontext stellt einen zentralen, nichtsdestotrotz von der bisherigen Forschung ungenügend adressierten Aspekt der Softwarenutzung dar. Es muss dabei klar sein, dass Aneignung ein kollektiver und kreativer Prozess gesehen werden muss. In diesem Beitrag stellen wir das Konzept der Aneignungsinfrastruktur vor. Dabei handelt es sich um eine sozio-technische Methode, die auf die Kombination hoch anpassbaren Softwaredesigns mit der Nutzung einer Kommunikations- und Kooperations-Infrastruktur innerhalb der Nutzercommunity sowie zwischen Benutzern und Entwicklern aufbaut. Die Kommunikations- und Kooperationsinfrastruktur ermöglicht es, nicht antizipierte Anforderungen des Nutzungskontexts zu fassen und in einen agilen Entwicklungsprozess einfließen zu lassen. Die dadurch möglich gemachte technische Flexibilität unterstützt den Umgang mit differenzierten und dynamisch wechselnden Anforderungen. Anhand des BSCWeasel-Projektes zeigen wir, wie eine entsprechende Aneignungsinfrastruktur realisiert werden kann.

1 Einführung

Die Forschung, die durch das *Produktivitätsparadox* angeregt wurde, hat das Verständnis für die wirtschaftlichen Effekte von Informationstechnologien enorm verbessert. So zeigen

Brynjolfsson und Hitt, dass die meisten Produktivitätseffekte nicht direkt auf Computer-Anwendungen zurückzuführen sind, sondern auf die spezifische Aneignung der Artefakte in der organisationalen Umwelt [Bryn98]. Aneignung von Technologien wird dabei als kreativer Prozess des Anwenders – sowohl Organisation, als auch Endbenutzers - verstanden, der auch über die zugehörigen intendierten Nutzungsmöglichkeiten der Softwareentwickler hinausgehen kann [Pipe05]. Der erfolgreiche Prozess der Technologie-Aneignung führt dabei zu neuen Arbeitsweisen, innovativen Geschäftsprozessen oder veränderten organisatorischen Strukturen. Da die organisationale Umgebung sich typischerweise während des Aneignungsprozesses ändert und eine erfolgreiche IT-Aneignung zusätzliche Auswirkung auf die Arbeitspraxis, Geschäftsprozess oder organisatorische Strukturen hat, sind die IT-Anforderungen im Prozess nicht stabil [Orli96, Pip+99]. Deshalb muss sich ein Ansatz, der die IT-Aneignung unterstützt, mit den Dynamiken des Anwendungskontexts auseinandersetzen. Daneben muss die Ausdifferenzierung von Anforderungen aus verschiedenen Anwendungsfeldern berücksichtigt werden. Darüber hinaus sollte das Potential der Benutzer berücksichtigt werden, die aufgrund ihrer Praxiskenntnis und ihrer Nutzungserfahrung im kreativen Handeln innovative Lösungen entwickeln [vHip88].

Im Folgenden zeigen wir einen Ansatz, der einen Brückenschlag zwischen entwicklungsorientierter und nutzungsorientierter Flexibilisierung schlägt. Dabei werden wir zunächst den aktuellen Stand in der Diskussion innerhalb von HCI und SE aufarbeiten. Dann werden wir ein allgemeines Framework für die Aneignungsunterstützung vorstellen und anschließend anhand des Groupwaresystems BSCWeasel zeigen, wie sich das darlegte Konzept in die Praxis umsetzen lässt.

2 Aneignung und Flexibilisierung

Im Folgenden wird der Zusammenhang zwischen Aneignungsprozessen und den technischen Konzepten zur Flexibilisierung darlegt. Das Studium der Literatur zeigt, dass das Thema der Flexibilisierung dabei zumeist unter zwei unabhängig voneinander betrachteten Perspektiven – der produktorientierten und der prozessorientierten Flexibilisierung - angegangen wird. Doch aufgrund der unter der jeweiligen Perspektive erzielten Fortschritte scheint es nun angebracht, die beiden Perspektiven wieder zu synthetisieren.

2.1 Aneignungsprozesse

Aneignungsprozesse bei Groupware wurden in verschiedenen Langzeitstudien untersucht [Orli96, Pip+99, Kar+98, Wulf99, Tör+03]. Diese Fallstudien geben ein gutes empirisches Bild der Aneignungsprozesse und den daraus resultierenden Änderungen in der Arbeitspraxis, Geschäftsprozessen und organisatorischen Strukturen.

Orlikowski und Hofmann bieten ein konzeptionelles Modell an, um die organisatorischen Änderungen, die aus der Aneignung von Groupware resultieren, zu verstehen [Orl+97]. Sie unterscheiden drei verschiedene Typen von im Prozess auftretenden Änderungen: *Anticipated changes* sind organisatorische Transformationen, die geplant sind und während der Groupware-Nutzung in der Organisation auch umgesetzt werden. Im Gegensatz dazu stehen die *opportunity-based changes*, die bei der Einführung nicht vorhergesehen waren, sondern auf der Entdeckung von Potentialen des Systems im Verlauf der Aneignung und darauf aufbauender zielgerichteter Weiterentwicklung beruhen. Demgegenüber sind *emergent changes* nicht nur in der Einführungsphase nicht vorhersehbar, sie finden insbesondere unter der Hand in dezentralen und ungeplanten Aktivitäten statt.

Die gelegheitsbasierten, und emergenten Änderungen bedürfen einer Aneignungsinfrastruktur, um sich als Potential lokal sich entwickelnden Innovation entfalten zu können. Diese Formen nicht vorhergesehener Aneignung geschehen immer dann, wenn ein Einzelner oder eine Gruppe von Benutzern eine technische Funktionalität auf innovative Weise in Praxis integrieren kann. Oft löst dabei eine externe Irritierung die Reflektion über die Anwendung und die herrschende Praxis aus. Typischerweise lernen dabei die Benutzer voneinander, wobei sich benutzer-getriebene Innovationen über lokale Wissensnetzwerke verbreitern.

Mit dem Aufkommen vernetzter, kooperativer Applikationen ist der Bedarf an flexiblen, gruppenanpassbaren Anwendungen gestiegen [Wul+95, Ben+95]. Zugleich stellt die Verteiltheit der Akteure stellt auch neue Anforderung an die Gestaltung der Anpassbarkeit der Systeme [Ober94]. Einige dieser Herausforderungen wurden durch Pipek mit seinem Konzept der Diskursinfrastrukturen angegangen [Pipe05]. Ein integrierter Ansatz zur Aneignungsunterstützung steht jedoch noch aus.

2.2 Produkt-orientierte Flexibilisierung

Die HCI-Forschung betrachtet Flexibilität hauptsächlich als ein Produktmerkmal, das es erlaubt, Anwendungen innerhalb ihres Nutzungskontextes anzupassen [Hen+91, Teeg00]. Das Anpassen findet nach der eigentlichen Design- und Implementierungsphase statt. Typischerweise beginnt es direkt nach der Installation oder während der Installation einer Anwendung. Das Anpassen erfolgt normalerweise als Zusammenarbeit von normalen Benutzern, lokalen Experten, System Support oder Helpdesk-Personal, wobei man drei Ebenen der Komplexität -Parametrisierung, Erweiterung und Neuprogrammierung - unterscheidet [Mørre97].

Die Forschung zu produkt-orientierter Flexibilisierung fokussiert dabei vor allem auf den Entwurf von geeigneten Softwarearchitekturen, der Gestaltung von Anpassungswerkzeugen und der kognitiven Anforderungen an den Endbenutzer. Verschiedene hoch anpassbare Systeme - kommerzielle wie Forschungsprototypen - wurden hierzu entwickelt, wie z.B. OVAL [Mal+95], oder FreEvolve [Stiem00].

Die Frage, wie Flexibilitätsanforderungen in geeigneter Weise erhoben werden, wird jedoch selten bis gar nicht adressiert. Dieser blinde Fleck liegt unter anderem daran, dass die Ansätze zur Gestaltung hoch anpassbarer Systeme Fragen der Flexibilität von Produkten nicht mit Fragen der Flexibilisierung von Entwicklungsprozessen verbinden.

2.3 Prozess-orientierte Flexibilisierung

Selbst die Gestaltung hoch anpassbarer Systeme kann nicht sämtliche Nutzungssituation und die daraus abgeleiteten Flexibilitäts-Anforderungen vorhersehen. Deshalb kann es immer wieder vorkommen, dass die bereitgestellten Anpassungsmöglichkeiten der Systeme nicht ausreichen. Deshalb sollte auch der Entwicklungsprozess auf nicht antizipierte Anforderungen vorbereitet sein. Insbesondere bedarf es hierfür einer Infrastruktur, die es erlaubt, solche nicht antizipierten Situationen im Nutzungskontext überhaupt erst einmal sichtbar zu machen.

Agile Vorgehensmodelle, wie eXtreme Programming (XP) [Beck00], gehen davon aus, dass Anforderungen sich permanent ändern können und haben deshalb verschiedene, aufeinander abgestimmte Techniken entwickelt, die es erlauben, den Änderungswünschen mit vertretbaren Kosten nachzukommen [Beck00].

Der Ansatz des Radical Tailoring [Mal+95] versucht, sämtliche zukünftige Anforderungen mittels hoch anpassbarer Produkte zu erschlagen. Demgegenüber verfolgt XP die Strategie, hierauf mit hoch agilen Entwicklungsprozessen zu reagieren. Damit stellt XP sicherlich die bekannteste und extremste Form einer prozess-orientierten Flexibilisierungsstrategie dar.

Da sich aber XP als reines Ausführungsorgan der Kundenanforderung versteht, kümmert sich XP nicht selbst um Aneignungsprozesse im Nutzungskontext. Insbesondere findet man keine Überlegungen zu einer geteilten Infrastruktur, die einen gemeinsamen Lernprozess von Benutzern und Entwicklern unterstützt. Stattdessen erhalten Entwickler nur indirekt Feedback über das *Customer on Side*-Prinzip [Beck00]. Erschwerend kommt hinzu, dass dieses Prinzip in der Praxis schwer zu realisieren ist [Rum+01].

Von partizipativ orientierten evolutionären Prozessmodellen wie STEPS [Flo+89] unterscheidet sich XP insbesondere im Verständnis der Beziehung von Entwicklern und Nutzern. Die Kernidee von STEPS beruht diesbezüglich darin, Softwareentwicklung als ein fortdauernden gegenseitigen Lernprozess aufzufassen, bei dem sich Software in enger Zusammenarbeit von Benutzern und Entwicklern in diversen Releases entwickelt. Die kontinuierliche Verbesserung sollte dabei auf Grund von konkreten, alltagsweltlichen Nutzungserfahrungen erfolgen. STEPS ist jedoch ein sehr allgemeines Modell. Die Realisierung der unterschiedlichen Teile des Modells wurde innerhalb von STEPS auch nicht weiter ausgearbeitet.

Insbesondere auf der konzeptionellen Ebene unberücksichtigt blieb in STEPS jedoch der Aspekt heterogener Umwelten, an die das Produkt im Nutzungskontext jeweils individuell angepasst werden muss. Eine Erweiterung in diese Richtung findet man aber im erweiterten STEPS Modell von Wulf und Rohde [Wul+95]. Auf der softwaretechnischen Ebene fanden Ideen des erweiterten STEPS-Modells Eingang in den Forschungsprototypen FreEvolve [Stie00]. Jedoch ist auch bei FreEvolve keine Aneignungsinfrastruktur integriert, die es erlaubt, nicht antizipierte Aneignungsphänomene und Anpassungsbedarfe zu erfassen.

Andere relevante Arbeiten sind Fischers konzeptionellen Überlegungen zur der Erweiterung seines Konzepts der End-User Modifiability [Fis+90]. Ähnlich dem erweiterten STEPS Modell schlägt er in den Arbeiten zum Prozessmodell SER vor, dass die produktorientierte Flexibilisierung Teil eines evolutionären Entwicklungsprozesses sein sollte [Fis+02]. Er geht dabei von einem Phasenmodell aus, bei dem sich Nutzung und kooperative Gestaltung abwechseln. Jedoch finden sich in seinen Arbeiten nur vereinzelt Hinweise darauf, wie eine solche Zusammenführung beider Flexibilisierungsstrategien aussehen könnte.

Zusammenfassend lässt sich feststellen, dass in den neueren Arbeiten ein Trend zu erkennen ist, Ansätze zur produkt-orientierter Flexibilisierung mit prozess-orientierter Flexibilisierung zu kombinieren. Jedoch ist in den Arbeiten kein klarer Ansatz herausgearbeitet, der die Lücke zwischen den beiden Perspektiven überbrückt.

3 Vermittlung zwischen prozess- und produkt-orientierter Flexibilisierung: Das Konzept einer Infrastruktur zur Aneignungsunterstützung

Im Folgenden wollen wir unser Konzept einer technischen Infrastruktur zur Unterstützung von Aneignungsprozessen in Organisationen umreißen. Der Begriff Infrastruktur adressiert dabei genau jenes offene Problem bisheriger Ansätze, produkt-orientierte mit prozess-orientierter Flexibilisierung zu verbinden.

Das Konzept verfolgt dabei zwei grundsätzliche Ziele:

- die Förderung des Wissensaustausch innerhalb der Nutzercommunity
- die Förderung einer reziproken Partizipation am Nutzungs- bzw. Entwicklungskontext

Durch die Förderung einer reziproken Partizipation soll dabei die Chance erhöht werden, innovative Nutzungen als auch eventuelle Nutzkrisen zu erkennen und aufgrund geeigneter Produktmodellierung (wie z.B. durch komponenten-basiertes Software Engineering) und geeigneter Prozessmodellierung (wie z.B. eXtreme Programming) auf eine effiziente Art und Weise reagieren und die Anwendung an den lokalen Kontext anpassen bzw. erweitern zu können.

Bei der Gestaltung der Infrastruktur gilt es zu beachten, dass sowohl die Nutzercommunity, als auch Benutzer und Entwickler üblicherweise räumlich sowie zeitlich verteilt sind, da Aneignung ein kontinuierlicher, nicht auf einen bestimmten Zeitraum beschränkter Prozess ist. Zum anderen gilt es zu berücksichtigen, dass hier zwischen Akteuren mit deutlich unterschiedlichen Arbeitspraxen, Interessen, Wissenshorizonten und Kulturen vermittelt werden soll.

Bei der Vermittlung zwischen heterogenen Kulturen, haben Star und Griesemer anhand einer Fallstudie gezeigt, dass sogenannten *Boundary Objects* eine besondere Rolle zur Lösung des Vermittlungsproblem zukommt [Sta+89]. *Boundary Objects* zeichnen sich dabei dadurch aus, dass sie auf der einen Seite eine gemeinsam geteilte Entität in den verschiedenen Kulturen darstellen, dass sie aber auf der anderen Seite von den jeweiligen Akteuren anders betrachtet und genutzt werden können. Strübing erläutert anhand eines Betatests des Programms ATLAS/ti¹, wie das Artefakt selbst die Rolle eines *Boundary Object* innerhalb des Gestaltungs- und Nutzungsdiskurs einnehmen kann [Strü99].

Im Falle der Aneignungsunterstützung von Softwareartefakten gilt es unseres Erachtens darum, dass Konzept von *Boundary Objects* mit dem softwareergonomischen Konzept der *Direct*

¹ Vgl. <http://www.atlasti.de/>

Manipulation zu verbinden. Das von Ben Shneidermann eingeführte Konzept besagt im Kern, dass die für den Nutzungskontext relevanten Dinge, die so genannte *Objects of Interest*, am User Interface visuell präsentiert werden und dort auch direkt bearbeitbar sein sollten [Shne83]. Umgekehrt zeigte sich in unserer Evaluation, dass die symbolische Interpretation des Artefakts durch das an der Benutzerschnittstelle sichtbare Verhalten geprägt wird und die visuelle Repräsentation in der Kommunikation zwischen den Akteuren eine besondere Bedeutung zukommt.

Für die technische Umsetzung der Aneignungsunterstützung lassen sich aus der Verbindung beider Konzepte folgende Anforderungen ableiten:

- sollten die Nutzungs- und Gestaltungsdiskurse am User Interface sichtbar gemacht werden und direkt aus der Nutzung zugänglich sein [Wul+01]
- sollte bei der Rolle des Artefakts als *Boundary Object* insbesondere seine visuelle Repräsentation und den Aspekt des *Direct Manipulation* mit einbeziehen
- sollte deshalb das direkte Referenzieren auf die *Objects of Interest* innerhalb der Nutzungs- und Gestaltungsdiskurse unterstützt werden

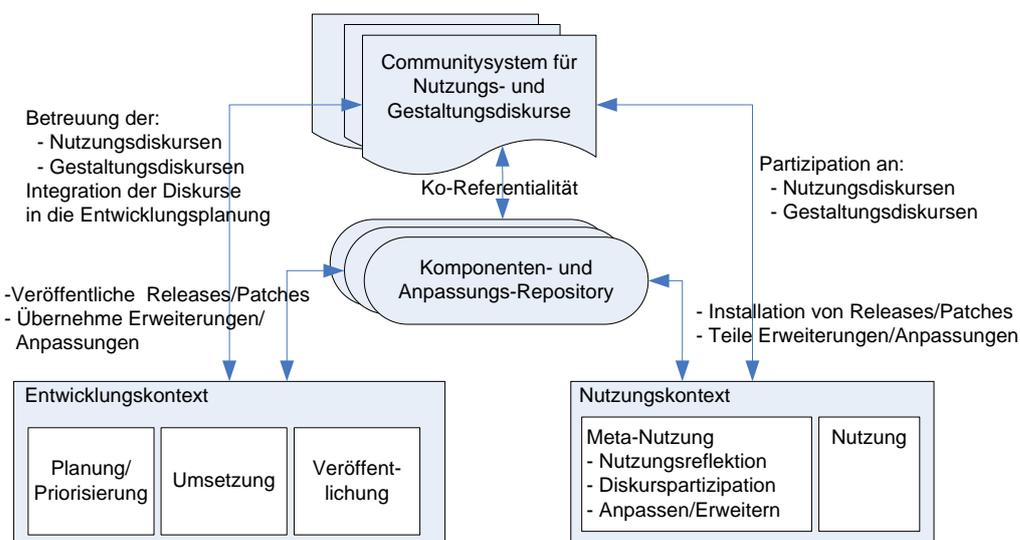


Abbildung 1: Schaubild einer geteilten, technischen Infrastruktur, die zwischen Nutzungs- und Entwicklungskontext vermittelt.

Abbildung 1 zeigt das Schaubild einer von Benutzern und Entwicklern geteilten Infrastruktur, die die verschiedenen Aspekte der Aneignungsunterstützung integriert und den Akteuren, eine an ihre Bedürfnisse angepasste Sicht auf die Meta-Nutzung erlaubt. Dabei sollten die Komponenten samt deren Anpassungen und die darauf sich beziehenden Diskurse einander ko-

referenziell verweisen [Wul+01], umso den Übergang zwischen Nutzungskritik und Nutzungsanpassung zu erleichtern.

Weite Teile dieser Architektur konnten im BSCWeasel-Projekt auf Basis von Eclipse umgesetzt werden. Die Realisierung des *Component Repository* wurde dabei von Eclipse mitgeliefert. Die *Discourse Infrastructure* wurde durch die im Folgenden beschriebenen Module *CHiC* und *PaDU* realisiert. Der Übergang zwischen Nutzungskritik und Nutzungsanpassung ist jedoch noch eine Herausforderung für zukünftige Arbeiten.

4 Umsetzung einer Aneignungsunterstützung innerhalb des BSCWeasel Projekts

Um unser Konzept zu erforschen, haben wir ein Groupware-System namens BSCWeasel entwickelt, das die hier vorgeschlagene Infrastruktur zur Aneignungsunterstützung als integralen Bestandteil besitzt. BSCWeasel ist ein Rich Groupware Client für das BSCW System, das an der Gesellschaft für Mathematik und Datenverarbeitung (GMD – heute integriert in die Fraunhofer-Gesellschaft) Mitte der 1990er Jahre entwickelt wurde [Ben+95].

BSCW ist eine erfolgreiche, vollständig browserbasierte Groupware-Anwendung. Die zugrunde liegende Architektur hat bestimmte Vorteile für die Benutzer, so z.B. bedarf sie keines zusätzlichen Installationsaufwands. Jedoch hat die *Thin Client*-Lösung auch verschiedene Nachteile gegenüber einer modular aufgebauten *Rich Client*-Lösung, insbesondere was die eigenständige Erweiterbarkeit durch den Nutzer anbetrifft [Ste+04].

Das BSCWeasel – ein Open Source Projekt², das im Frühjahr 2004 begonnen wurde –ergänzt deshalb das BSCW um solch einen modular aufgebauten *Rich Client* auf Eclipse-Basis (siehe Abbildung 2). In der ersten Version des BSCWeasel wurden dabei einige Hauptfunktionen des BSCW umgesetzt, die über die XML RPC API des BSCW Systems angesprochen wurden. Die erste Realisierung konnte dabei fast vollständig durch die von der *Eclipse Rich Client*-Plattform (RCP) bereitgestellten Komponenten und einige weitere schon vorhandene Komponenten realisiert werden. Später wurde der *Rich Client* durch das Hinzufügen und das Adaptieren bestehender Komponenten um neue Funktionalität erweitert (z.B. ein Chat auf XMPP/Jabber Basis).

² Vgl. www.bscweasel.de

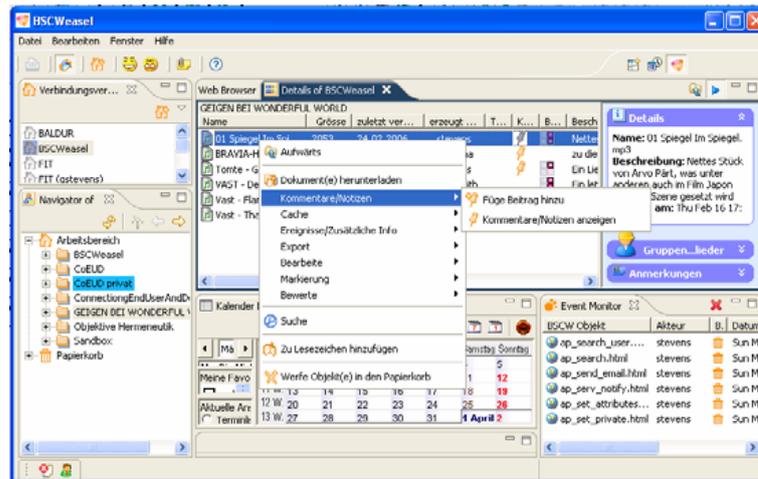


Abbildung 2: Snapshot des BSCWeasel Client

In das BSCWeasel wurden zusätzlich zwei (aktuell noch getrennte) Teilsystemen – ChiC und PaDU - integriert, die den Aneignungs- und Gestaltungsdiskurs zwischen den Benutzern bzw. Benutzern und Entwicklern verbessern sollen.

4.1 CHiC - Community Help In Context

Das eine Teilsystem ist die *Community Help In Context* (ChiC) [Ste+06a]. Die Kernidee von ChiC besteht dabei darin, das Konzept kontext-sensitiver Hilfe mit dem Wiki-Konzept zu verbinden. Das integrierte Hilfesystem wird so von einem unidirektionalen Medium zwischen Hersteller und Benutzer zu einem Community-orientierten Medium. Er erlaubt den Benutzern zu jeglicher Funktion des Anwendungssystems einen Beitrag zu erzeugen, zu ändern, zu kommentieren oder zu Bewerten und mit anderen Benutzern auszutauschen. Das entwickelte System macht sich dabei zu Nutze, dass heutige Anwendungssysteme nach dem *Direct Manipulation* Prinzip [Schne83] aufgebaut sind. Durch die algorithmische Inspektion der Widgetstruktur der laufenden Anwendung erlaubt das ChiC-System, die Nutzungsdiskurse an die betreffenden *Objects of Interest*, den Widgets, zu heften. Das hierbei angewendete Verfahren, das auf Prinzipien der Metaprogrammierung basiert, hat darüber hinaus den Vorteil, dass es auch in beliebigen Anwendungssystemen auf Eclipse-RCP-Basis eingesetzt werden kann. Die technischen Details von ChiC sind in [Ste+06a] genauer dargelegt.

4.2 PaDU – Participatory Design in Use

Um die reziproke Partizipation zwischen Benutzern und Entwicklern zu fördern, haben wir ein professionelles *Issue-Tracking*-System in die BSCWeasel-Applikation integriert, in dem

Benutzer und Entwickler in gleicher Weise partizipieren können [Ste+06b]. Um den Einstieg für den Endbenutzer zu vereinfachen, haben wir das System dabei mit einer speziell für Endbenutzer angepassten Benutzerschnittstelle ausgestattet. Dies erlaubt Benutzern des BSCWeasel-Systems nun einen speziellen Zugang aus ihrem unmittelbaren Nutzungskontext heraus.

Die Nutzerbeteiligung wird durch ein Formular zur Beschreibung des Nutzungsproblems unterstützt: Das Formular ist dabei eine Abwandlung des ursprünglichen *Critical Incident-Dialog* (Castillo), so wie er in der Studie von Hartson, Castillo et al. benutzt wurde [Har+96]. Im Gegensatz zu Hartson, Castillo et al. erlaubt PaDU auch das graphische Annotieren des Nutzungskontexts. Hierzu wird vom aktuellen Zustand automatisch ein Screenshot angelegt, in den Dialog eingefügt und mit einem speziellen Annotationswerkzeug verknüpft. Dies erlaubt dem Benutzer, nicht nur textuell, sondern auch graphisch auf die *Objects of Interest* zu verweisen.

Da Systementwicklung ein im hohen Maße kommunikativer Aushandlungsprozess ist, war es bei der Gestaltung von PaDU wichtig, die Kommunikationsprozesse für alle Beteiligten transparent zu machen. Deshalb wird z.B. der Benutzer nach dem Senden eines Feedback-Reports direkt an die Stelle in *Issue-Tracking-System* geführt, wo sein Beitrag samt aller Ergänzungen Dritter verwaltet wird. Dieses Feature erlaubt es dem Benutzer, den Fortgang seines Beitrags weiterzuverfolgen, die Kommentare von anderen Benutzern/Entwicklern zu lesen bzw. selber neue Kommentare zu verfassen.

Die technischen Details und die Einordnung von PaDU in die Landschaft anderer Systeme zur Beteiligung von Anwendern während der Nutzung ist in [Ste+06b] dargelegt.

5 Diskussion der Aneignungsunterstützung im BSCWeasel Projekt

Das BSCWeasel wurde an der Universität Siegen entwickelt. Die Kerngruppe bestand aus zwei Entwicklern, die zu verschiedenen Zeitpunkten durch Studententeams unterstützt wurden. Eine erste anfängliche Version des BSCWeasel wurde ab Mai 2005 von den Entwicklern und den Studententeams benutzt. Spätere Versionen wurden innerhalb der Forschungsgruppe (ca. 15 Mitglieder) und einer befreundeten Forschungsgruppe (ca. 15 Forscher) ca. 100 km von Siegen entfernt veröffentlicht.

Das BSCWeasel-Projekt folgt einen Aktionsforschungsansatz. Hieraus ergibt sich, dass primär formative Evaluationen betrieben werden, um die Ergebnisse in die Verbesserung der Gestaltung einfließen zu lassen. Sie beruht auf der Untersuchung von Paper-Mockups, semi-strukturierten und informellen Interviews, gesammelten Feldnotizen und BSCWeasel bezogener Email-Korrespondenz. Zusätzlich wurden zwei Usabilitystudien nach dem DaTech-Verfahren durchgeführt. Die erste Studie wurde ohne die Aneignungsinfrastruktur im April 2005 mit 9 Nutzern durchgeführt. Die zweite wurde im Januar 2006 mit der Aneignungsinfrastruktur mit 6 Nutzern durchgeführt.

Hinsichtlich des Nutzungsgrads des BSCWeasel an der Universität Siegen und dem Forschungsinstitut haben wir 11 regelmäßige Nutzer gezählt. Alle diese Nutzer waren zuvor auch intensive BSCW Nutzer und haben im BSCWeasel spezielle Funktionen entdeckt, die ihnen die alltägliche Arbeit vereinfachen.

Von September 2005 bis 2006 wurden 74 Gestaltungsvorschläge mittels der Aneignungsinfrastruktur gemeldet. Die meisten der dort adressierten Problem verweisen auf Nutzungsprobleme, die zwar die Nutzer irgendwie gemeistert haben, bei denen jedoch eine Erweiterung bzw. ein Redesign der Groupware das Nutzerverhalten ändern oder die Effizienz steigern könnte.

Basierend auf unseren Studien innerhalb des BSCWeasel-Projekts konnten wir drei Dimensionen identifizieren, die bei der Gestaltung einer Infrastruktur zur Aneignungsunterstützung berücksichtigt werden sollte. Wir wollen die einzelnen Dimensionen im Einzelnen darlegen. Die Ausarbeitung konkreter Gestaltungsrichtlinien ist jedoch noch eine zu leistende Aufgabe.

5.1 Die epistemologische Dimension: Krisensituationen und Erforschungsprozesse

Die heutige Forschung betrachtet die verschiedenen Aspekte von Krisensituation³ in der Nutzung und damit einhergehende Reflektionsprozesse meist noch getrennt. So wird die Konsultierung des Hilfesystems, das Anpassung des Systems oder die Partizipation am Gestaltungsdiskurs in Open-Source-Projekten getrennt voneinander studiert. In unseren Interviews ist uns jedoch aufgefallen, dass sich diese Trennung aus Nutzersicht nicht so scharf darstellt. Vielmehr fanden es die Nutzer in einer Krisensituation häufig schwierig, die richtige Unterstützungsfunktionalität auszuwählen.

³ Im Folgenden soll hier der englische Begriff der *breakdown situation* mit Krisensituation übersetzt werden.

Einen Hinweis auf das zugrunde liegende Problem fanden wir in einen Kommentar eines Nutzers. Er kommentierte sein Erweiterungsvorschlag mit den Hinweis: „*Funktion nicht vorhanden oder nicht gefunden, d.h. ich bin ein DAU ;-)*“. Im ersten Abschnitt des Satzes wird das fehlende Feature als ein Mangel des Systems konstatiert, was einen Kontakt zu den Entwicklern legitim erscheinen lässt. Im zweiten Abschnitt wird das Problem jedoch auf mangelndes eigenes Wissen über das System zurückgeführt, was die Konsultation eines Hilfesystems legitimieren würde. Der springende Punkt hier ist aber, dass aus der Perspektive des Nutzers in der Krisensituation nicht entschieden werden kann, welcher Fall hier tatsächlich vorliegt.

Um die Struktur von Krisensituation und deren Umgang durch den Nutzer besser verstehen zu können, haben wir deren Manifestationen in den ‚Thinking-Aloud‘-Protokollen, die während der Usability Studien angefertigt wurden, eingehender studiert. Dabei haben wir auf die von Ullrich Oevermann entwickelte Methode der Sequenzanalyse zurückgegriffen [Oeve97]. Aus Platzgründen kann an dieser Stelle die Ergebnisse nur komprimiert anhand einer Passage illustriert werden.

„*Na, komm!* [das routinisierte Handeln gerät in eine Krise] - *Ja, jetzt wäre die Frage,* [der Nutzer fängt an über die aktuelle Situation zu reflektieren] *jetzt bin ich hier im Navigator und wenn ich den Inhalt jetzt da drüben sehen wollte, ne* [im Sinne von: „nicht wahr?“], *ganz einfach, also dann, vielleicht ... weiß ich nicht, vielleicht kann man das da reinschieben.* [er entwickelt eine Hypothese über die Problemsituation] *Da geht irgendwas nicht. (T: zustimmend Hmm)* [experimentiert herum]. *Nee, so! O.k., prima, kriege ich jetzt nicht hin. Ich hatte jetzt nur gedacht, vielleicht könnte man - sozusagen wie im Windows Explorer ...* [er kommt zu den Schluss, dass seine Annahme nicht stimmt und fängt an sie durch eine neue zu ersetzen] [er probiert verschiedene Hypothesen aus] *so wie ich mir das vorstelle, scheint das nicht zu gehen* [kommt zu einer Überzeugung] *aber das ist egal, macht nix.*[beschließt die Sache nicht weiter zu erforschen]“ (Passage aus einem Thinking-Aloud Protokoll).

Wir können unsere Interpretation einer Krisensituation in der Softwarenutzung wie folgt zusammenfassen: Das Hauptmerkmal einer Krisensituation ist der Zweifel, der in aktuellen Situation gründet⁴. In seiner allgemeinen Form kann er durch den höchst indexikalischen Ausdruck charakterisiert werden: „*Hier und jetzt ist irgendwas ungewöhnlich*“. Dies löst einen Forschungsprozess aus, in welchem Hypothesen entwickelt werden, die auf der einen Seite über

⁴ Anhand dieses Beispiel wird auch klar, dass mit Krisensituation keine existentielle Krise gemeint ist. Vielmehr bedeutet eine Krisensituation zunächst einmal nur, dass die üblichen Handlungsroutinen nicht greifen.

die Nutzungsroutine hinausgehen müssen (da diese ja gerade in die Krise geraten ist), auf der anderen Seite an diese zurückgebunden bleiben (da nur von den bestehenden Überzeugungen die neue Situation interpretiert werden kann). Dieses Spannungsverhältnis kann nur aufgelöst werden, wenn man die entwickelten Hypothesen irgendwie an der Realität prüft. Dabei ergibt sich ein offener zyklischer Prozess von Reflektion und Aktion, der strukturell viele Ähnlichkeiten zu dem aufweist, was der Pragmatist H. Joas mit "Kreativen Handeln" bezeichnet hat [Joas96]. Insbesondere kann zu Beginn dieses Prozesses nicht entschieden werden, welche Handlungsstrategie letztendlich zielführend ist. Entsprechend kann es vorkommen, dass der Akteur innerhalb dieses Prozesses verschiedene Handlungsstrategien ausprobiert.

Auf dieser Ebene kann das Artefakt *epistemisch* die Funktion eines *Boundary Objects* übernehmen, wenn im Rahmen des Umgangs mit dem Artefakt gemachte Erfahrungen als Bindeglied zwischen seiner symbolischen Interpretation und der in die Krise geratenen Handlungsroutine fungieren kann. Als Folge des Forschungsprozesses kann dabei sowohl die symbolische Interpretation als auch die Handlungsroutine transformieren werden.

5.2 Die mediale Dimension: Visuelle Annotation und die Referenzierung des Nutzungskontext

Wie oben beschrieben gründet eine Krisensituation in einer momentanen Situation und verweist höchst indexikalisch auf die aktuellen *Objects of Interests*. Dies hat auch Konsequenzen für die mediale Dimension einer Aneignungsunterstützung. So zeigte sich in der Evaluation der Beiträge, dass sich die Nutzer in ihren Beiträgen implizit oder explizit auf das Softwareartefakt und die aktuelle Nutzungssituation beziehen. Dabei wird häufig auf die aktuelle Benutzerschnittstelle Bezug genommen, in der die Krisensituation auftrat. Des Weiteren haben wir festgestellt, dass die Funktionalität, einen Screenshot vom aktuellen Zustand der Applikation zu erstellen, ein sehr hilfreiches Werkzeug dafür ist, ein ‚Etwas‘ des ‚Hier und Jetzt‘ einer aktuellen Krisensituation einzufangen. Des Weiteren wird das eingebaute graphische Annotierungswerkzeug nicht nur dazu verwendet, auf die Problemsituation zu verweisen, sondern auch dazu, ein alternatives Design zu illustrieren.

Auf dieser Ebene kann das Artefakt *medial* die Funktion eines *Boundary Objects* einnehmen, indem es eine gemeinsame Interpretationsbasis herstellt, mittels derer sich die indexikalischen Bezüge innerhalb Kommunikation zwischen den verschiedenen Akteuren verorten lassen.

5.3 Die politische Dimension: Ich, meine Community und der Hersteller

Die Entwicklung und Nutzung von Technologien ist eingebettet und beeinflusst von den vorhandenen Machtstrukturen. Als Folge davon besitzt die Aneignung und kreative (Um-)Nutzung auch eine politische Dimension. Dies ist zwar ein allgemein bekanntes Faktum, trotzdem wird dieser Aspekt selten im Kontext von User Innovation diskutiert.

In unseren Evaluationen haben wir verschiedene Indikatoren für die Wichtigkeit der politischen Dimension und den Aspekt des Vertrauens gefunden. Dabei haben wir aus Nutzersicht drei verschiedene Akteure identifiziert, die eine kritische Rolle im Aneignungsprozess spielen: die Gruppe von Freunden und Kollegen, die weltweite Nutzercommunity und den Hersteller. Dabei wird die Rolle des Herstellers von den Nutzern unterschiedlich beurteilt. Auf der einen Seite vertrauen einige Nutzer den Informationen der Hersteller, da sie von offizieller Seite geprüft worden sind. Jedoch haben wir auch die gegenteilige Meinung gefunden. Exemplarisch wird diese Haltung in den Kommentar eines Nutzers ausgedrückt: *„Ich würde den Hilfetexten von anderen Nutzern mehr vertrauen, insbesondere weil der Text von jemand geschrieben wurde, der sich in einen ähnlichen Nutzungskontext befindet.“* Gegenüber großen Herstellern fehlt auch das Vertrauen, dass ihr Feedback Ernst genommen wird und dass dieser etwas bewirken kann.

Ein anderer wichtiger Punkt bei den Nutzern ist die Angst, dass ihre Beiträge von irgendjemand falsch interpretiert werden. Hierbei kann eine zu große Transparenz auch problematisch werden. Ein Ausschnitt aus einer Email Korrespondenz soll diesen Aspekt verdeutlichen. Hierbei haben wir eine Benutzeranfrage per Email bekommen: *“Ein ‚Killer-Feature‘ wäre m.M.n. sicher weiter eine einfache Backup / Transfer-Möglichkeit von Server zu Server.“* Auf Grund unseres Entwicklungsansatzes haben wir die Anfrage auf das öffentliche Design-Diskussionsforum gestellt und den Nutzer per Email darüber informiert: *„Ich habe Deine Anregung auch mal unter [URL] aufgenommen.“* Da die Anfrage jedoch eine politische Implikationen hat, hat uns zweite Email erreicht, in der der Nutzer folgende äußerte *“Eine Bitte: Kannst Du meinen Request nochmal löschen und ich erstelle ihn neu, dann kann ich es etwas präziser formulieren und wir vermeiden ggf. Konflikte mit [Hersteller], die den ‚Server-to-Server‘-Transfer meines Wissens als kommerzielle Dienstleistung anbieten.“* Dieses Beispiel zeigt auf, dass die Kommunikation und einhergehend damit auch die Entwicklung von Gestaltungsideen von dem jeweiligen politischen Kontext abhängen.

Insbesondere kann auf dieser Ebene das Artefakt *politisch* die Funktion eines *Boundary Object* einnehmen, auf die sich die verschiedenen Gestaltungs- und Nutzungsinteressen beziehen und

miteinander ausgehandelt werden können. Der Erfolg eines Artefakts ergibt sich dabei aus seiner Fähigkeit, die unterschiedlichen Interessen der einzelnen Akteure im ausreichenden Maße zu integrieren.

6 Zusammenfassung

Der Flexibilisierung von IT-Landschaften kommt eine immer größere Bedeutung zu. In der Forschung werden jedoch meist produkt-orientierte und prozess-orientierte Flexibilisierungsstrategien noch getrennt von einander betrachtet.

Synthetisiert man jedoch beide Richtungen, erkennt man, dass ein integrierendes Glied fehlt, das in der praktischen Umsetzung zwischen beiden Konzepten vermittelt. Bei der Vermittlung kommt dem Endbenutzer eine entscheidende Rolle zu, da er zum einen mit den (geänderten) Anforderungen im Nutzungskontext unmittelbar konfrontiert und zum andern in der Lage ist, auch eigenständig Innovation zu entwickeln.

Das Konzept der Infrastrukturen zur Aneignungsunterstützung erweitert die Ideen der *User Innovation* [vHip88, vHi+02] um das Konzept anpassbarer Software [Tege00, Hen+91], wobei es sich die Eigenschaft der *Direct Manipulation* [Shne83] heutiger Anwendungssysteme zunutze macht, um Meta-Nutzungsfunktionalitäten direkt in das User Interface einzuweben.

Auf der analytischen Ebene konnte das Konzept der *Boundary Objects* [Sta+89] fruchtbar gemacht werden um die einzelnen Dimensionen einer Aneignungsunterstützung zu thematisieren. Dabei zeigt sich, dass die materielle Struktur des Artefakts nicht nur in der Nutzung, sondern auch im Diskurs hierüber eine nicht zu vernachlässigende Rolle spielt. Insbesondere zeigt die genaue Analyse von Krisensituation, dass die Trennung der verschiedenen Aspekte der Nutzungsreflektion aus der Perspektive des Nutzers problematisch ist.

Im BSCWeasel-Projekt konnte gezeigt werden, dass sich weite Teile des Konzepts der Infrastrukturen zur Aneignungsunterstützung mit Hilfe der durch Eclipse bereitgestellten Möglichkeiten auf eine effiziente Art und Weise umsetzen lassen [Ste+04, Ste+06a, Ste+06b].

Erste praktische Erprobungen der Infrastruktur beim BSCWeasel-Projekt deuten auch an, dass die hier vorgestellte technische Infrastruktur schon heute dazu beitragen kann, das Potential der Nutzer in einem kontinuierlichen Entwicklungsprozess besser zu nutzen. Jedoch hat sich auch bei der technischen Umsetzung gezeigt, dass die Architektur von Eclipse nicht darauf anlegt ist,

den Übergang von Nutzung zur Meta-Nutzung zu fördern. Inwieweit dies nachträglich durch geeignete Erweiterung behoben werden kann, ist Teil der zukünftigen Arbeiten der vorgestellten, prototypischen Umsetzung der Infrastruktur.

References

- [Beck00] Beck, K.: Extreme Programming Explained: Embrace Change. 2000, Pearson Education: Addison-Wesley.
- [Ben+95] Bentley, R., Dourish, P.: Medium versus mechanism: Supporting collaboration through customisation. In: Proceedings of ECSCW'95, 1995. Sweden: Kluwer.
- [Bryn98] Brynjolfsson, E., Hitt, L.M.: Beyond the productivity paradox. Communications of the ACM, 1998. 41(8): S. 49-55.
- [Fis+02] Fischer, G., Ostwald, J.: Seeding, Evolutionary Growth, and Reseeding. In: Proc. of PDC 02, CPSR, 2002, S. 292-298.
- [Fis+90] Fischer, G., Girgensohn, A.: End-user modifiability in design environments. In: Proceedings of the SIGCHI. 1990, ACM Press, 1990, S. 183 - 191.
- [Flo+89] Floyd, C., Reisin, F.-M., Schmidt, G.: STEPS to Software Development with Users Source. In: Proc. of the European SE Conf., Springer-Verlag, 1989, S. 48-64.
- [Har+96] Hartson, H.R., Castillo, J.C., Kelso, J., Kamler, J., Neale, W.: Remote Evaluation: The Network as an Extension of the Usability Laboratory. In Proc. of CHI'96, 1996, S. 228-235
- [Hen+91] Henderson, A., Kyng, M.: There's No Place Like Home: Continuing Design. In: Use, in Design at Work, 1991, S. 219 - 240.
- [Joas96] Joas, H., Die Kreativität des Handelns, 1996, Suhrkamp.
- [Kar+98] Karsten, H., Jones, M.. The long and winding road: Collaborative IT and organisational change. In: Int. Conference on CSCW'98. ACM Press, 1998, S. 28-38.

- [Mal+95] Malone, T.W., Lai, K.-Y., Fry, C.: Experiments with Oval: a radically tailorable tool for cooperative work. In: ACM TOIS. 13 (2), 1995, S. 177 - 205.
- [Mørc97] Mørch, A.: Three Levels of End-user Tailoring: Customization, Integration and Extension. In: Computers and Design in context, MIT Press, 1997, S. 51 – 76.
- [Ober94] Oberquelle, H., Situationsbedingte und benutzerorientierte Anpassbarkeit von Groupware. In: Menschengerechte Groupware, Teubner, 1998, S. 31-50.
- [Orl+97] Orlikowski, W.J., Hofman, J.D.: An Improvisational Model for Change Management.. In: Sloan Management Review, 1997, S. 11-21.
- [Orli96] Orlikowski, W.J.: Evolving with Notes: Organizational change around groupware technology. In Groupware & Teamwork, J. Wiley, 1996, S. 23 – 60.
- [Oeve97] Oevermann, U.: Thesen zur Methodik der werkimmanenten Interpretation vom Standpunkt der objektiven Hermeneutik, Manuskript, gehalten auf der 4. Arbeitstagung der Arbeitsgemeinschaft objektive Hermeneutik e.V. 1997: Frankfurt a. M
- [Pipe05] Pipek, V.: From Tailoring to Appropriation Support: Negotiating Groupware Usage. PhD Thesis, University of Oulu: Oulu, Finland, 2005.
- [Pip+99] Pipek, V., Wulf, V.: A Groupware's Life. In: Proc. of ECSCW '99, Kluwer, 1999, S. 199 – 219.
- [Rum+01] Rumpe, B., Schröder, A.: Quantitative Untersuchung des Extreme Programming Prozesses, Technischer Bericht TUM-I0110 and ViSEK/006/D2001, 2001.
- [Shne83] Shneiderman, B.: Direct manipulation: a step beyond programming languages. In: *IEEE Computer* 16(8), 1983, S. 57-69.
- [Sta+89] Star, S.L., Griesemer, J.R.: Institutional Ecology, Translations and Boundary Objects. In: *Social Studies of Science*, 19, 1989, S. 387-420.
- [Ste+04] Stevens, G., Budweg, S, Pipek, V.: The "BSCWeasel" and Eclipse-powered Cooperative End User Development. In: Workshop at the CSCW'04, 2004.

- [Ste+06a] Stevens, G., Wiedenhöfer, T.: CHIC - A pluggable solution for community help in context. In: Proc. of the NordCHI 2006 (in Druck).
- [Ste+06b] Stevens, G. Draxler, S.: Partizipation im Nutzungskontext. In: Konferenzband Mensch & Computer 2006 (in Druck).
- [Stiem00] Stiernerling, O.: Component-Based Tailorability. Dissertation, Institut für Informatik III., Rheinische Friedrich-Wilhelms-Universität Bonn, 2000.
- [Strü99] Strübing, J.: Digging the Gold Mine. Vortragsmanuskript, gehalten an der Spring School on STS, Zurich, March, 3, 1999.
- [Teeg00] Teege, G.: Users as Composers: Parts and Features as a Basis for Tailorability. In: Computer Supported Cooperative Work, 9(1), 2000, S. 101-122.
- [Tör+03] Törpel, B., Pipek, V., Rittenbruch, M.: Creating Heterogeneity - Evolving Use of Groupware in a Network of Freelancers. In: Computer Supported Cooperative Work, 12(1-2), 2003.
- [vHip+02] von Hippel, E., Katz, R.: Shifting Innovation to Users via Toolkits, In: Management Science, 48(7), 2002, S. 821-834.
- [vHip88] von Hippel, E.: The sources of innovation, Oxford University Press, 1988.
- [Wul+01] Wulf, V., Golombek, B.: Exploration environments: concept and empirical evaluation. In: Proc. of the GROUP. 2001, S. 107-116.
- [Wul+95] Wulf, V., Rohde, M.: Towards an Integrated Organization and Technology Development. In: Proc. of Designing Interactive Systems. 1995, S. 55-64.
- [Wulf99] Wulf, V.: Evolving Cooperation when Introducing Groupware – A Self-Organization Perspective. In: Cybernetics and Human Knowing, 6(2), 1999, S. 55 – 75.

INDIVIDUAL AND HUMAN-ASSISTED COMPUTER SELF-EFFICACY: AN EMPIRICAL EXAMINATION

Jason Bennett Thatcher

College of Business and Behavioral Science - Clemson University
Clemson, SC 29634 USA
jthatch@clemson.edu

Michael J. Gundlach

College of Business - Florida State University
Tallahassee, FL 32306-1110 USA
mjg1052@garnet.acns.fsu.edu

D. Harrison McKnight

Eli Broad College of Business - Michigan State University
East Lansing, MI 48824-1121 USA
mcknight@bus.msu.edu

Mark Srite

Sheldon B. Lubar School of Business - University of Wisconsin-Milwaukee
Milwaukee, WI 53201 USA
msrite@uwm.edu

Abstract

Researchers have found computer self-efficacy to be important to technology adoption. Past research has treated computer self-efficacy (CSE) as a unitary concept. This study proposes that CSE has two dimensions—individual and human-assisted. Using items drawn from the Compeau and Higgins' [CoHi95b] CSE instrument, the paper examines each dimension's relationship to computer anxiety and perceived ease of use of information technology. The paper contributes to the literature by showing that the Compeau and Higgins instrument measures two CSE dimensions with distinct effects on computer anxiety and perceived ease of use of information technology. Implications for research and practice are offered.

1 Introduction

Successfully adopting technologies is crucial for effective enterprises, and computer self-efficacy is a key factor in adoption. Learning to utilize information technology (IT) rests on the shoulders of individual organizational members. At this individual level of analysis, understanding factors influencing IT use has been of great interest to information systems researchers [CoHH99, Davi89, MaHZ96]. Numerous studies have demonstrated that computer self-efficacy (CSE) plays an important role in influencing IT perceptions and use [HiSM87, Igl95, MaYJ98]. CSE refers to an individual's "judgment of one's capability to use a computer" [CoHi95b, p. 192]. In general, research suggests that CSE is an important antecedent to beliefs and emotions that influence IT use. For example, CSE positively influences perceptions about the ease of use (EOU) of IT [VeDa96] and negatively impacts computer anxiety [CoHH99].

Although CSE has been frequently researched, critics of the CSE literature suggest that, "the results obtained to date have, in some cases, been either equivocal or contradictory" [MaYJ98, p. 126]. For example, using an established CSE scale, Compeau and Higgins [CoHi95a] found support for a model linking CSE to outcome expectancy, which refers to a person's estimate that a behavior will lead to an outcome [Band77]. However, using a new CSE scale, Johnson and Marakas [JoMa00] found stronger support for CSE exerting a positive influence on outcome expectancy. After examining these contradictions, we feel that the problematic findings are likely rooted in how studies have operationalized the CSE construct [Band77]. Some have employed self-developed measures [Busc95, HeSt95, JoMa00], while others have modified existing measures [TaTo95, WeMa93], creating a lack of consistency in the way CSE has been operationalized. Furthermore, CSE has been treated as a unitary concept, even though, as discussed later in the paper, its measures reflect two distinct concepts. Treating multi-dimensional concepts as unitary constructs can cause empirical problems, as found in other domains [Rubi73]. A lack of conceptual clarity and measurement issues may be responsible for fragmented findings affecting the development of the CSE concept [MaYJ98]. This paper examines two questions: 1) What are the conceptual dimensions of CSE?; and 2) Can the Compeau and Higgins' [CoHi95b] CSE scale be adapted to operationalize these conceptual dimensions?

2 Theoretical Background

Drawing on social cognitive theory [Band77], CSE research brings *self-efficacy* to the domain of IT. Self-efficacy refers to “people’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performances” [Band86]. Self-efficacy’s implications have been studied extensively in many fields such as education, psychology, social psychology, health, athletics, and management [Band97].

Computer self-efficacy refers to individuals’ judgment of their capabilities to use computers [MaYJ98]. Research suggests that individuals who express higher CSE levels are likely to express more positive beliefs about, and more frequently use, IT [CoHH99]. CSE may be conceptualized at two levels – general and task-specific [MaYJ98]. *General CSE* refers to individuals’ beliefs about their ability to use a computer across situations or applications. *Task-specific CSE* refers to individuals’ beliefs about their ability to use a specific information technology. *Task-specific CSE* may be influenced by training or experience [AgSS00]. General and task-specific CSE do not differ from each other in terms of their conceptual makeup, only in terms of level of specificity. Rather, they are the same basic construct applied to either general or specific technologies. A number of recent studies have examined the construct of computer self-efficacy. Unfortunately a consensus does not seem to have emerged as to how to operationalize CSE. In other research, Compeau and Higgins’ ten-item scale has been used in its entirety [GuNd06, Haya04, ShPC03] as well as reduced to a smaller subset [ThCH06, Hasa06, ChLu04]. A number of other scales have also been employed [Hasa06, ReDM05, StHe03, YiIm04, FaNW03].

2.1 Attribution Theory and CSE

Individuals may attribute computer self-efficacy beliefs to internal and external sources [Band88, KeCH99, MaYJ98, SiMG95]. Attribution theory helps explain individuals’ causal explanations for events or performance. It suggests that individuals’ beliefs about locus of causality influence perceptions about their performance in various situations. Locus of causality refers to whether individuals believe that their ability to perform an action (such as using an IT) rests on external factors or resides within themselves [Mart95, p. 9]. Thus, external attribution relates to human-assisted CSE. When individuals make external attributions, they assess their capability in light of external factors such as peer or technical support. In terms of CSE, external attributions reflect beliefs about one’s capability to perform a task on a computer *with*

assistance or support from an external source. When individuals make internal attributions, they believe that they can exert personal control over performance or other outcomes (i.e., without external help). In terms of CSE, internal attributions reflect individuals' beliefs about their ability to independently accomplish a task using a computer. Hence, internal attribution relates to individual (unassisted) CSE.

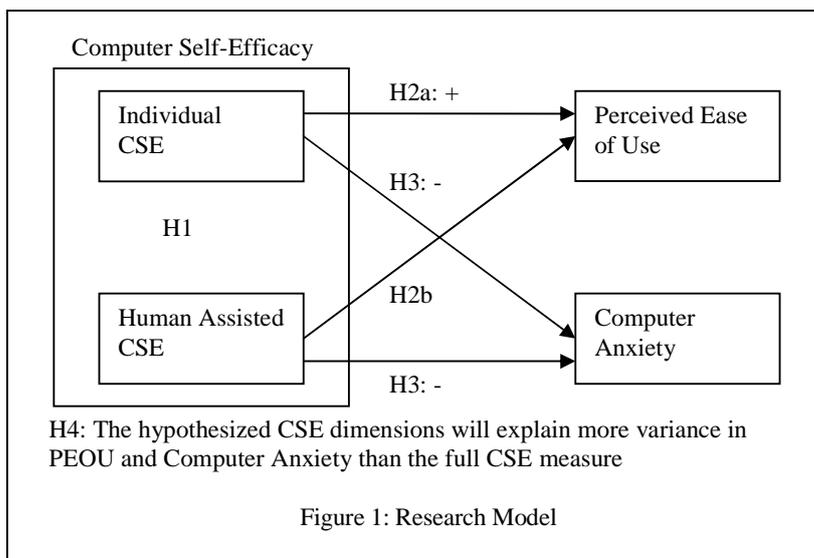
Applying attribution theory to CSE terms yields two construct definitions. *Individual CSE* means a belief about one's ability to independently accomplish a task on a computer. A little child demonstrates individual self-efficacy by proudly saying, "Look, Mom, I can do it all by myself". *Human-assisted CSE* means a belief about one's ability to perform a task on a computer with support from another person. This type of CSE reflects the need for help, as when, frustrated at the computer, you ask for help from colleagues or support personnel. Distinguishing between internal and external sources of efficacy beliefs is important because individuals who make internal attributions express more confidence in their ability to perform a task and report more positive beliefs and attitudes than do those who make external attributions [Band77, Schu84, ScGu86, SiMG95]. In this way, splitting CSE along attributional lines extends theory beyond the levels-of-specificity distinction of general and task-specific CSE. Distinguishing between individual and human-assisted CSE is important because it distinguishes the two underlying attributional thought structures that influence how CSE will operate. Because of their attributional differences, individual and human-assisted CSE should apply to both general and task-specific CSE situations.

Rather than develop a new CSE measure, this study focuses on refining an existing instrument to measure the internal and external dimensions of CSE. By doing so, we build on prior empirical research. The Compeau and Higgins' [CoHi95b] instrument was designed to capture the magnitude and strength of CSE across situations. The instrument has received varying support through its wide use in the information systems literature [AgKa00, AgSS00, CoHH99]. Although designed to assess individuals' overall CSE level, we believe that the Compeau and Higgins' items represent distinct attributional sources of CSE. By carefully evaluating the instrument's items in terms of dimensionality, and then validating these dimensions across general and specific CSE, we believe researchers will possess more effective tools to evaluate CSE across settings. Hence, the next section of the paper develops hypotheses that predict the multi-dimensionality of Compeau and Higgins' instrument, and evaluates the

nomological network of relationships between the dimensions of CSE and two effects of CSE-computer anxiety and perceived ease of use.

3 Research Model and Hypotheses

The research model was developed as follows (see Figure 1). First, the researchers examined the face validity of the Compeau and Higgins' CSE instrument to determine whether items fell within the internal and external dimensions of efficacy beliefs. Face validity is a method most often used to assess whether measurement items adequately represent the concept they are intended to measure, thereby assuring the important logical link between empirical measures and theoretical concepts, on which the veracity of logical positivist science depends [Schw80]. Face validity was assessed prior to empirical analysis in order to identify *a priori* which items



best measured each attributional CSE dimension.

Two of the researchers separately examined the wording and phrasing of the instrument in order to identify items that implied different sources of attributional

causation. The items were examined in light of: (a) what the researchers knew about attribution theory and (b) the definitions of individual and human-assisted CSE presented in the Theoretical Background section. The researchers consistently identified six items that uniquely represented the two distinct dimensions of CSE – the individual (I) and human-assisted (H). Their coding for items comprising each dimension is shown in Table 1. Overall, the researchers agreed on the placement of all but one of the items (Cohen's Kappa = .92).

Meeting to discuss their item codings, the two researchers agreed that the remaining four items did not have sufficient clarity to be clustered with either dimension or to form a distinct conceptual dimension; rather, these items represented such beliefs as familiarity with the

technology, resources such as time required to complete a task, or an automated help facility built into the system. Ignoring these items (coded O in Table 1), the scale appears to have two distinct dimensions representing individual and human-assisted CSE. Thus, an analysis of these measures in light of attribution theory suggests that:

Hypothesis 1. The Compeau and Higgins' [CoHi95b] computer self-efficacy instrument measures two distinct dimensions of efficacy beliefs.

Computer Self Efficacy Items	Rater 1	Rater 2
<i>I could complete my job using the technology if ...</i>		
1. ... there was no one around to tell me what to do. (I)	I	I
2. ... I had never used a package like it before. (I)	I	I
3. ... I had only the software manuals for reference. (I)	I	I
4. ... I had seen someone else using it before trying it myself.	H	O
5. ... I could call someone for help if I got stuck. (H)	H	H
6. ... someone else helped me get started. (H)	H	H
7. ... I had a lot of time to complete the job for which the software was provided.	O	O
8. ... I had just the built-in help facility for assistance	O	O
9. ... someone showed me how to do it first. (H)	H	H
10. ... I had used similar packages like this one before to do the job.	O	O
I = Items comprising the individual dimension of CSE. H = Items comprising the human-assisted dimension of CSE. O = Other sources of efficacy such as resource availability or experience. Cohen's Kappa, a measure of interrater reliability, was .92		

Table 1. Compeau and Higgins' (1995) General Self-Efficacy Measure

The Information Systems literature was then reviewed to identify perceptions of IT with which the individual and human-assisted dimensions would demonstrate distinct relationships. In this way, nomological validity of the constructs could be demonstrated. Nomological validity means whether or not the construct does what is expected within its nomological network [Schw80], such as relating to other constructs as theory suggests [WeMa93]. If two constructs perform differently in their nomological network, this provides additional evidence that they are discriminant constructs.

To establish the nomological validity of the CSE dimensions, we examined two well-established correlates of CSE – the perceived ease of use (EOU) of IT and computer anxiety. We selected EOU and computer anxiety as dependent variables for two reasons. First, CSE exhibits distinct relationships with EOU and computer anxiety. When individuals express high levels of CSE, research suggests that they report high levels of EOU [Venk00, VeDa96]. Inversely, when individuals express low levels of CSE, studies consistently show individuals report higher levels of computer anxiety [JoMa00, MaYJ98]. Second, EOU and computer anxiety were used as outcome variables in important CSE research. In an important extension

of the technology acceptance model, Venkatesh [Venk00] found that CSE was a significant positive correlate of EOU. Also, in their seminal CSE research, Compeau and Higgins [CoHH99, CoHi95a] found that computer anxiety and general computer self-efficacy were negative correlates. Because EOU and computer anxiety are well established constructs in CSE research, we believe they serve as useful outcome variables for research examining the nomological validity of dimensions of CSE. EOU is defined as “the degree to which the prospective user expects the target system to be free of effort” [DaBW89]. Venkatesh and Davis [VeDa96] found that general CSE was a significant positive correlate of EOU of IT. Since a belief that one can independently accomplish a computer task (individual CSE) will make a system seem simple to use, we believe that the individual CSE dimension will positively affect perceived EOU. In other words, when people have higher levels of individual CSE, they should report higher levels of EOU because they have confidence in their personal ability to use IT.

Hypothesis 2a. The individual dimension of computer self-efficacy will have a significant, positive influence on perceived ease of use.

Human-assisted CSE reflects beliefs about the ability to use IT with external assistance. Although individuals may believe they can perform a task with assistance, they may not believe that a technology will be easy for them to use. Beliefs linked to ability should improve as the sources of successful performances are attributed to internal causes grounded within one’s self (e.g., ability or effort) rather than external circumstances [Band77, Band88, GiSR89, HeST95, MaHZ96, SiMG95]. Because perceptions of EOU should be based primarily on beliefs in individuals’ capabilities without receiving potential human assistance, we believe that the human-assisted dimension of CSE should not influence the perceived ease of use of IT. EOU perceptions tend to involve ease of use by the unaided person, indicating perceptions about how easy it is for the individual to use the technology without help. The EOU items match this definition of the concept.

Hypothesis 2b. The human-assisted dimension of computer self-efficacy will not significantly affect perceived ease of use.

Computer anxiety reflects negative emotional arousal linked to actual or potential computer use [IgPH89]. Computer anxiety stems from irrational fears about the implications of computer use such as the loss of important data or fear of making other mistakes [Siev88]. Compeau et al. [CoHH99] found that CSE was a significant negative correlate of computer anxiety. Given that

computer anxiety reflects irrational fears about IT use, it is reasonable to expect that it will be influenced by individual beliefs about using IT alone or with human assistance. Either type of CSE should influence computer anxiety. A high level of individual CSE should lower anxiety about using the computer because it implies one has personal control over the computer. Human-assisted CSE should lower anxiety because one would feel more assured (and thus less anxious) knowing that help is available. As a result, we expect that the individual and human-assisted CSE dimensions will each demonstrate significant negative relationships with computer anxiety.

Hypothesis 3. Individual and human-assisted computer self-efficacy will demonstrate significant negative relationships with computer anxiety.

Even if hypothesized relationships to EOU and computer anxiety are demonstrated, to be a useful extension of CSE research, the distinction between individual and human-assisted CSE must offer more explanatory power than the original, full CSE instrument. We believe it will. A broad, dual construct will tap both the individual and human-assisted aspects of CSE. Two separate CSE measures should provide more explanatory power than one conflicted measure.

Hypothesis 4. The hypothesized dimensions of computer self-efficacy will explain more of the variance in perceived ease of use and computer anxiety than will the original measure of computer self-efficacy.

4 Method

Stinchcombe [Stin68] argued that two somewhat different tests of a theory provide stronger support for the theory than does one test. To determine the distinction between the two types of CSE, we propose that they must pass two tests. First, they must factor separately, and, second, they must predict at least one related consequent in distinct ways (see Hypotheses 2a, 2b). To add contextual variety, two studies were conducted to assess the dimensionality of the CSE measure and to test the nomological validity of the CSE dimensions. Study 1 focused on general CSE, while Study 2 examined specific CSE. We felt it possible that the model constructs would behave differently under these two conditions. If the proposed dimensions operate as hypothesized under both conditions, this would more strongly support the two-dimension CSE model because it is a more rigorous test.

Study 1 operationalized constructs in terms of general CSE, anxiety, and the perceived ease of use of computers. The sample consisted of undergraduate business students enrolled in a required introductory computing course in the College of Business at a large public university in the Southeastern United States. The male to female ratio was 58% to 42%. Surveys were distributed during the first week of the course. A list-wise deletion yielded 153 usable responses (72.8%). Study 2 operationalized the constructs in terms of task-specific self-efficacy, anxiety and the perceived ease of use associated with Oracle Developer 2000. The sample consisted of senior-level MIS undergraduate students enrolled in an upper division systems analysis and design course in their major at the same university in the Southeastern United States. The male to female ratio was 69% to 31%. One hundred seventy-five surveys were distributed to students in a course that required using Oracle Developer 2000. The initial survey captured task-specific efficacy beliefs. At the conclusion of the Oracle module six weeks later, a second survey was administered that measured anxiety evoked by, and the ease of use of, Oracle Developer 2000. A list-wise deletion yielded 149 usable responses (74.5%). Across both studies, respondents completed questionnaires during regularly scheduled class times. Power analysis suggested that the sample sizes were more than sufficient to detect medium effect sizes based on the number of predictors used in the study [Gree91].

Measures were drawn from the information systems literature. CSE was measured using the ten-item scale developed by Compeau and Higgins [CoHi95b]. Computer anxiety was measured using four items developed by Heinssen, Glass, and Knight [HeGK87]. Four items were used to evaluate ease of use [Davi89]. Consistent with prior CSE research, we controlled for age, gender, and years of computer experience [MaYJ98] in the tests of Hypotheses 2 and 3.

5 Data Analysis and Results

We used Partial Least Squares (PLS), a structural equation modeling (SEM) technique, to assess the dimensionality and nomological validity of the CSE dimensions. This involved predicting EOU and CA with CSE (Hypotheses 2-4). PLS was used for two reasons. It does not require normality and allows researchers to estimate models using ordinal data derived from scales [Wold82]. When determining sample size, Barclay et al. [BaHT95] suggest a “rule of thumb” of ten times the most complex construct’s number of indicators or the largest number of paths leading to a latent construct. With more than 140 cases per study, the datasets satisfy suggested

guidelines for using PLS [BaHT95]. The model was then evaluated in two steps – measurement (to evaluate the measures’ dimensionality) and structural (to examine nomological validity). When using PLS, internal composite reliability (ICR), the average variance extracted (AVE), and items’ loadings and cross-loadings are used to evaluate the convergent and discriminant validity of the latent constructs in the research model. Composite reliability is calculated by squaring the sum of loadings, then dividing it by the sum of squared loadings, plus the sum of the error terms. Interpreted like a Cronbach’s alpha, an internal composite reliability (ICR) of .70 is sufficient for research [FoLa81]. With the exception of the full CSE measure in Study 2 (ICR = .67), results presented in Tables 3a and 3b suggest that the measures displayed adequate convergent validity. The AVE measures the variance captured by the indicators relative to measurement error [FoLa81]. To demonstrate convergent validity, the AVE should be greater than .50 [BaHT95]. To evaluate discriminant validity, Fornell and Larcker [FoLa81] suggest that the square root of the AVE may be compared with the correlations among the latent variables. To be discriminant, the square root of a construct’s AVE should be greater than its correlation with any other construct. Across studies, the AVE square roots presented in Table 2 suggest that the constructs demonstrated adequate discriminant validity. In each study, it is interesting to note that the three-item measures of individual and human-assisted CSE yielded higher AVE’s than the 10-item general CSE measure. A second way to evaluate discriminant validity is to examine the factor loadings of each indicator [Chin88]. To be discriminant, each indicator should load higher on the construct of interest than on any other variable. Inspection of loadings and cross-loadings presented in Tables 3a and 3b suggests that the items load on the appropriate constructs. Inspection of the loadings and cross-loadings suggests that the individual and human-assisted CSE items were discriminant. Hence, card sorting and analysis of the measurement model support Hypothesis 1.

PLS results provided mixed empirical support for Hypotheses 2 through 4 (see Table 4). PLS structural model results may be interpreted like a regression analysis. Each R^2 indicates the amount of variance explained in the latent construct [BaHT95]. Path coefficients can be read like standardized betas resulting from ordinary least squares regression. To test the significance of path coefficients, a bootstrapping procedure was used to generate t-statistics [Chin98]. Hypothesis 2, which suggested that individual and human-assisted CSE should demonstrate distinct relationships to the ease of use of IT, was supported. However, analysis provided limited support for Hypotheses 3 and 4. Results will be discussed in more detail below.

Study 1- Full Self-Efficacy Model							Study 1 - Dimensions Self-Efficacy Model							
	CSE	CA	EOU	Age	Exp	Gen		HACSE	ICSE	CA	EOU	Age	Exp	Gen
CSE	0.75						HACSE	0.89						
CA	-0.41	0.89					ICSE	0.53	0.85					
EOU	0.38	-0.35	0.90				CA	-0.38	-0.36	0.89				
Age	-0.04	-0.01	-0.07	-			EOU	0.25	0.42	-0.35	0.90			
Exp.	0.27	-0.27	0.17	0.04	-		Age	-0.03	-0.08	-0.01	-0.07	-		
Gen.	0.14	-0.23	0.23	0.23	0.15	-	Exp.	0.28	0.21	-0.27	0.17	0.04	-	
							Gen.	0.05	0.18	-0.23	0.23	0.23	0.15	-
Study 2- Full Self-Efficacy Model							Study 2 - Dimensions Self-Efficacy Model							
	CSE	CA	EOU	Age	Exp	Gen		HACSE	ICSE	CA	EOU	Age	Exp	Gen
CSE	0.68						HACSE	0.86						
CA	-0.31	0.82					ICSE	0.38	0.80					
EOU	0.27	-0.35	0.90				CA	-0.20	-0.38	0.82				
Age	0.03	0.15	-0.02	-			EOU	0.11	0.35	-0.35	0.90			
Exp.	0.02	0.03	-0.03	-0.16	-		Age	0.08	-0.02	0.15	-0.02	-		
Gen.	-0.06	-0.07	-0.04	0.25	-0.14	-	Exp.	0.03	-0.05	-0.03	-0.03	-0.16	-	
							Gen.	-0.03	-0.06	-0.07	-0.04	0.25	-0.14	-

^a The diagonal is the square root of the average variance extracted. To be discriminant, the diagonal item should be greater than corresponding off-diagonal elements.

Key: CSE = Computer Self-Efficacy, CA = Computer Anxiety, EOU = Perceived Ease of Use, Exp. = Experience, Gen = Gender, HACSE = Human-Assisted CSE, ICSE = Individual CSE

Table 2 - Correlation of Latent Constructs^a

Study 1 - Full Self-Efficacy Model							Study 1 - Dimensions Self-Efficacy Model							
ITEMS	CSE	CA	EOU	Gen	Exp	Age	ITEMS	HACSE	ICSE	CA	EOU	Gen	Exp	Age
CSE1	0.69	-0.29	0.33	0.12	0.15	-0.10	CSE5	0.92	0.53	-0.34	0.23	0.05	0.23	-0.03
CSE2	0.71	-0.38	0.36	0.15	0.17	-0.08	CSE6	0.84	0.49	-0.30	0.24	0.06	0.29	-0.02
CSE3	0.75	-0.27	0.39	0.19	0.22	-0.02	CSE9	0.90	0.38	-0.36	0.18	0.01	0.21	-0.02
CSE4	0.82	-0.29	0.33	0.19	0.18	0.05	CSE2	0.41	0.87	-0.38	0.36	0.15	0.17	-0.08
CSE5	0.80	-0.34	0.23	0.05	0.23	-0.03	CSE3	0.52	0.83	-0.27	0.39	0.19	0.22	-0.02
CSE6	0.77	-0.30	0.24	0.06	0.29	-0.02	CSE1	0.41	0.87	-0.29	0.33	0.12	0.15	-0.10
CSE7	0.77	-0.29	0.19	0.06	0.20	0.03	CA1	-0.30	-0.34	0.89	-0.39	-0.25	-0.21	0.04
CSE8	0.76	-0.28	0.31	0.06	0.21	-0.06	CA2	-0.35	-0.32	0.90	-0.29	-0.17	-0.26	0.02
CSE9	0.72	-0.36	0.18	0.01	0.21	-0.02	CA3	-0.35	-0.31	0.89	-0.25	-0.20	-0.25	-0.08
CSE10	0.71	-0.21	0.17	0.14	0.19	-0.04	CA4	-0.34	-0.33	0.88	-0.32	-0.22	-0.24	0.03
CA1	-0.34	0.89	-0.39	-0.25	-0.21	0.04	EOU1	0.24	0.38	-0.33	0.92	0.19	0.11	-0.02
CA2	-0.36	0.90	-0.29	-0.17	-0.26	0.02	EOU2	0.16	0.37	-0.29	0.89	0.24	0.18	-0.12
CA3	-0.36	0.89	-0.25	-0.20	-0.25	-0.08	EOU3	0.27	0.38	-0.32	0.90	0.18	0.18	-0.05
CA4	-0.32	0.88	-0.27	-0.22	-0.23	-0.05	EOU4	0.19	0.36	-0.34	0.91	0.21	0.16	-0.08
EOU1	0.35	-0.33	0.92	0.19	0.11	-0.02	Gen	0.04	0.18	-0.23	0.23	1.00	0.15	0.23
EOU2	0.29	-0.29	0.89	0.24	0.18	-0.12	Exp	0.28	0.21	-0.27	0.17	0.15	1.00	0.04
EOU3	0.35	-0.32	0.90	0.18	0.18	-0.05	Age	-0.03	-0.08	-0.01	-0.07	0.23	0.04	1.00
EOU4	0.33	-0.30	0.91	0.22	0.15	-0.08								
Gen	0.14	-0.23	0.23	1.00	0.15	0.23								
Exp	0.27	-0.27	0.17	0.15	1.00	0.04								
Age	-0.04	-0.01	-0.07	0.23	0.04	1.00								
ICR	0.75	0.89	0.90	-	-	-	ICR	0.81	0.89	0.89	0.90	-	-	-

Table 3a - Loadings, Cross Loadings, and Reliabilities

Study 2 - Full Self-Efficacy Model							Study 2 - Dimensions Self-Efficacy Model							
ITEMS	CSE	CA	EOU	Gen	Exp	Age	ITEMS	HACSE	ICSE	CA	EOU	Gen	Exp	Age
CSE1	0.50	-0.33	0.41	-0.04	-0.06	-0.03	CSE5	0.85	0.35	-0.16	0.15	-0.07	-0.03	0.21
CSE2	0.54	-0.24	0.19	0.01	-0.10	0.06	CSE6	0.87	0.39	-0.22	0.06	-0.01	0.05	0.09
CSE3	0.70	-0.35	0.26	-0.11	0.03	-0.08	CSE9	0.87	0.24	-0.14	0.09	-0.01	0.06	-0.08
CSE4	0.71	-0.16	0.15	-0.07	-0.03	0.21	CSE2	0.22	0.81	-0.33	0.41	-0.04	-0.06	-0.03
CSE5	0.79	-0.22	0.06	-0.01	0.05	0.09	CSE3	0.27	0.81	-0.24	0.19	0.01	-0.10	0.06
CSE6	0.74	-0.14	0.09	-0.01	0.06	-0.08	CSE1	0.42	0.79	-0.35	0.26	-0.11	0.03	-0.08
CSE7	0.67	-0.11	0.08	0.02	0.07	0.13	CA1	-0.09	-0.35	0.84	-0.31	-0.03	0.01	0.15
CSE8	0.73	-0.19	0.16	-0.13	0.04	0.00	CA2	-0.17	-0.43	0.90	-0.37	-0.02	-0.04	0.17
CSE9	0.64	-0.21	0.35	-0.04	0.07	-0.09	CA3	-0.19	-0.23	0.81	-0.30	-0.20	-0.01	0.06
CSE10	0.70	-0.25	0.23	0.00	-0.03	-0.03	CA4	-0.22	-0.23	0.73	-0.17	0.03	-0.05	0.12
CA1	-0.22	0.84	-0.31	-0.03	0.01	0.15	EOU1	0.11	0.32	-0.21	0.89	-0.03	-0.06	-0.09
CA2	-0.32	0.90	-0.37	-0.02	-0.04	0.17	EOU2	0.09	0.28	-0.37	0.89	-0.04	-0.03	0.07
CA3	-0.23	0.81	-0.30	-0.20	-0.01	0.06	EOU3	0.11	0.33	-0.39	0.92	-0.05	-0.03	-0.07
CA4	-0.25	0.73	-0.17	0.03	-0.05	0.12	EOU4	0.10	0.35	-0.31	0.91	-0.04	0.00	0.03
EOU1	0.27	-0.21	0.89	-0.03	-0.06	-0.09	Gen	-0.03	-0.06	-0.07	-0.04	1.00	-0.14	0.25
EOU2	0.20	-0.37	0.89	-0.04	-0.03	0.07	Exp	0.03	-0.05	-0.02	-0.03	-0.14	1.00	-0.16
EOU3	0.25	-0.39	0.92	-0.05	-0.03	-0.07	Age	0.08	-0.02	0.15	-0.02	0.25	-0.16	1.00
EOU4	0.26	-0.31	0.91	-0.04	0.00	0.03								
Gen	-0.06	-0.07	-0.04	1.00	-0.14	0.25								
Exp	0.02	-0.02	-0.03	-0.14	1.00	-0.16								
Age	0.03	0.15	-0.02	0.25	-0.16	1.00								
ICR	0.67	0.82	0.90	-	-	-	ICR	0.86	0.80	0.82	0.90	-	-	-

Table 3b - Loadings, Cross Loadings, and Reliabilities

	Study 1. General Computer Self-Efficacy Dependent Variable				Study 2. Specific Computer Self-Efficacy Dependent Variable			
	Ease of Use		Computer Anxiety		Ease of Use		Computer Anxiety	
	Path	T-Statistic	Path	T-Statistic	Path	T-Statistic	Path	T-Statistic
Full CSE Model								
Age	-0.11	1.70*	0.02	0.16	0.02	0.25	0.19	2.34*
Gender	0.20	2.52**	-0.16	2.16*	-0.02	0.24	-0.14	1.52~
Computer Experience	0.06	0.61	-0.15	1.97*	-0.04	0.55	-0.02	0.21
Computer Self-Efficacy	0.33	5.63**	-0.34	3.93**	0.36	4.43**	-0.38	4.97**
	(R ² = 0.19)		(R ² = 0.22)		(R ² = 0.13)		(R ² = 0.18)	
Dimensions Model								
Age	-0.09	1.44~	0.02	0.11	0.00	0.00	0.18	2.17*
Gender	0.17	2.31*	-0.17	2.35*	-0.02	0.26	-0.14	1.71*
Computer Experience	0.07	0.80	-0.14	1.81*	-0.02	0.21	-0.03	0.38
Individual CSE	0.35	4.43**	-0.18	2.73**	0.39	4.33**	-0.37	4.16**
Human-Assisted CSE	0.03	0.32	-0.23	2.49*	0.00	0.36	-0.08	0.81
	(R ² = 0.21)		(R ² = 0.23)		(R ² = 0.14)		(R ² = 0.20)	

^a Significance was calculated using a bootstrapping technique. For Study 1, 152 samples were generated. For Study 2, 148 samples were generated. ~ = p<.10, * = p<.05, ** = p<.01

Table 4. PLS Results^a

6 Discussion & Implications

This study suggests that CSE is a multi-faceted construct comprised of two dimensions with distinct attributional sources. In general, results supported the hypotheses. Across independent samples, operationalizations of CSE, and analytic techniques (i.e., card sorting and partial least squares analysis), supported the notion that the Compeau and Higgins' scale provided reliable, distinct measures of individual and human-assisted dimensions of CSE (H1). Even though the results of each study may be deemed anomalous, because it could be a function of the type of specificity of the questions asked, the results of the two studies taken together show a consistent pattern that one would not expect to see by chance. This eliminates the plausible alternative explanation of the study's findings that the results are an artifact of the type of CSE studied. Overall, our findings suggest that IT users may attribute their CSE to individual (i.e., internal) and human-assisted (i.e., external) sources.

Our findings support the notion that the individual CSE dimension correlates with general and specific beliefs about IT. PLS analysis supported Hypotheses 2a and 2b, which state individual CSE should exert a significant positive effect on EOU (Study 1: path = .35, $p \leq .01$, Study 2: path = .39, $p \leq .01$), whereas human-assisted CSE should not affect EOU (Study 1: path = 0.03, n.s., Study 2: path = 0.00, n.s.). This finding is consistent with the self-efficacy construct's foundations in social cognitive theory and attribution theory. These theories suggest that self-efficacy attributable to one's own skill set and/or knowledge base should lead to positive perceptions of ability in a specific context. According to Bandura [Band77], successful performances perceived as resulting from internal causes rather than the circumstance should enhance beliefs about future performance in specific situations such as the ease of use of IT. Additionally, studies in attribution theory [Heid58, Kell71, Mart95, Wein95] have shown that people who perceive themselves as sources of change in their lives should have more stable beliefs across situations [Kloo88, Wein85]. Hence, our findings underscore the importance of individual CSE for encouraging positive perceived ease of use of IT.

Even when organizations offer human support for using information technology, our findings suggest that users will experience anxiety when using a specific IT. Mixed support was found for Hypothesis 3, that states that both individual and human-assisted CSE should negatively affect computer anxiety. At a general level, Study 1 results suggest that both CSE dimensions negatively affect computer anxiety. Human-assisted CSE demonstrated a slightly stronger negative relationship with computer anxiety (path = -.23, $p \leq .01$) than individual CSE did (path

= -.18, $p \leq .01$). This finding implies that at a general level, computer anxiety can be offset by beliefs about one's own capabilities and by forms of human support, and the latter may be very applicable to assisting those experiencing computer anxiety across technologies or software packages. However, when considering specific CSE, Study 2 results suggest that anxiety associated with using a specific IT was diminished solely by individuals' internal judgments of capabilities (path = -0.37, $p \leq .01$); they were not influenced by their awareness of human support for using the software (path = -0.08, $p \leq .01$). Apparently, believing they could do the task with someone's help did not effectively mitigate respondents' anxiety about using Oracle Developer2000. Hence, this finding underscores the importance of building individuals' sense of competence for using software independently of human support. While this finding contradicts Hypothesis 3, it does indicate that for the task specific situation, individual CSE acts different from human-assisted CSE, supporting the overall thesis that these constructs are distinct. Although the individual and human assisted dimensions used fewer items, they offered comparable explanatory power to the full CSE instrument (H4). The CSE dimensions' scales explained only .02 more variance in EOU and computer anxiety than the full CSE instrument. However, because 3-item scales explained approximately the same amount of variance as the full 10-item CSE instrument, this finding suggests that the CSE dimensions' instruments may offer a more parsimonious set of CSE scales for future research.

This paper has implications for Information Systems researchers and practitioners. For research, this paper's findings respond to concerns about the validation of instruments used in Information Systems research [StKa97, ZmBo91]. Our findings clarify CSE's theoretical underpinnings and present parsimonious measures of individual and human-assisted CSE dimensions. Moreover, specifying the dimension(s) of CSE in which researchers are interested can only strengthen the internal, external, and nomological validity of future studies. By identifying reliable, parsimonious measures of CSE and related constructs, results within and across studies may be more effectively interpreted and compared. This is an interesting and difficult issue to comment on, however, since these concerns apply to CSE researchers as well as to the broader community of self-efficacy researchers [Band88]. Coming to a consensus about what CSE means conceptually and how to measure it consistently will take time, requiring practical considerations about the nature of different information technologies, theoretical considerations about the construct's foundations, and communication and debate between CSE researchers.

For applied communities, this research directs attention to how firms evaluate IT training programs. CSE is a frequently used measure of IT training programs' effectiveness. If programs seek to raise participants' overall level of IT competence, trainers should evaluate trainees' general self-efficacy. However, if trainers seek to encourage independent use of IT, our findings suggest trainers should evaluate training programs based on internal (individual) CSE attributions. To build internal efficacy attributions, education research suggests programs emphasize providing positive feedback on prior performance and ability [Schu83, Schu82]. However, because this study does not link training or CSE to performance with IT, additional research is needed which ties specific forms of training and feedback to CSE's dimensions and to outcomes such as IT use in organizations.

7 Conclusion

This present study underscores CSE's multidimensional nature and the associated difficulties with operationalizing its meaning. Rather than a critique of prior research, this study should be interpreted as clarifying CSE's conceptualization and measurement and providing evidence of the usefulness of a refined version of a well-established CSE instrument.

Works Cited

- [AgKa00] Agarwal, R. and E. Karahanna: Time flies when your having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, 2000. 24(4): p. 665-694.
- [AgSS00] Agarwal, R., V. Sambamurthy, and R.M. Stair: The evolving relationship between general and specific computer self-efficacy - An empirical assessment. *Information Systems Research*, 2000. 11(4): p. 418-430.
- [Band97] Bandura, A.: *Self-Efficacy: The Exercise of Control*. 1997, New York: W.H. Freeman.
- [Band77] Bandura, A.: Self-efficacy: Towards a unifying theory of behavioral change. *Psychological Review*, 1977. 84: p. 191-215.
- [Band88] Bandura, A.: Self-regulation of motivation and action through goal systems, in *Cognitive Perspectives on Emotion and Motivation*, V. Hamilton, G.H. Bower, and N.H. Frijda, Editors. 1988, Kluwer Academic Publishers: Dordrecht, Netherlands.
- [Band86] Bandura, A.: *Social Foundations of Thought and Action: A Social Cognitive Theory*. 1986, Englewood Cliffs, NJ: Prentice Hall.

- [BaHT95] Barclay, D., Higgins, C. and Thompson, R.: The Partial Least Squares (PLS) Approach to Causal Modeling: Personal Computer Adoption and Use as an Illustration (with commentaries), *Technology Studies* (2:2), 1995, pp. 285-324.
- [Busc95] Busch, T.: Gender differences in self-efficacy and attitudes toward computers. *Journal of Educational Computing*, 1995. 12(2): p. 147-158.
- [ChLu04] Chan, S. and M. Lu: Understanding Internet Banking Adoption and Use Behavior: A Hong Kong Perspective. *Journal of Global Information Management*, 2004. 12(3): p. 21-43.
- [Chin88] Chin, W.: "The Partial Least Squares Approach to Structural Equation Modeling" in G.A. Marcoulides (Ed.), *Modern Methods for Business Research*, Lawrence Erlbaum Associates, Mahwah, New Jersey, 1988.
- [CoHH99] Compeau, D., C.A. Higgins, and S. Huff: Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly*, 1999. 23(2): p. 145-158.
- [CoHi95a] Compeau, D.R. and C.A. Higgins: Application of social cognitive theory to training for computer skills. *Information Systems Research*, 1995. 6(2): p. 118-143.
- [CoHi95b] Compeau, D.R. and C.A. Higgins: Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 1995. 19(2): p. 189-211.
- [Davi89] Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 1989. 13(3): p. 319-340.
- [DaBW98] Davis, F.D., R.P. Bagozzi, and P.R. Warshaw: User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 1989. 35(8): p. 982-1003.
- [FaNW03] Fagan, M., S. Neill, and B. Wooldridge: An Empirical investigation into the Relationship Between Computer Self-Efficacy, Anxiety, Experience Support and Usage. *The Journal of Computer Information Systems*, 2003/2004. 44(2): p. 95-104.
- [FoLa81] Fornell, C., and Larcker, D.F.: Evaluating Structural Equations Models with Unobservable Variables and Measurement Error, *Journal of Marketing Research*, 1981. 18(1): 1981, pp. 39-50.
- [GiSR89] Gist, M.E., C.E. Schowrerer, and B. Rosen: Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology*, 1989. 74: p. 884-891.
- [Gree91] Green, S.B.: How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, 1991. 26(3): p. 499-510.
- [Hass06] Hassan, B.: Effectiveness of Computer Training. *Journal of Organizational and End User Computing*, 2006. 18(1): p. 50-68.
- [Haya04] Hayashi, A., et al.: The Role of Social Presence and Moderating Role of Computer Self Efficacy in Predicting the Continuance Usage of E-Learning Systems. *Journal of Information Systems Education*, 2004. 15(2): p. 139-154.
- [Heid58] Heider, F.: *The Psychology of Interpersonal Relations*. 1958, New York: Wiley.
- [HeGK87] Heinssen, R.K., C.R. Glass, and L.A. Knight: Assessing computer anxiety: Development and validation of the computer anxiety rating scale. *Computers and Human Behavior*, 1987. 3: p. 49-59.
- [HeSt95] Henry, J. and R. Stone: A Structural Equation Model of Job Performance Using a Computer-Based Order Entry System. *Behaviour and Information Technology*, 1995. 14(3): p. 163-173.

- [HiSM87] Hill, T., N.D. Smith, and M.F. Mann: Role of efficacy expectations in predicting the decision to use advanced technologies: The case of computers. *Journal of Applied Psychology*, 1987. 72(2): p. 307-313.
- [Igli95] Igbaria, M. and A. Iivari: The effects of self-efficacy on computer usage. *Omega*, 1995. 23(6): p. 587- 605.
- [IghPH89] Igbaria, M., F.N. Pravir, and S.L. Huff: Microcomputer applications: An empirical look at usage. *Information and Management*, 1989. 16: p. 187-196.
- [JoMa00] Johnson, R.D. and G.M. Marakas: The role of behavioral modeling in computer skills acquisition - Toward refinement of the model. *Information Systems Research*, 2000. 11(4): p. 402-417.
- [KeCh99] Kelley, H., D. Compeau, and C.A. Higgins: Attribution Analysis of Computer Self-Efficacy. in *Americas Conference on Information Systems*. 1999. Milwaukee, Wisconsin: Association for Information Systems.
- [Kell71] Kelley, H.H.: *Attribution in Social Interactions*. 1971, Morristown, NJ: General Learning Press.
- [Kloo88] Kloosterman, P.: Self-confidence and motivation in mathematics. *Journal of educational Psychology*, 1988. 80: p. 345-351.
- [MaYJ98] Marakas, G.M., M.Y. Yi, and R.D. Johnson: The multilevel and multifaceted character of computer self-efficacy: Toward clarification on the construct and an integrative framework for research. *Information Systems Research*, 1998. 9(2): p. 126-163.
- [Mart95] Martinko, M.J.: The nature and function of attribution theory within the organizational sciences, in *Attribution Theory: An Organizational Perspective*, M. Martinko, Editor. 1995, St. Lucie Press: Delray Beach, FL. p. 7-16.
- [MaHZ96] Martinko, M.J., J.W. Henry, and R.W. Zmud: An attributional explanation of individual resistance to the introduction of information technologies in the workplace. *Behavior and Information Technology*, 1996. 15(5): p. 313-330.
- [ReDM05] Reed, K., D.H. Doty, and D. May: The Impact of Aging on Self-Efficacy and Computer Skill Acquisition. *Journal of Managerial Issues*, 2005. 17(2): p. 212-228.
- [Rubi73] Rubin, Z.: *Liking and loving: An invitation to social psychology*. 1973, New York: Holt, Rinehart, and Winston.
- [Schu83] Schunk, D.H.: Developing children's self-efficacy and skills: THE roles of social comparative information and goal setting. *Contemporary Educational Psychology*, 1983. 8: p. 76-86.
- [Schu82] Schunk, D.H.: The effects of effort attributional feedback on children's achievement: A self-efficacy analysis. *Journal of educational Psychology*, 1982. 74: p. 548-556.
- [Schu84] Schunk, D.H.: Self-efficacy perspective on achievement behavior. *Educational Psychologist*, 1984. 45(3): p. 819-825.
- [ScGu86] Schunk, D.H. and T.P. Gunn: Self-efficacy and skill development: Influence of task strategies and attributions. *Journal of Educational Research*, 1986. 79: p. 238-244.
- [Schw80] Schwab, D.P.: Construct validity in organizational behavior, in *Research in Organizational Behavior*, B.M. Staw and L.L. Cummings, Editors. 1980, JAI Press: Greenwich, CN. p. 3-43.

- [ScPC03] Sheng, Y.P., M. Pearson, and L. Crosby: Organizational Culture and Employees' Computer Self-Efficacy: An Empirical Study. *Information Resources Management Journal*, 2003. 16(3): p. 42-58.
- [Siev88] Sievert, M.E., et al.: Investigating computer anxiety in an academic library. *Information Technology and Libraries*, 1988. 7(9): p. 243-252.
- [SiMG95] Silver, W.S., T.R. Mitchell, and M.E. Gist: Responses to successful and unsuccessful performance: The moderating effect of self-efficacy on the relationship between performance and attributions. *Organizational Behavior and Human Decision Processes*, 1995. 62(3): p. 286-299.
- [Stin68] Stinchcombe, A.L.: Constructing social theories. 1968, New York: Harcourt, Brace, and World. 15-56.
- [StHe03] Stone, R. and J. Henry: The Roles of Computer Self-Efficacy and Outcome Expectancy in Influencing the Computer End User's Organizational Commitment. *Journal of End User Computing*, 2003. 15(1): p. 38-53.
- [StKa97] Straub, D. and E. Karahanna: Knowledge Worker Communications and Recipient Availability: Toward a Task Closure Explanation of Media Choice. *Organization Science*, 1997(In Press): p. 1-37.
- [TaTo95] Taylor, S. and P.A. Todd: Understanding information technology usage: A test of competing models. *Information Systems Research*, 1995. 6(2): p. 144-176.
- [ThCH06] Thompson, R., D. Compeau, and C. Higgins: Intentions to Use Information Technologies: An Integrative Model. *Journal of Organizational and End User Computing*, 2006. 18(3): p. 25-46.
- [Venk00] Venkatesh, V.: Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, 2000. 11(4): p. 342-365.
- [VeDa96] Venkatesh, V. and F.D. Davis: A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*, 1996. 27(3): p. 451-482.
- [WeMa93] Webster, J. and J.J. Martocchio: Turning work into play: Implications for microcomputer software training. *Journal of Management*, 1993. 19(1): p. 127-146.
- [Wein85] Weiner, B.: An Attributional Theory of Achievement Motivation and Emotion. *Psychological Review*, 1985. 92(4): p. 548-573.
- [Wein95] Weiner, B.: *Judgements of Responsibility*. 1995, New York: Guilford Press.
- [WiBr94] Williams, L.J. and B.K. Brown, Method variance in organizational behavior and human resources research on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes*, 1994. 57: p. 185-209.
- [Wold82] Wold, H.: Soft modeling: the basic design and some extensions, in *Systems Under Indirect Observation: Causality, Structure, Prediction*, K.G. Joreskog and H. Wold, Editors. 1982, North Holland: New York. p. 1-47.
- [YiIm04] Yi, M. and K. Im: Predicting Computer Task Performance: Personal Goal and Self-Efficacy. *Journal of Organizational and End User Computing*, 2004. 16(2): p. 20-37.
- [ZmBo91] Zmud, R. and A.C. Boynton: Survey measures and instruments in MIS: Inventory and appraisal, in *The Information Systems Challenge: Survey Research Methods*, K.L. Kraemer, J.I. Cash, Jr., and J.F.J. Nunamaker, Editors. 1991, Harvard Business School: Boston, MA.

Einführung in den Track

Customer und Supplier Relationship Management

Prof. Dr. Freimut Bodendorf

Universität Erlangen-Nürnberg

Prof. Dr. Roland Holten

Universität Frankfurt a.M.

Prof. Dr. Andreas Geyer-Schulz

Universität Karlsruhe (TH)

Das Electronic Business umfasst sowohl operative als auch strategische Aspekte der Geschäftstätigkeit eines Unternehmens mit seinen Partnern. Die Gestaltung und Unterstützung der längerfristigen Geschäftsbeziehungen durch Informations- und Kommunikationssysteme gewinnen dabei zunehmend an Bedeutung. Die u. a. durch Outsourcing-Strategien getriebene Intensivierung der Kooperation von Unternehmen in der vernetzten Wirtschaft führt zu erhöhten Anforderungen an das Lieferantenmanagement. Auf der anderen Seite werden auch Kunden mehr und mehr als Partner des Unternehmens betrachtet, mit denen nicht nur Transaktionen effizient abzuwickeln, sondern auch insbesondere dauerhafte Beziehungen aufzubauen und zu gestalten sind.

Das Beziehungsmanagement mit Kunden und Lieferanten des Unternehmens erfordert einerseits Informationen über diese Partner und andererseits Instrumente zur Kommunikation, Koordination und Kooperation. Der Track widmet sich Strategien, Modellen, Umsetzungskonzepten und Werkzeugen für ein umfassendes Customer und Supplier Relationship Management.

Programmkomitee:

Dr. Frank Bensberg, Universität Münster

Prof. Dr. Christoph Lattemann, Universität Potsdam

Dr. Vanessa Niemeyer, Universität Regensburg

Prof. Dr. Richard Pibernik, MIT-Zaragoza

Dr. Susanne Robra-Bissantz, Universität Erlangen-Nürnberg

Dr. Eric Sucky, Universität Frankfurt a. M.

Optimierung des Customer Lifetime Value auf Basis einzelvertraglicher Entscheidungen

Ein dynamisches Optimierungsmodell

Dr. Ulrich Faisst

Bain & Company, Inc.
D-80335 Munich
ulrich.faisst@bain.com

Martin Gneiser

Lehrstuhl WI-IF (Prof. Buhl)
Universität Augsburg
D-86159 Augsburg
martin.gneiser@wiwi.uni-augsburg.de

Dr. Nina Kreyer

IBM Global Business Services
London, SE1 9PZ, UK
nina.kreyer@uk.ibm.com

Dr. Nina Schroeder

O2 Germany
D-80992 München
nina.schroeder@o2.com

Abstract

Der Customer Lifetime Value (CLV) hat sich als zentrale Größe zur Beurteilung und Gestaltung von Kundenbeziehungen etabliert. Während intuitiv verständlich ist, dass Verträge mit einem positiven Wertbeitrag stets angeboten werden, stellt sich die Frage, ob und wann sich auch der Abschluss von Verträgen mit einem negativen Wertbeitrag lohnt, um zukünftig profitable Geschäfte mit dem Kunden zu tätigen und den CLV zu maximieren. Um solche Strategien adäquat zu unterstützen, sind geeignete Entscheidungsmodelle notwendig. Der folgende Beitrag greift diesen Bedarf auf und schlägt ein dynamisches Optimierungsmodell vor, das Unternehmen bei Entscheidungen über das Angebot oder die Ablehnung einzelner Verträge unterstützt und leistet damit einen Beitrag zu einer wertorientierten Unternehmensführung.

1 Einleitung

Zur Maximierung des Customer Lifetime Value (CLV) über die gesamte Kundenlebensdauer stehen Entscheider im täglichen Geschäft vor der Frage, welche Verträge mit welchen Kunden (nicht) zu kontrahieren sind. Hierzu kann es sinnvoll sein, bspw. zu Beginn der Kundenbeziehung aus einer Einzelgeschäftsicht unwirtschaftliche Kontrakte zu schließen (und bspw. Kunden ein kostenfreies Girokonto zur Verfügung zu stellen), um im späteren Verlauf der Beziehung ertragreiche Geschäfte zu realisieren. Trotz der recht umfangreichen Forschung auf dem Gebiet des CLV sind Modelle, die solche Entscheidungen unterstützen, bislang rar. Dieser Beitrag greift diesen Bedarf auf und entwickelt aufbauend auf eine kurze Literaturanalyse ein Modell, welches den Abschluss verschiedener Kontrakte als Sequenz einzelner Investitionsentscheidungen analysiert und mit Hilfe eines dynamischen Optimierungsmodells den CLV eines Kunden maximiert.¹

Die zentrale Forschungsfrage des Beitrags lautet:

Wie kann die Kundenbeziehung als Sequenz einzelner Entscheidungen über die Akzeptanz bzw. Ablehnung einzelner Kontrakte dargestellt werden und wie kann eine optimale Entscheidungsreihenfolge gewährleistet werden, die den Abschluss einzelner unwirtschaftlicher Verträge zulässt, den CLV langfristig jedoch maximiert?

Während Entscheidungsträger im Management i.d.R. eine ganzheitliche Sicht auf den Kunden fordern, die alle Verträge des gesamten Kundenlebenszyklus (sowie ggf. weitere Elemente wie die Kundenhistorie oder relationale Daten) berücksichtigt, steht der einzelne Vertriebsmitarbeiter (täglich) vor der Entscheidung, ob der Abschluss einer konkreten, vom Kunden nachgefragten (Dienst-) Leistung wirtschaftlich ist. Das vorgestellte Modell soll diese scheinbar gegensätzlichen Ziele verbinden und sowohl die Erreichung der strategischen Ziele des Managements (hier: der Maximierung des CLV über die gesamte Kundenbeziehung) als auch der operativen Ziele des Vertriebs bei der Entscheidung über den Abschluss einzelner Verträge sicherstellen.

Darüber hinaus werden folgende Forschungsfragen adressiert:

¹ Im Rahmen des Beitrags wird häufig auf Beispiele aus der Finanzdienstleistungsbranche zurückgegriffen; das Modell lässt sich aber grundsätzlich auch auf andere Branchen übertragen. Analysen zu den immateriellen Werten der Unternehmen [vgl. GuLe03; GuLS04; NaRa05] machen die Relevanz eines effektiven Kundenmanagements gerade für Finanzdienstleister deutlich. Die Tatsache, dass Banken von 52% profitablen, 23% nicht-profitablen und 25% profit-neutralen Kunden ausgehen [Mess05, S. 50] und vielfach 20% der Kunden 80% der Umsätze bzw. 75% der Gewinne generieren [vgl. KaKP05; Koch04; Plat04] zeigt zusätzlich, dass ein fundiertes, wertorientiertes Management der Kundenbeziehung auch und gerade in der Finanzdienstleistungsbranche notwendig ist. Dies gelingt jedoch nur in wenigen Fällen und die sich durch CRM bietenden Potenziale bei Finanzdienstleistern werden nicht oder nur unzureichend ausgeschöpft [vgl. KaKP05].

Wie lässt sich die optimale Dauer einer Kundenbeziehung ermitteln?

Zur Maximierung des CLV muss sichergestellt werden, dass die Kundenbeziehung erst endet, wenn die ertragreichen Geschäfte mit dem jeweiligen Kunden abgeschlossen wurden. Ab dem Zeitpunkt, ab dem keine weitere Steigerung des CLV möglich ist, sollte bspw. auf zusätzliche Marketing-Aktivitäten verzichtet werden.

Lassen sich generelle Regeln ableiten, die Unternehmen bei der Entscheidung für bzw. gegen das Angebot von Verträgen unterstützen und wie hoch ist der Wertbeitrag, den ein Kontrakt mindestens erreichen muss, um aus Sicht des Unternehmens ökonomisch sinnvoll zu sein?

2 Stand der Forschung und Literatureinbettung

Im Rahmen des Kunden-Beziehungsmanagements [vgl. JoSe05, S.11] propagieren einschlägige Publikationen seit vielen Jahren die Vorteilhaftigkeit kundenzentrischer Strategien [vgl. Pe-Wa82, ReiSa90], die im Verkauf eines Produkts oder einer Dienstleistung erst den Anfang (statt wie lange Zeit eher das Ende) einer Beziehung zwischen dem Unternehmen und einem Kunden sehen. Statt einer isolierten Betrachtung verschiedener Einzelgeschäfte soll durch die umfassende und ganzheitliche Analyse der Kundenbeziehung ein nachhaltiger Beitrag zum Unternehmenserfolg geleistet werden. In diesem Zusammenhang wurde der Begriff des Customer Relationship Management (CRM) geprägt. Hierunter wird im Folgenden ein bereichsübergreifender, kundenorientierter Managementansatz verstanden, bei dem mit Hilfe moderner IKS die Initiierung, Gestaltung, Erhaltung sowie Verstärkung von Kundenbeziehungen durch ganzheitliche Marketing-, Vertriebs- und Servicekonzepte unter Berücksichtigung von Profitabilitätskriterien angestrebt wird.²

2.1 Der CLV als zentrale Größe zur Bewertung der Kundenbeziehung

Einer der Gründe dafür, dass CRM-Initiativen vielfach die in sie gesetzten Erwartungen nicht erfüllen, ist in Mess- und Prognoseproblemen der durch CRM erreichten bzw. erreichbaren Erfolge zu suchen. Während sich der Erfolg produktzentrierter Konzepte vergleichsweise einfach ermitteln lässt, verlangt der Übergang zu kundenzentrischen Strategien neue Messinstrumente und -methoden, da hier Ergebnisse ex ante schwer(er) prognostizierbar sind Ursache-Wirkungs-Zusammenhänge nicht immer eindeutig identifiziert werden können. Als Prädiktoren

² Für weitere Definitionen des CRM-Begriffs sei auf [AlPÖ05], [BrMi03] und [HiWi04], S. 16f verwiesen.

der Profitabilität kundenzentrischer Strategien werden häufig vorökonomische Zielgrößen wie Vertrauen, Loyalität, Commitment, Kundennähe oder –zufriedenheit genannt. In der einschlägigen Literatur wird dabei insbesondere der Kundenzufriedenheit – als zentralem Faktor zur Beeinflussung von Wiederholungskäufen, die wiederum unmittelbar auf die Umsätze wirken – sowie dem Zusammenhang zwischen der Kundenzufriedenheit und Faktoren wie der Loyalität und der Profitabilität hohe Aufmerksamkeit geschenkt [vgl. Kraf99; Gerp00; Kotl03, Kap. 3; PfFa04; ReKu00; ReKu02; ReKr01]. Obwohl eine umfangreiche wissenschaftliche Auseinandersetzung mit den oben genannten Prädiktoren stattfindet, sind die Ergebnisse nach wie vor ambivalent und es bestehen unterschiedliche Auffassungen darüber, in welchem Ausmaß diese Faktoren tatsächlich dazu geeignet sind, die Profitabilität einer Kundenbeziehung zu beurteilen. So zeigt eine aktuelle Studie, dass finanzwirtschaftliche Metriken im Vergleich zu „weichen“ Faktoren eine bessere Prognose zukünftiger Verkäufe erlauben [vgl. NaRa05]. Dagegen konnten Anderson et al. einen stark positiven Zusammenhang zwischen der Kundenzufriedenheit und dem Unternehmenserfolg aufweisen [vgl. AnFL94] und auch Gerpott zeigt, dass ein positiver Zusammenhang zwischen der Kundenzufriedenheit und der Wiederkaufabsicht besteht, wobei die Wirkung der Kundenzufriedenheit auf die *tatsächliche* Wiederkauftrate nicht untersucht wurde [vgl. Gerp00]. Edvardsson et al. kommen in ihrer Studie zu dem Ergebnis, dass von einer hohen Loyalität (und Kundenzufriedenheit) im industriellen Sektor negative Effekte auf den Unternehmenserfolg ausgehen, während beides in serviceorientierten Branchen positiv auf den Erfolg wirkt [EJGS00]. Neben anderen Arbeiten zeigen auch Bolton et al. in einer aktuellen Untersuchung, dass die Zusammenhänge zwischen den verschiedenen Prädiktoren und dem Unternehmenserfolg i.A. komplex und nur schwer messbar sind [vgl. BoLV04; s.a. Hall96; MiKa01]. Vor diesem Hintergrund sollten Entscheider – statt bspw. einer ausschließlichen Fokussierung auf die Verlängerung bestehender Kundenbeziehungen oder der Erhöhung der Loyalität – die tatsächlich durch den Kunden erreichbaren Werte ermitteln und auf dieser Basis Entscheidungen über die optimale Gestaltung der Kundenbeziehung treffen [vgl. KuRB04; ThRK04]. Aufgrund der mit vielen herkömmlichen Verfahren zur Kundenbewertung³ verbundenen Probleme (wie bspw. die mangelnde Berücksichtigung zukünftiger Perioden, der hohe Einfluss subjektiver Einschätzungen oder die mangelhafte Quantifizierbarkeit der verschiedenen Einflussfaktoren) wurde in den letzten Jahren verstärkt das Konzept des CLV propagiert, mit dessen Hilfe die Bewertung der mit dem Kunden (direkt oder indirekt) zu erwirtschaftenden

³ Vgl bspw. [KrA199] für einen Überblick zu diesen Verfahren.

Einzahlungsüberschüsse über die gesamte Kundenbeziehungsdauer möglich wird. Im Allgemeinen wird der CLV – der originär auf Dwyer zurückgeht [vgl. Dwyer89] – als Summe der abdiskontierten Cash-flows während der gesamten Kundenbeziehung definiert. Dabei leistet der CLV einen monetären Ergebnisbeitrag, in den alle mit dem Kunden verbundenen direkten und indirekten, gegenwärtig und künftig erreichbaren Zahlungen über die gesamte Kundenbeziehungsdauer einfließen.⁴ Das Konzept des CLV genießt in der wissenschaftlichen Auseinandersetzung hohe Aufmerksamkeit und wird zur Analyse verschiedener unternehmerischer Probleme angewandt.⁵ Im Folgenden wird ein Modell vorgestellt, welches ausgehend vom einzelnen Kunden sowohl das Management bei der Gestaltung von CRM-Strategien unterstützt, als auch auf der Ebene einzelner Kunden optimale Investitionsentscheidungen zur Maximierung des CLV erlaubt.

2.2 Kundenbeziehungsmuster und Kaufverhalten

Im Rahmen des Beitrags wird von einem sog. „Kundenmigrations“-Szenario ausgegangen. Dieser Begriff wird u.a. bei Dwyer verwendet, der bei den von ihm entwickelten CLV-Modellen zwischen einem „Customer-retention environment“ und einem „Customer-migration environment“ unterscheidet [vgl. Dwyer89]. Die einschlägige Forschung unterscheidet zwischen Modellen, in denen Kunden alle Bedarfe bei einem Unternehmen decken und solchen, in denen Kunden zur Bedarfsdeckung mehrere Anbieter nutzen. Innerhalb der erstgenannten Modelle geht man davon aus, dass Kunden, die den Anbieter wechseln „für immer“ verloren sind und bezeichnet es daher als „lost-for-good“-Szenario. Kehrt ein Kunde zu einem späteren Zeitpunkt wieder zu einem Anbieter zurück, wird er als neuer Kunde betrachtet und seine Historie vernachlässigt. Im Gegensatz dazu scheint die zweite, sog. „always-a-share“-Situation, in der Kunden ihren Gesamtbedarf bei verschiedenen Unternehmen decken realistisch und soll daher im Rahmen der folgenden Modellierung unterstellt werden.⁶

3 Grundlagen der Modellierung

⁴ Für eine ausführliche Auseinandersetzung mit dem CLV-Konzept und seinen verschiedenen Forschungsrichtungen – die nicht Gegenstand dieses Beitrags sind – sei auf BeNa98; B1Th01; BoLV02; DhGl03; Dwyer89; GiKo02; GuLe03; LiSh05; JaSi02; JoSe04; KuRB04; PfHC05; RuZL02; ScPe95; ZeRL01 verwiesen.

⁵ Vgl. dazu bspw. BeNa01; Dwyer89; GuLe03; GuLS04; JaSi02; KeWa95; KuRB04

⁶ Anmerkung: In diesem Szenario werden (ehemalige) Kunden, die aktuell keine Produkte nachfragen, als Bestandskunden betrachtet und für den Fall, dass sie zu einem späteren Zeitpunkt erneut in Kontakt mit dem Anbieter treten, nicht wie Neukunden behandelt.

Das vorgestellte Modell unterstellt eine Entscheidungssequenz zwischen einem Finanzdienstleister und einem Kunden, in deren Verlauf der Kunde verschiedene Produkte des Finanzdienstleisters nachfragt. Der Finanzdienstleister entscheidet mit dem Ziel der CLV-Maximierung über eine Angebotsunterbreitung, was wiederum die Zufriedenheit des Kunden beeinflusst und dazu führt, dass der Kunde während seiner Kundenbeziehung – mit bestimmten Wahrscheinlichkeiten – einen oder mehrere Verträge mit dem Anbieter abschließt.

3.1 Modellannahmen

Dem Modell liegen einige Annahmen zu Grunde, die nachfolgend dargestellt werden sollen, bevor das Modell spezifiziert und die Ergebnisse präsentiert werden.

- A1 Ein Kunde fragt während seines Lebenszyklus $m=1, \dots, n$ verschiedene Kontrakte zu unterschiedlichen Zeitpunkten t_m (mit $t_1=0$) nach. Die Art der Kontrakte und die Zeitpunkte zu denen diese nachgefragt werden, sind bekannt.
- A2 Das Unternehmen entscheidet zu jedem Zeitpunkt, ob er den vom Kunden nachgefragten Kontrakt anbietet. Diese Entscheidung wird durch die Entscheidungsvariable s_m repräsentiert, wobei gilt: $s_m = \begin{cases} 1 & \text{der Vertrag } m \text{ wird dem Kunden angeboten} \\ 0 & \text{der Vertrag } m \text{ wird dem Kunden nicht angeboten} \end{cases}$.

Weiterhin wird angenommen, dass die Entscheidung des Unternehmens keinen Einfluss auf bereits mit dem Kunden geschlossene Verträge hat.

- A3 Die Zustandsvariable $x_m \in [0;1]$ repräsentiert die Zufriedenheit des Kunden zum Zeitpunkt t_m . Die Zufriedenheit des Kunden zu Beginn der Kundenbeziehung ist dem Unternehmen bekannt, so dass gilt: $x_1 = x_s$.
- A3a Wenn der Kunde zum Zeitpunkt t_{m+1} keinen Kontrakt schließen will, hängt die Variable $x_{m+1} \in [0;1]$ nur von der Zufriedenheit des Kunden in der Vorperiode t_m ab.⁷ Für diesen Fall wird angenommen, dass die Zufriedenheit des Kunden mit jeder Periode, in der er keinen neuen Vertrag nachfragt, um einen Faktor $\beta_m \in (0;1)$ abnimmt.⁸

⁷ An dieser Stelle sei darauf hingewiesen, dass der Periodenbegriff hier nicht – wie in anderen Arbeiten z.T. unterstellt – davon ausgeht, dass Perioden Zeiträume gleicher Dauer abgrenzen, sondern die Periodenlänge hier durchaus unterschiedlich sein kann.

⁸ Diese Annahme ist u.a. vor dem Hintergrund plausibel, dass Kunden, die über einen längeren Zeitraum keine Verträge mit dem (angestammten) Anbieter abschließen, vermutlich eine engere Beziehung mit einem anderen Unternehmen eingehen und hierdurch eine Unzufriedenheit mit dem bisherigen Finanzdienstleister zum Ausdruck bringen. Darüber hinaus lässt sich mit Hilfe des Parameters β_m auch eine intuitive Darstellung der Kundenmigrationssituation erreichen.

A3b Wenn der Kunde in t_{m+1} einen Kontrakt nachfragt, hängt seine Zufriedenheit von seiner Zufriedenheit in der Vorperiode t_m und davon ab, ob das Unternehmen seine Anfrage positiv oder negativ beantwortet. In diesem Fall wird zusätzlich angenommen, dass die Kundenzufriedenheit von einem „Zufriedenheitsparameter“ $\alpha_m \in R_+$ beeinflusst wird, die den Nutzen/Disnutzen des Kunden ausdrückt, wenn ihm das Unternehmen für einen nachgefragten Kontrakt kein Angebot macht.⁹

Die Transformationsfunktionen $f_{\alpha_m} : R_+^2 \rightarrow R_+$ und $f_{\beta_m} : R_+ \rightarrow R_+$ spezifizieren die verbal beschriebene Überführung der Kundenzufriedenheit von einer Periode zur nächsten:¹⁰

$$x_{m+1} = \begin{cases} f_{\alpha_m}(x_m, s_m) = \frac{1}{1 + \alpha_m} (x_m + \alpha_m \cdot s_m) & \text{wenn der Vertrag } m \text{ vom Kunden nachgefragt wird} \\ f_{\beta_m}(x_m) = \beta_m \cdot x_m & \text{wenn der Kunde Vertrag } m \text{ nicht nachfragt} \end{cases}$$

Die Funktion f_{α_m} wurde so gewählt, dass die Kundenzufriedenheit (dargestellt durch den Parameter α_m) je nachgefragtem Kontrakt mit einem Vertragsangebot steigt und mit einer Ablehnung der Kundenanfrage sinkt. Die Transformationsfunktionen sind darüber hinaus so gestaltet, dass eine sinkende Kundenzufriedenheit durch Angebote in Folgeperioden wieder ausgeglichen (oder übertroffen) werden kann. Zur Illustration wird in Abbildung 1 der Einfluss, den die Entscheidung, dem Kunden einen Vertrag zum Zeitpunkt t_m (nicht) anzubieten auf dessen Zufriedenheit in t_{m+1} hat, dargestellt. Die Abbildung zeigt, dass Kunden mit einer höheren Ausgangszufriedenheit ($x_m=0,8$) stärker auf einen ablehnenden Bescheid reagieren ($x_{m+1}=0,27$, $\Delta=0.53$), als Kunden, deren Zufriedenheit geringer ist ($x_m=0,3 \rightarrow$ Ablehnung $\rightarrow x_{m+1}=0,1$, $\Delta=0.2$). Ergo führt eine Zusage bei einem Kunden mit geringer Zufriedenheit zu höheren „Zufriedenheitszuwachsen“ ($x_m=0,3 \rightarrow$ Zusage $\rightarrow x_{m+1}=0,77$, $\Delta=0.47$) als bei vergleichsweise zufriedenen Kunden ($x_m=0,8 \rightarrow$ Zusage $\rightarrow x_{m+1}=0,93$, $\Delta=0.13$).¹¹

⁹ Mit Hilfe des Parameters α_m lassen sich kontraktsspezifische Nutzen/Disnutzen darstellen, die bspw. reflektieren, dass ein Kunde eine höhere Unzufriedenheit aufweist, wenn ihm ein für ihn besonders wichtiges Produkt verweigert wird, als bei der Ablehnung eines für ihn eher unwichtigen Produktes.

¹⁰ Neben den beiden an dieser Stelle spezifizierten Transformationsfunktionen existieren selbstverständlich auch andere, mit den Annahmen kompatible Funktionen, so dass die hier gewählten Funktionen lediglich der Illustration des Vorgehens dienen.

¹¹ Abhängig vom Wert der Variable α_m fallen die maximale Kundenzufriedenheit begrenzenden (und in der Abbildung gepunkteten) Linien steiler (für $\alpha_m < 2$) bzw. flacher (für $\alpha_m > 2$) aus. Hieraus erkennt man bereits, dass mit Hilfe des Parameters α_m neben dem bislang unterstellten Kundenmigrations-Szenarios auch eine „lost-for-good“-Situation dargestellt werden kann, in dem von einem vergleichsweise hohen α_m ausgegangen wird. In

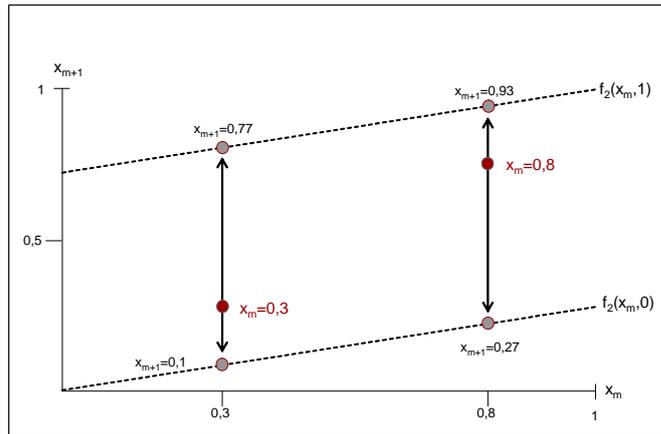


Abb. 1: Illustration der Transformationsfunktion mit $\alpha_m=2$

A4 Die Wahrscheinlichkeit, dass ein Kunde ein Angebot für einen Kontrakt m nachfragt ($p(x_m)$), hängt von seiner Zufriedenheit x_m ab, es gilt also: $p(x_m) = x_m$ mit $p(x_m) \in [0;1]$.

A5 Das Unternehmen gilt als risikoneutraler Entscheider. Da es lediglich darüber entscheiden kann, ob es einen vom Kunden nachgefragten Kontrakt anbietet, dieser dann aber mit einer bestimmten Wahrscheinlichkeit $q_m \in (0;1)$ den Abschluss tätigt¹², werden angebotene Kontrakte nicht automatisch geschlossen. Der Wertbeitrag V_m misst den Wert eines Vertrags zum Zeitpunkt t_m und ergibt sich aus dem Produkt der Entscheidungsvariable s_m , der Wahrscheinlichkeit q_m , dass der Kunde, das Angebot annimmt und dem erwarteten Kontraktwert $E(KW_m)$. Letzterer ergibt sich aus der Summe der aus dem Vertrag resultierenden Cash-flows abdiskontiert zum Zeitpunkt t_m in dem der Vertrag geschlossen wird. Der Zins r wird über die Zeit hinweg als konstant angenommen und Kosten, die ggf. für die Angebotserstellung anfallen, werden vernachlässigt, so dass gilt:

$$V_m(s_m) = s_m \cdot q_m \cdot E(KW_m) = s_m \cdot q_m \cdot E\left(\sum_{i=1}^T CF_{m,i} \cdot (1+r)^{t_m-t_i}\right) \quad (1)$$

A6 Der Wert eines Kunden zum Zeitpunkt t_m ($= CLV_m$) ergibt sich aus der Summe aller Vertragswerte des Kunden, abdiskontiert auf den Zeitpunkt t_m wie folgt:

$$CLV_m = \sum_m^n V_m \cdot (1+r)^{-t_m} \quad \text{mit } r \in [0;1]. \quad (2)$$

diesem Fall nähern sich beiden Szenarien an und es wird (aufgrund der hohen Unzufriedenheit des Kunden) zunehmend unwahrscheinlich, dass der Kunde weitere Produkte nachfragt.

¹² Die Variable q_m hat keinen Einfluss auf die Zufriedenheit des Kunden, sondern dient an dieser Stelle lediglich dazu, den Fall, dass ein Kunde ein Angebot des Finanzdienstleisters ablehnt, abzubilden.

A7 Das Unternehmen verfolgt das Ziel, den CLV eines Kunden zu maximieren.

Die wiederholte Interaktion zwischen einem Kunden und dem Unternehmen kann als dynamisches Optimierungsproblem modelliert werden, in dem das Unternehmen in jeder Periode entscheidet, ob er ein (potenziell) vom Kunden nachgefragtes Produkt anbietet. Der Zusammenhang zwischen den verschiedenen Modellvariablen s_m , x_m , den jeweiligen Transformationsfunktionen und den Wertbeiträgen V_m einzelner Kontrakte kann der folgenden Abbildung entnommen werden:

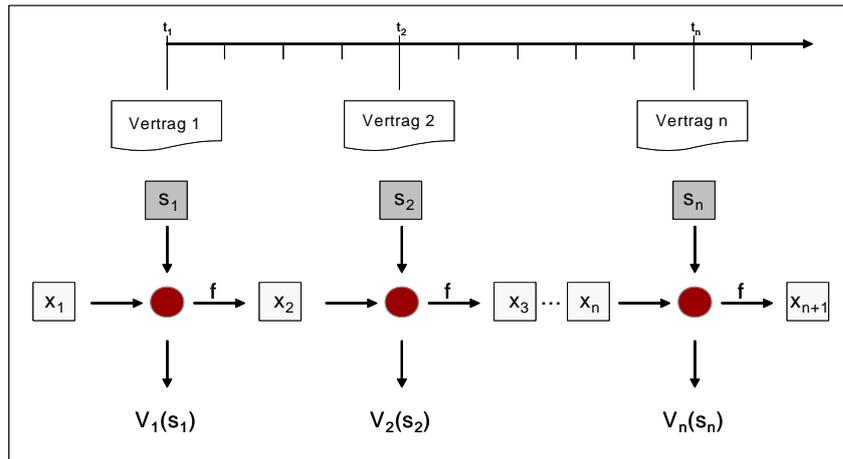


Abb. 2: Zusammenhang der Modellvariablen

Aufbauend auf den obigen Annahmen kann nun ein Entscheidungsbaum (s. Abb. 3) für die Nachfrage des Kunden nach den jeweiligen Kontrakten zu unterschiedlichen Zeitpunkten abgeleitet werden:

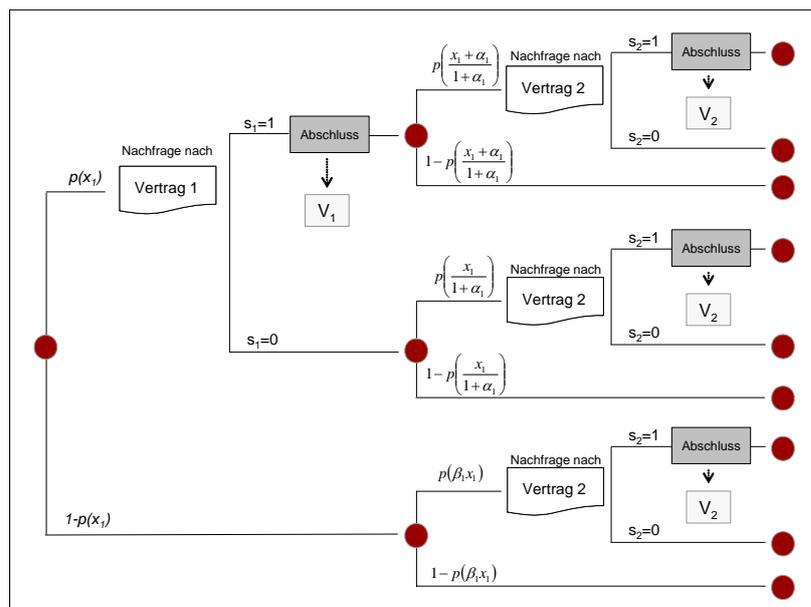


Abb. 3: Entscheidungsbaum aus Sicht des Unternehmens

Zur Entscheidung darüber, ob einem Kunden ein nachgefragter Vertrag angeboten wird, müssen alle Äste des Entscheidungsbaums berücksichtigt werden. Die Lösung des Optimierungsproblems lässt sich mit Hilfe der (Lösung der) Bellmanschen Funktionalgleichung, die wie folgt lautet, ermitteln:

$$CLV_m^*(x_m) = p(x_m) \cdot V_m(s_m) + \lambda_{m+1} \cdot [p(x_m) \cdot CLV_{m+1}^*(f_{\alpha_m}(x_m, s_m)) + (1 - p(x_m)) \cdot CLV_{m+1}^*(f_{\beta_m}(x_m))] \Rightarrow \max! \quad (3)$$

mit $\lambda_{m+1} = (1 + r)^{t_m - t_{m+1}}$ und $CLV_m^* : R \rightarrow R$.

3.2 Modellergebnisse

Durch die Lösung der obigen Zielfunktion lässt sich die optimale Entscheidungssequenz, die implizit die optimale Kundenzufriedenheit und Aussagen über die optimale Dauer der Kundenbeziehung beinhaltet, ermitteln [vgl. Bell57].¹³ Darüber hinaus lässt sich eine Entscheidungsregel für das Unternehmen ableiten, ob es den jeweils nachgefragten Kontrakt anbieten soll oder nicht. Nachfolgend werden diese Ergebnisse detailliert dargestellt.

3.2.1 Ermittlung optimaler Entscheidungssequenzen zur CLV-Maximierung

Die Lösung der Zielfunktion zeigt, dass sich die optimalen Entscheidungssequenzen s^* und die optimalen Kundenzufriedenheiten x^* wie folgt ergeben:

$$s^* = (s_1^*(x_s), s_2^*(x_2), \dots, s_n^*(x_n))^T \Leftrightarrow x^* = (x_s, x_2, \dots, x_n)^T = (x_s, f_{\alpha_1}(x_s, s_1), \dots, f_{\alpha_{n-1}}(x_{n-1}, s_{n-1}))^T \quad (4).$$

Diese optimalen Sequenzen determinieren den maximal erreichbaren Kundenwert des jeweiligen Kunden. Es zeigt sich, dass die Entscheidung, ob ein Kontrakt angeboten werden sollte von der Zufriedenheit des Kunden abhängt. Es sind also Situationen denkbar, in denen auf ein Angebot verzichtet wird, während es in einem anderen Fall (bei anderer Kundenzufriedenheit aber ansonsten identischen Voraussetzungen) angeboten wird. Darüber hinaus zeigt sich, dass es für Unternehmen rational sein kann, einem Kunden kein Angebot zu machen und die daraus resultierenden negativen Folgen (wie eine sinkende Kundenzufriedenheit) in Kauf zu nehmen.

Obwohl dieses Ergebnis aus einer ökonomischen Perspektive einleuchtet, bedeutet die Umsetzung der daraus resultierenden Strategie umfassende Änderungen bei bestehenden Geschäftspraktiken. So würden Fälle auftreten, in denen Kunden (zu denen u.U. eine langjährige Ge-

¹³ Die optimale Entscheidungssequenz beschreibt dabei eine optimale Anfangsentscheidung und zustandsabhängige bedingt-optimale Entscheidungen für die Kontrakte $m=2, \dots, n$.

schäftsbeziehung besteht) aufgrund der ermittelten optimalen Politik kein Angebot für eine nachgefragte Leistung erhalten würden. Darüber hinaus müssten Vertriebsmitarbeiter auf den Abschluss von Geschäften verzichten, um die langfristigen Ziele des Unternehmens (die Maximierung des CLV über die gesamte Beziehungsdauer) zu erreichen. Dies ist insbesondere bei Unternehmen, die ihre Vertriebsmitarbeiter (wie bspw. Versicherungen) derzeit in erster Linie über Abschlussprovisionen steuern, nur schwer durchsetzbar. Gleichzeitig zeigen die Projekterfahrungen der Autoren, dass derzeit viele Unternehmen der Finanzdienstleistungsbranche ihre bestehenden Provisionierungssysteme überprüfen und neue, die kundenzentrischen Strategien unterstützenden Provisionierungs- bzw. Vertriebssteuerungssystemen entwickeln. Das vorgestellte Modell unterstützt die Bemühungen dahingehend, als dass bspw. Provisionen nur für solche Verträge gezahlt werden sollten, deren Abschluss sich auch auf Basis des vorgestellten Modells empfiehlt. Ggf. könnte darüber hinaus auch ein Incentivierungssystem entwickelt werden, welches die Erhöhung der Abschlusswahrscheinlichkeit auf Seiten des Kunden belohnt.

3.2.2 Ermittlung der optimalen Dauer der Kundenpflege

Wie die Wahrscheinlichkeit $p(x_m)$ mit der der Kunde einen Vertrag nachfragt zeigt, gibt es (eigentlich) keinen Zeitpunkt, zu dem eine Kundenbeziehung als beendet bezeichnet werden kann, da Kunden zu beliebigen Zeitpunkten Kontrakte nachfragen können. Betrachtet man jedoch die oben ermittelte Sequenz optimaler Entscheidungen über Vertragsangebote, so erkennt man, dass ab einem bestimmten Zeitpunkt keine Vertragsangebote durch das Unternehmen erfolgen werden. Dieser Zeitpunkt wird mit t_{opt} bezeichnet und es gilt: $t_{opt} = t_{m-1}$ mit $s_i^*(x_i) = 0 \forall i \geq m$. Auch wenn ab diesem Zeitpunkt selbstverständlich die Kundenbetreuung für die bis zu diesem Zeitpunkt geschlossenen Verträge (die bereits in den jeweiligen vertragsspezifischen Cashflows berücksichtigt sind) weitergeführt wird, sollten ab diesem Zeitpunkt keine zusätzlichen Investitionen zur Kundenpflege (bspw. im Sinne von Präsenten oder Vergünstigungen) erfolgen, da diese unwirtschaftlich wären. Dieses Ergebnis bietet große Einsparpotenziale, da sich Marketing-Aktivitäten nun präziser steuern und gezielt auf einzelne Kunden ausrichten lassen, statt – wie bisher – große Streuverluste bei Kampagnen in Kauf nehmen zu müssen.

3.2.3 Entscheidungen über das Angebot einzelner Kontrakte

Die beiden obigen Ergebnisse unterstützen insbesondere Entscheidungen, die sich auf die Kundenbeziehung während des gesamten Kundenlebenszyklus beziehen. Darüber hinaus sind je-

doch auch Instrumente wichtig, die Vertriebsmitarbeiter bei der Entscheidung darüber unterstützen, ob ein aktuell vom Kunden nachgefragter Kontrakt angeboten werden sollte. Darüber hinaus kann die Analyse auf einzelvertraglicher Basis auch Aufschluss darüber geben, welcher minimale Vertragswert aus Sicht des Finanzdienstleisters gerade noch tolerabel ist, um bspw. Entscheidungen über Rabatte oder andere Vergünstigungen zu treffen. Dieser minimale Vertragswert lässt sich – abhängig von der Wahrscheinlichkeit, mit der ein Kunde das Angebot annimmt – wie folgt ermitteln:

$$E(KW_m^{\min}) = - \left(\lambda_{m+1} \cdot \left[CLV_{m+1}^* \left(\frac{x_m + \alpha_m}{1 + \alpha_m} \right) - CLV_{m+1}^* \left(\frac{x_m}{1 + \alpha_m} \right) \right] \right) \cdot q_m^{-1} \quad (5).$$

Im Rahmen von Verhandlungen über die Konditionen eines Vertragsangebots können hierdurch die aus Sicht des Unternehmens gerade noch akzeptablen Vertragswerte ermittelt und damit Kundengespräche effektiver geführt und der Vertrieb beim Abschluss von Geschäften unter der Berücksichtigung kundenzentrischer Ziele unterstützt werden.

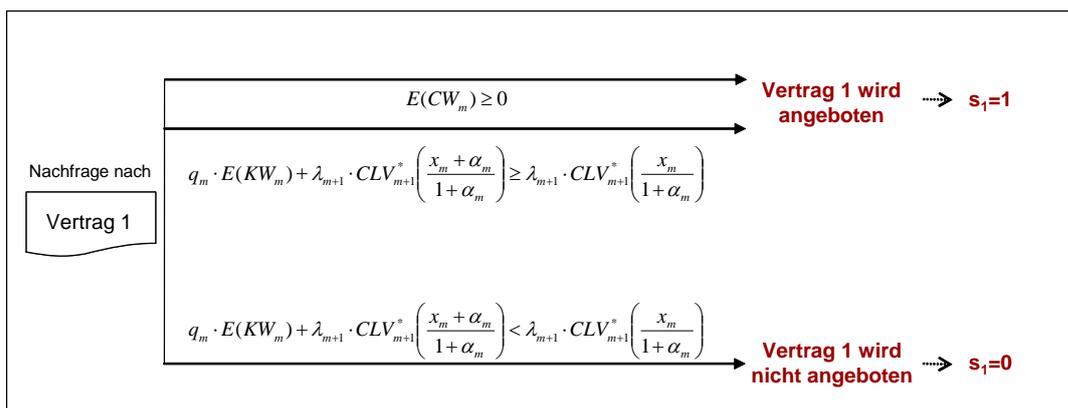


Abb. 4: Entscheidungsregeln für einen vom Kunden nachgefragten Vertrag

Abbildung 4 verdeutlicht, ausgehend von einem vom Kunden nachgefragten Vertrag, die daraus resultierende Entscheidungsregel, ob dieser Vertrag durch den Finanzdienstleister angeboten werden sollte. Die Grafik zeigt, dass – unabhängig von Kundenwertbetrachtungen – immer dann ein Vertrag angeboten wird, wenn der erwartete Kontraktwert $E(KW_m)$ positiv ist. Darüber hinaus werden auch dann Angebote erstellt, wenn der Vertragswert zwar negativ ist, dies aber durch die erwarteten positiven Cash-flows späterer Geschäfte (über-)kompensiert werden kann *und* dieser CLV größer ist als der CLV des Kunden, wenn das Geschäft nicht angeboten wird. Dieses Ergebnis steht im Einklang mit derzeit am Markt beobachteten Strategien. So bietet bspw. die HypoVereinsbank ein sog. „Starter Paket“ für junge Kunden an, welches – neben einem kostenfreien Giro-Konto – eine Reihe attraktiver und kostenfreier Zusatzleistungen bein-

haltet. Dahinter verbirgt sich das Kalkül, dass junge Kunden vergleichsweise günstig akquiriert werden können und (aufgrund der erreichten Kundenbindung und –zufriedenheit) im Verlauf ihrer späteren Kundenbeziehung ertragreiche Geschäfte tätigen welche die negativen Cashflows der ersten Zeit kompensieren. Auch wenn beim Starter Paket nicht alle der im Modell definierten Annahmen gelten und das Entscheidungskalkül daher nicht exakt übertragen werden kann, zeigt das Beispiel, dass es in der unternehmerischen Realität eine Reihe interessanter Anwendungsbereiche für die entwickelte Methodik gibt und sich hierdurch fundierte Entscheidungen über die Kundenbetreuung und –bearbeitung treffen lassen. Darüber hinaus lassen sich im Modell auch Szenarien ableiten, in denen Kunden – obwohl sie einen positiven CLV besitzen – kein nachgefragtes Angebot unterbreitet würde, da sich hieraus ein negativer Vertragswert ergeben kann, der den CLV (der nach wie vor größer Null sein kann) senken würde. Dieses Ergebnis unterscheidet sich von Empfehlungen gängiger Forschungsarbeiten, die i.d.R. davon ausgehen, dass Kunden immer bedient werden, so lange wie der Gesamt-CLV positiv bleibt.

4 Zusammenfassung und Ausblick

Das vorgestellte Modell zeigt, dass sich scheinbar konträre Ziele (wie die eher langfristige Perspektive des Managements und die eher abschlussorientierte, zeitpunktbezogene Sicht des Vertriebs) integriert optimieren lassen und sowohl der CLV über die gesamte Kundenbeziehungsdauer maximiert als auch Entscheidungen über das Angebot einzelner Verträge im Einklang mit den Zielen des CRM getroffen werden können. Dabei konnten – neben intuitiv einsichtigen Empfehlungen wie der, bspw. dann Verträge anzubieten, wenn der aus dem Vertrag resultierende Kontraktwert positiv ist – auch Situationen identifiziert werden, die (wie bspw. die Ablehnung eines Vertrags, obwohl der Kunde über einen positiven CLV verfügt) zunächst fraglich erscheinen, aber vor dem Hintergrund eines ökonomischen Kalküls plausibel sind. Das Modell kann sowohl bei Neukunden als auch bei Bestandskunden angewandt werden. Weiterhin lässt sich das Modell sowohl für strategische Entscheidungen des Managements (z.B. bei der Entwicklung ökonomisch fundierter CRM-Strategien) als auch für die Steuerung, Kontrolle und Unterstützung des Vertriebs einsetzen. So lassen sich mit Hilfe des Modells z.B. Empfehlungen von Kundenberatern (die über langjährige Erfahrungen und gute Kenntnisse der einzelnen Kunden verfügen) zum Abschluss vermeintlich unwirtschaftlicher Geschäfte prüfen und es können

Regeln abgeleitet werden, unter welchen Bedingungen auch Verträge, die isoliert betrachtet einen negativen Wertbeitrag leisten, kontrahiert werden sollten.

Neben den beschriebenen Vorteilen soll an dieser Stelle auch auf die dem Modell immanenten Risiken und Limitationen eingegangen und hieraus zusätzlicher Forschungsbedarf abgeleitet werden.

Die Optimierung des CLV auf Basis von Einzelverträgen stellt hohe Anforderungen an die verwendeten Daten. Während sich diese in Segmenten wie dem Retail-Geschäft, für das viele Informationen verfügbar und die Kunden einander eher ähnlich sind, z.B. mit Hilfe adäquater Data-Mining-Methoden [vgl. Bens02] vermutlich (zumindest in guter Näherung) ermitteln lassen, kann das Modell bei zunehmender Spezifität der Kunden nur mit verhältnismäßig hohem Aufwand implementiert werden. Darüber hinaus sollte untersucht werden, inwieweit die Annahme, nach der eine (ablehnende) Entscheidung über ein Kontrahierungsangebot keinen Einfluss auf die bestehenden Geschäfte hat, ggf. relativiert werden muss. Gleiches gilt für die bisher im Modell fehlende Berücksichtigung des relationalen CLV-Potenzials, da zu vermuten ist, dass Kunden sich über Angebote des Finanzdienstleisters austauschen und insbesondere unzufriedene Kunden wenig dazu beitragen werden, andere Kunden von den Vorteilen des Finanzdienstleisters zu überzeugen und das Image des Anbieters schädigen können. Darüber hinaus wären Forschungen hilfreich, welche die Rolle des Vertriebsmitarbeiters (der i.d.R. viele wertvolle Informationen über die von ihm betreuten Kunden besitzt) genauer untersuchen und bspw. Instrumente entwickeln, wie dieser sein Wissen im Sinne des Unternehmens in das Entscheidungsmodell einbringen kann, ohne dabei den Verlust von Einfluss oder Kompetenzen befürchten zu müssen. Bisher wurde im Modell lediglich die Kundenzufriedenheit als wesentlicher, die Kaufentscheidung beeinflussender Faktor berücksichtigt. Wie im Abschnitt 2 gezeigt, gibt es darüber hinaus auch eine Reihe weiterer, den Wert einer Kundenbeziehung beeinflussende Faktoren (wie Vertrauen, Loyalität oder aufgebaute Wechselbarrieren), deren Berücksichtigung einen Beitrag zur besseren Adaption des Modells an die unternehmerische Realität leisten würde. Darüber hinaus wären empirische Arbeiten wünschenswert, die helfen, das Modell zu untermauern (oder zu verwerfen) und weiter zu entwickeln.

Literaturverzeichnis

- [AnFL94] Anderson, Eugene W.; Fornell, Claes; Lehmann, Donald R.: Customer satisfaction, market share, and profitability. Evidence from Sweden. In: Journal of Marketing, Vol. 56, 1994, S. 53-66.
- [Bell57] Bellmann, Richard E.: Dynamic Programming, Princeton University Press, Princeton, 1957.
- [BeNa98] Berger, Paul D.; Nasr, Nada: Customer Lifetime Value: Marketing Models and Applications. In: Journal of Interactive Marketing, Vol. 12, Nr. 1, 1998, S. 18..
- [BeNa01] Berger, Paul D.; Nasr, Nada: The allocation of promotion budget to maximize customer equity. In: Omega. The International Journal of Management Science, Vol. 29, 2001, S. 49-61.
- [Bens02] Bensberg, Frank: Data Mining im Rahmen des analytischen CRM. Methoden und Anwendungen. In: Mayer, R.; Kemper, H.-G. (Hrsg.): Business Intelligence in der Praxis Erfolgreiche Lösungen für Controlling, Vertrieb und Marketing, Hrsg.: Bonn 2002, S. 63-87.
- [BlTh01] Blattberg, Robert C.; Thomas, Jacquelyn: Customer Equity: Building and Managing Relationships as Valuable Assets, Boston, 2001.
- [BoLV04] Bolton, Ruth N.; Lemon, Katherine N.; Verhoef, Peter C.: The Theoretical Underpinnings of Customer Asset Management: A Framework and Propositions for Future Research. In: Journal of the Academy of Marketing Sciences, Vol. 32, Nr. 3, 2004, S. 271-292.
- [Dwy89] Dwyer, Robert F.: Customer Lifetime Valuation to Support Marketing Decision Making. In: Journal of Direct Marketing, Vol. 8, Nr. 2, 1989, S. 8-15.
- [EJGS00] Edvardsson, Bo; Johnson, Michael D.; Gustafsson, Anders; Strandvik, Tore: The Effects of Satisfaction and Loyalty on Profits and Growth: Products Versus Services. In: Total Quality Management, Vol. 11, Nr. 7, 2000, S. 917-927.

- [Gerp00] Gerpott, Torsten, J.: Kundenbindung – Konzepteinordnung und Bestandsaufnahme der neueren empirischen Forschung. In: Die Unternehmung, Vol. 54, Nr. 1, 2000, S. 23-42.
- [GiKo02] Gierl, Heribert; Koncz, Julia: Customer Lifetime Value. In: Dallmer, H (Hrsg.): Das Handbuch des Direct Marketing & More. 8. Auflage, Gabler, Wiesbaden, 2002, S. 939-956.
- [GuLe03] Gupta, Sunil; Lehmann, Donald R.: Customer As Assets. In: Journal of Interactive Marketing, Vol. 17, Nr. 1, 2003, S. 9-24.
- [GuLS04] Gupta, Sunil; Lehmann, Donald R.; Stuart, Jennifer A.: Valuing Customers, in: Journal of Marketing Research, Vol. 41, Nr. 1, 2004, S. 7-18.
- [Hall96] Hallowell, Roger: The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. In: International Journal of Service Industry Management, Vol. 17, Nr. 1, 1996, S. 27-40.
- [HiWi04] Hippner, Hajo; Wilde, Klaus, D.: Grundlagen des CRM – Konzepte und Gestaltung, Gabler-Verlag, Wiesbaden, 2004.
- [JaSi02] Jain, Dipak; Singh, Siddarth: Customer Lifetime Value Research in Marketing: A Review and Future Directions. In: Journal of Interactive Marketing, Vol. 16, Nr. 2, 2002, S. 34-46.
- [JoSe05] Johnson, Michael D.; Selnes, Fred: Diversifying Your Customer Portfolio. In: MIT Sloan Management Review, Spring, 2005, S. 11-14.
- [KaKP05] Karakostas, Bill; Kardaras, Dimitri; Papathanassiou, Eleutherios A.: The state of CRM adoption by the financial services in the UK: an empirical investigation. In: Information & Management, Vol. 42, Nr. 6, 2005, S. 853-863.
- [KeWa95] Keane, Timothy, J.; Wang, Paul: Applications for the Lifetime Value Model in Newspaper Publishing. In: Journal of Direct Marketing, Vol. 9, Nr. 2, 1995, S. 59-66.
- [Koch04] Koch, Richard: Das 80/20-Prinzip, Campus-Verlag, Berlin, 2004.

- [Kotl03] Kotler, Philip: Marketing Management. 11. Auflage, Prentice Hall, New York, 2003.
- [Kraf99] Krafft, Manfred: Der Kunde im Fokus: Kundennähe, Kundenzufriedenheit, Kundenbindung – und Kundenwert? In: Die Betriebswirtschaft, Vol. 59, Nr. 4, 1999, S. 511-530.
- [KrAl99] Krafft, Manfred; Albers, Sönke: Ansätze zur Segmentierung von Kunden – Wie geeignet sind herkömmliche Konzepte? In: Zeitschrift für betriebswirtschaftliche Forschung, Vol. 52, 1999, S. 515-536.
- [KuRB04] Kumar, V.; Ramani, Girish; Bohling, Timothy.: Customer Lifetime Value Approaches and Best Practice Applications. In: Journal of Interactive Marketing, Vol. 8, Nr. 3, 2004, S. 60-72.
- [Mess05] Messner, Wolfgang: CRM-Investitionen – wann sind sie rentabel? In: Die Bank, Nr. 3, 2005, S. 50-54.
- [LiSh05] Liu, Duen-Ren; Shih, Ya-Yueh.: Integrating AHP and data mining for product recommendation based on customer lifetime value. In: Information & Management, Vol. 42, Nr. 3, 2005, S. 387-400.
- [MiKa01] Mittal, V.; Kamakura, W.A.: Satisfaction, Repurchase Intent, and Moderating Effect of Customer Characteristics. In: Journal of Marketing Research, Vol. 38, Nr. 1, 2001, S. 177-194.
- [NaRa05] Nagar, Venky; Ragan, Madhav: Measuring Customer Relationships: The Case of the Retail Banking Industry. In: Management Science, Vol. 51, Nr. 6, 2005, S. 904-919.
- [PeWa92] Peters, Thomas, J.; Waterman, Robert, H.: In search of excellence, Warner Books, New York, 1982.
- [PfFa04] Pfeifer, P.; Farris, P.: The elasticity of customer value to retention: the duration of a customer relationship. In: Journal of Interactive Marketing, Vol. 18, Nr. 2, 2004, S. 20-31.

- [PfHC05] Pfeifer, Phillip, E.; Haskins, Mark, E.; Conroy, Robert, M: Customer Lifetime Value, Customer Profitability, and the Treatment of Acquisition Spending. In: Journal of Managerial Issues, Vol. 17, Nr. 1, 2005, S. 11-25.
- [ReiSa90] Reichheld, F.; Sasser, W.: Zero Defections: Quality comes to service. In: Harvard Business Review, 68 Jg., Vol. 5, S. 105-111.
- [ReKr01] Reinartz, Werner; Krafft, Manfred: Überprüfung des Zusammenhangs von Kundenbindungsdauer und Kundenertragswert. In: Zeitschrift für Betriebswirtschaft, Vol. 71., 2001, S. 1263-1281.
- [ReKu00] Reinartz, Werner; Kumar, V.: On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing. In: Journal of Marketing, Vol. 64, Nr. 4, 2000, S. 17-35.
- [ReKu02] Reinartz, Werner; Kumar, V.: The Mismanagement of Customer Loyalty, Harvard Business Review, July 2002, S. 86-94.
- [RuZL02] Rust, Roland T. / Zeithaml, Valerie A. / Lemon, Katherine: Driving Customer Equity: How Customer Lifetime Value is Reshaping Corporate Strategy, New York, 2002.
- [ScPe95] Schmittlein, David C.; Peterson, Robert A.: Customer Base Analysis: An Industrial Purchase Process Application. In: Marketing Science, Vol.13, Nr. 1, 1995 S. 41-68.
- [ThRK04] Thomas, Jacquelyn S.; Reinartz, Werner; Kumar, V.: Getting the Most out of All Your Customers. In: Harvard Business Review, Juli-August 2004, S. 117-123.
- [ZeRL01] Zeithaml, Valarie A.; Rust, Roland T.; Lemon, Katherine N.: The Customer Pyramid: Creating and Serving Profitable Customers. In: California Management Review, Vol. 43, Nr. 4, 2001, S. 118-142.

Spezifizierung des Kano-Modells zur Messung von Kundenzufriedenheit

Hans Ulrich Buhl, Dennis Kundisch, Annette Renz, Nicola Schackmann

Lehrstuhl WI-IF
Universität Augsburg
86135 Augsburg
{hans-ulrich.buhl,dennis.kundisch}@wiwi.uni-augsburg.de

Zusammenfassung

„Weiche“ Faktoren wie die Kundenzufriedenheit werden bei IT-Projektentscheidungen in ihrer Bedeutung für den Unternehmenserfolg oft vernachlässigt oder nur unzureichend mit einbezogen. In diesem Beitrag wird ein Modell vorgestellt, welches auf Basis des Ansatzes von Kano dazu dient, die Auswirkungen der Einführung neuer Leistungen oder der Steigerung ihrer Qualität hinsichtlich deren Wirkung auf die Gesamtzufriedenheit der Zielkundensegmente konsistent zu bewerten. Die so bestimmten Zufriedenheitswerte können als Input für eine Schätzung der Auswirkungen von Maßnahmen auf die Kundenloyalität dienen und einen Beitrag zur Objektivierung von Projektentscheidungen leisten.

1 Einleitung¹

Die Erzeugung von Kundenzufriedenheit ist kein unternehmerischer Selbstzweck, sondern soll der Steigerung des Unternehmenswertes dienen. Zufriedenheitssteigernde Maßnahmen werden dabei zunehmend als strategische Investitionen in die Kundenbindung angesehen, die sich im Laufe der Zeit amortisieren und den Erhalt und die Steigerung des Unternehmenswerts sichern müssen. [MaSt00, S. 630] Da diese Maßnahmen jedoch überwiegend immaterielles und somit schwer quantifizierbares Vermögen schaffen, das mit den heutigen Controlling-Systemen nur unzureichend erfasst werden kann, [Kraf99, S. 513] wird in der Praxis die Bedeutung der Kundenzufriedenheit in Ermangelung valider Zahlen oftmals unterschätzt und bei einer Entscheidung für oder gegen ein Projekt gerade bei einer kurzfristigen Interpretation des Shareholder-

¹ Die Autoren danken drei anonymen Gutachtern für die zahlreichen wertvollen Hinweise zur Verbesserung des Beitrags, welche im Hinblick auf die Platz- und Zeitrestriktionen so weit möglich aufgegriffen wurden.

Value-Konzeptes nicht in ausreichendem Maße gewürdigt. In vielen Unternehmen besteht ein Bedarf, die Ergebniswirkungen solcher Investitionen besser abschätzen zu können.

Es stellt sich die Frage, wie eine Bestimmung der Gesamtzufriedenheit einzelner Kunden bzw. von (Ziel-)Kundensegmenten mit bestehenden und insbesondere geplanten Unternehmensleistungen auf Basis der (erwarteten) Zufriedenheit mit Einzelleistungen erfolgen kann. Der Ansatz von Kano stellt dabei einen in der Praxis weit verbreiteten Ansatz zur Messung von Kundenzufriedenheit dar. [BHMS96; KaHu01; Saue00] Gleichwohl stehen für eine Transformation der Ergebnisse einer Kundenbefragung nach Kano in „harte“ und vergleichbare Aussagen über den Nutzenbeitrag durch die Steigerung der Kundenzufriedenheit auf Basis mehrerer Maßnahmenprogramme bisher noch keine geeigneten Methoden zur Verfügung.

Dies war auch der Ausgangspunkt eines gemeinsam mit der Deutschen Bank durchgeführten Projektes. Ziel des Projektes war die Entwicklung eines Entscheidungsunterstützungstools zur Bewertung von Maßnahmenprogrammen für Funktionserweiterungen und -modifikationen im geschlossenen Bereich der Private Banking Webseite. Der Ansatz von Kano erschien für die Bewertung des Kundenzufriedenheitssteigerungspotenzials eines Maßnahmenprogramms zwar grundsätzlich geeignet, war aber wegen der lediglich qualitativen Formulierung nicht direkt anwendbar, sondern bedurfte einer Spezifikation (dies stellt auch eine in der Literatur identifizierte Forschungslücke dar [KaHu01, S. 142f.]). Auf Basis des erwarteten Effekts verschiedener Maßnahmenprogramme auf die Kundenzufriedenheit wurde in darauf aufbauenden Schritten der Einfluss der Kundenzufriedenheit auf die Loyalität und letztendlich auf den Unternehmenserfolg geschätzt. Im vorliegenden Beitrag steht nur der erste Schritt – die konsistente Bewertung des Kundenzufriedenheitseffekts von Maßnahmenprogrammen – im Vordergrund, da hier u. E. verallgemeinerbare Ergebnisse erzielt werden konnten.

Der Beitrag ist wie folgt gegliedert. Auf Basis einer kurzen Vorstellung der theoretischen Basis für das Konstrukt Kundenzufriedenheit und des Kano-Modells (Abschnitt 2) wird darauf aufbauend ein Bewertungsmodell entwickelt, welches dazu dient, Unternehmenseinzelleistungen hinsichtlich ihres Zufriedenheitsbeitrags zu bewerten und konsistent zu einem Gesamtzufriedenheitswert zu aggregieren (Abschnitt 3). Eine kurze Zusammenfassung sowie ein Ausblick auf den weiteren Forschungsbedarf runden den Beitrag ab (Abschnitt 4).

2 Kundenzufriedenheit und das Kano-Modell

Als grundsätzlicher Beitrag zum Verständnis der Kundenzufriedenheit hat das Confirmation/Disconfirmation-Paradigma (C/D-Paradigma) weite Verbreitung in der Kundenzufriedenheitsforschung erlangt. [Oliv97, S. 99] Laut des C/D-Paradigmas ist die Kundenzufriedenheit eine Einstellung, die aus einem Soll-Ist-Vergleich resultiert. Unter der „Ist-Leistung“ wird dabei die wahrgenommene Qualität eines Produkts oder einer Leistung verstanden. Die Sollkomponente als subjektive Größe stellt dagegen einen Vergleichsstandard dar und drückt sich bspw. in einer bestimmten Erwartungshaltung des Kunden aus. Sie unterliegt verschiedenen Einflussgrößen, zu denen vorhandene und bewusste Bedürfnisse ebenso zählen wie bereits gemachte Erfahrungen mit einer Leistung oder Mund-zu-Mund-Propaganda. Entspricht die wahrgenommene Leistung genau dem Vergleichsstandard, so ist das sog. Konfirmationsniveau der Zufriedenheit erreicht. Übertrifft die wahrgenommene Leistung die Erwartungen, spricht man von positiver Diskonfirmation (Zufriedenheit), im gegenteiligen Fall von negativer Diskonfirmation (Unzufriedenheit). [HoSt01, S. 20] Das C/D-Paradigma basiert hauptsächlich auf der kognitiven und weniger auf der affektiven Evaluation der Performance eines Produkts oder einer Leistung und ist daher eher zur Anwendung geeignet, wenn funktionale Anforderungen – wie dies beim Internetangebot eines Finanzdienstleisters oft der Fall sein dürfte – bei der Evaluation im Vordergrund stehen. [WiLe03, S. 346]

Die merkmalsorientierte Messmethode der Kundenzufriedenheit von [KTST84] unterstellt, dass nicht bei allen Produkten ein linearer Zusammenhang zwischen der Funktionalität des Produkts und der Kundenzufriedenheit besteht. Kano et al. gehen zudem davon aus, dass ein Kunde mit einzelnen Aspekten eines Produkts (un-)zufrieden sein kann, und dass sich die Gesamtzufriedenheit mit diesem Produkt aus seinen Teilurteilen zusammensetzt. Dafür teilen Kano et al. sämtliche Produktbestandteile oder -eigenschaften in *Basis-*, *Leistungs-* und *Begeisterungsanforderungen* ein, die in unterschiedlicher Weise zur Entstehung von Kundenzufriedenheit beitragen.

Zahlreiche Autoren verwenden das Kano-Modell und übertragen es auf die Kundenzufriedenheitsforschung. [BHMS96; HoSt01; KaHu01; MaSt00]² Statt der Funktionalität eines Produkts wie bei Kano wird bei der Übertragung auf die Kundenzufriedenheit die Erwartungserfüllung betrachtet. [BHMS96, S. 118; HoSt01, S. 33] In diesem Kontext können die drei Kategorien von Kano folgendermaßen beschrieben werden:

² Zur Kritik an der Kano-Methode vgl. [KaHu01, S. 142f.; Saue00, S. 159].

Basisfaktoren (BaF) sind „Mussfaktoren“, deren Angebot und Erfüllung als selbstverständlich vorausgesetzt wird, sodass die BaF nicht mehr explizit vom Kunden nachgefragt werden. Die Erfüllung der Erwartungen dringt beim Kunden oftmals nicht ins Bewusstsein. Insofern wird das Konfirmationsniveau der Kundenzufriedenheit im besten Fall erreicht, jedoch nicht überschritten. Den Kunden fällt eine Nichterfüllung ihrer als selbstverständlich verstandenen Erwartungen sehr deutlich auf, sodass eine Nichterfüllung der BaF zu einem starken Absinken der Kundenzufriedenheit unter das Konfirmationsniveau führt.

Begeisterungsfaktoren (BeF) werden nicht explizit vom Kunden nachgefragt, da es sich um innovative Faktoren handelt, an die der Kunde noch gar nicht denkt. Er hat also noch keine Erwartungen gebildet, sodass der Vergleich von Ist und Soll immer positiv ausfällt und die Zufriedenheit oberhalb des Konfirmationsniveaus liegt. Die Nichterfüllung der BeF führt nicht zu Unzufriedenheit, da der Kunde ihre Erfüllung nicht erwartet bzw. ihm das Fehlen einer Innovation nicht auffällt.

Leistungsfaktoren (LF) werden vom Kunden verlangt. Werden sie angeboten, führen sie zu einer Steigerung der Kundenzufriedenheit, werden sie nicht oder nur in unzureichender Qualität angeboten, führen sie zu Unzufriedenheit. Es wird dabei ein linearer Zusammenhang zwischen dem Erfüllungsgrad der Erwartungen und dem Zufriedenheitsniveau vermutet; bei exakter Erfüllung der Erwartungen wird gerade das Konfirmationsniveau erreicht.

Die so beschriebenen Zusammenhänge sind in der folgenden Abbildung dargestellt (vgl. Abb. 1).³ Die Abszisse der Graphik drückt hierbei die Erwartungserfüllung der einzelnen Leistungskomponenten aus, also den Soll-Ist-Vergleich zwischen erwarteter und wahrgenommener Leistung, die Ordinate zeigt die Ausprägungen der dazugehörigen Kundenzufriedenheit (dabei stellt der Null-Punkt den Referenzpunkt dar [HvNH98, S. 1229]). Im Zeitablauf verschieben sich in der Regel die Zugehörigkeiten einzelner Leistungen zu den drei Kategorien derart, dass sich die BeF zu LF und die LF zu BaF verändern. Dies begründet sich daraus, dass Eigenschaften, die heute noch unerwartete Begeisterung beim Kunden auslösen, morgen bereits als normal betrachtet und vom Kunden erwartet werden. [HoSt01, S. 27ff.] Damit einher geht auch ein Ab-

³ Der Vollständigkeit halber sollen auch noch die Indifferenz- und die Unzufriedenheitsfaktoren genannt werden. Indifferenzfaktoren (IF) werden vom Kunden weder erwartet noch genutzt. Diese haben auf die Kundenzufriedenheit keinen Einfluss. [Tont00, S. 729] Bei Investitionsentscheidungen müssen IF jedoch dann Berücksichtigung finden, wenn diese eine Funktionalität implementieren, auf welche weitere zufriedenheitsrelevante Funktionalitäten aufbauen. Unzufriedenheitsfaktoren sind Funktionalitäten, welche beim Kunden allein auf Grund des Vorhandenseins Unzufriedenheit auslösen. [Tont00, S. 729] Solche Faktoren können über den Kano-Fragebogen identifiziert werden und werden auf Grund ihrer Wirkungsweise in einem vorgeschlagenen Maßnahmenprogramm offensichtlich nicht enthalten sein. Beide Faktoren werden in der folgenden Analyse nicht betrachtet.

sinken der Kundenzufriedenheit über die Zeit, sofern keine neuen Investitionen getätigt werden – ein Umstand, der auch im Projekt mit der Deutschen Bank festgestellt werden konnte. Die gestrichelten Pfeile in der Abbildung sollen diesen Übergang von Leistungen in die niedrigeren Kategorien im Zeitablauf andeuten.

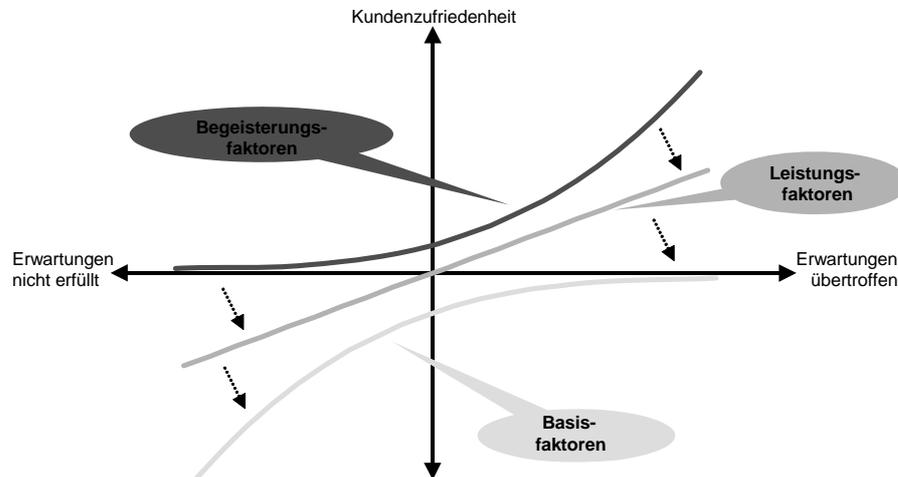


Abb. 1: Kano-Modell übertragen auf die Kundenzufriedenheit [BHMS96, S. 118]

Bei einer genaueren Analyse der Übertragungen des Kano-Gedankenguts auf die Kundenzufriedenheit wurde im Projekt deutlich, dass die grafische und rein qualitative Darstellung trotz aller Plausibilität der zu Grunde liegenden Aussagen einige Schwächen aufweist. Es bleibt bspw. die Frage offen, warum im Bereich der BaF trotz exakter Erfüllung der Erwartungen nicht das Konfirmationsniveau entsteht. Zudem wird unterstellt, dass auch eine Nichterfüllung der Erwartungen bei den BeF auftreten kann, obwohl hier der Kunde per definitionem gar keine Erwartungen gebildet hat. Da sich in der Literatur keine exaktere Modellbeschreibung sowie quantitative Weiterentwicklung des Modells findet – zur qualitativen Auswertung vgl. bspw. [Sae00, S. 40ff.] –, greifen wir dieses Defizit in der Kundenzufriedenheitsforschung auf.

3 Bewertungsmodell zur Kundenzufriedenheit

Folgende allgemeine Überlegungen liegen unserem Ansatz zu Grunde: Ein Finanzdienstleister bietet verschiedene Leistungen oder Leistungsbündel in seinem Webauftritt an, die zu Zufriedenheit führen. Diese Zufriedenheit beeinflusst das Verhalten der Kunden und somit letztendlich (über mehrere Zwischenschritte und im Zusammenhang mit weiteren Wirkungsfaktoren) den Unternehmenserfolg. [HoSt01, S. 61] Im Fokus dieser Arbeit steht der Zusammenhang der

Zufriedenheit bzw. Unzufriedenheit mit einzelnen Unternehmensleistungen und dem daraus resultierenden Zufriedenheitsbeitrag (ZB) sowie die aus den ZB zu ermittelnde Gesamtzufriedenheit. Dabei werden zunächst die Beziehungen zwischen den einzelnen Leistungen und den jeweiligen ZB nach Kano hergestellt. Im Anschluss daran erfolgt eine Aggregation der einzelnen ZB innerhalb einer Kano-Kategorie und schließlich über alle drei Kategorien zu einem Gesamtzufriedenheitswert.

3.1 Zufriedenheit mit einzelnen Unternehmensleistungen

Die Gesamtheit der Kunden wird nicht gleichermaßen auf die Beeinflussung ihrer Kundenzufriedenheit auf Grund bestimmter Leistungen reagieren, sodass – um diese Heterogenität durch Segmentierung zu berücksichtigen – im Folgenden von mehreren (jeweils homogenen) Kundensegmenten ausgegangen wird, bei denen einzelne Leistungen des Unternehmens stets in ähnlicher Weise in ZB dieses Kundensegments resultieren. Dies schließt auch mit ein, dass verschiedene Kundensegmente teilweise dieselbe Leistung in unterschiedliche Kano-Kategorien einteilen.

Basierend auf den Ausführungen von [HoSt01] und [BHMS96] liegen bei einer quantitativen Darstellung des Kano-Modells der Kundenzufriedenheit die Werte auf der Abszisse zwischen den Extremen „Nichterfüllung“ und „Übererfüllung“ der Erwartungen (vgl. Abb. 1) und entstehen – wie im C/D-Paradigma postuliert – als Ergebnis eines Vergleichs des tatsächlichen Leistungsniveaus $k_{t,i}$ mit dem Erwartungsniveau $e_{l,t,i}$, den das (Ziel)Kundensegment $l = 1, \dots, L$ (mit $L \in \mathbb{N}^+$) in jeder Periode t für jede Unternehmensleistung $i = 1, \dots, n$ zieht. Dabei wird angenommen, dass das Leistungsniveau im Zeitablauf konstant ist, die Erwartungen jedoch in jeder Periode monoton ansteigen (vgl. Abschnitt 2). Eine Veränderung des kundensegmentunabhängigen Leistungsniveaus wäre in diesem Sinne gleichbedeutend mit einer neuen Funktionalität.

Der Vergleich von Erwartungs- und Leistungsniveau liefert einen Wert für die Erfüllung der Erwartungen $x_{l,t,i}$ als Ergebnis. Des Weiteren werden aus Modellierungsgründen sowohl das Leistungs- als auch das Erwartungsniveau auf das Intervall $[0,1]$ normiert. Für das Leistungsniveau bedeutet dies, dass es sowohl einen bestmöglichen Wert für die Leistungsqualität (1) gibt als auch einen Wert, der die schlechtest denkbare Leistungsqualität (0) ausdrückt. Bei den Erwartungen gilt das Gleiche – sie sind ebenfalls nach oben und unten begrenzt. 1 bedeutet dabei höchste Ansprüche von Seiten der Kunden, während der Wert 0 angibt, dass der Kunde keine Erwartungen an die Leistung stellt. Es gilt:

- (1) $k_{t,i} = k_i \quad k_i \in [0,1]$
- (2) $e_{l,t-1,i} \leq e_{l,t,i} \quad e_{l,t,i} \in [0,1]$
- (3) $x_{l,t,i} = k_i - e_{l,t,i}, \quad x_{l,t,i} \in [-1,1]$

Auf Grund der vorgenommenen Normierungen für k_i und $e_{l,t,i}$ ergibt sich für $x_{l,t,i}$ ein Wertebereich von $[-1,1]$, wobei ein Wert von $x_{l,t,i} = -1$ bedeutet, dass die Erwartungen des Kundensegments „überhaupt nicht erfüllt“, $x_{l,t,i} = 1$, dass sie „völlig übererfüllt“, und $x_{l,t,i} = 0$, dass die Erwartungen „gerade erfüllt“ werden und somit das Konfirmationsniveau erreicht wird. Die Festlegung von Maximal- und Minimalwerten für die Erwartungserfüllung erscheint intuitiv plausibel, geht man davon aus, dass auf der einen Seite ein völlig unzufriedener Kunde im Zeitablauf nicht noch unzufriedener werden kann und auf der anderen Seite die Begeisterung eines zufriedenen Kunden ebenfalls eine Obergrenze erreichen wird.

Der ZB resultierend aus einer Unternehmensleistung existiert in jeder Periode t abhängig vom Grad der Erwartungserfüllung $x_{l,t,i}$ bezüglich dieser Leistung, wobei der funktionale Zusammenhang zwischen dem Soll-Ist-Vergleich, d. h. der Erfüllung der Erwartungen, und des ZB in jeder der drei Kategorien anders verläuft. An dieser Stelle wollen wir die bisherige qualitative Darstellung des Kano-Modells weiterentwickeln. Für eine präzisere und quantifizierbare Übertragung und Anwendbarkeit des Kano-Modells im Bereich der Kundenzufriedenheit werden zunächst die sich aus unserer Kritik ergebenden Anforderungen formuliert, um dann insbesondere die Funktionsverläufe der BaF und BeF gegenüber dem Kano-Modell der Kundenzufriedenheit von [HoSt01] und [BHMS96] zu modifizieren.

Dafür wird zunächst die Variable $s_{l,t,i,j}$ für den ZB, welchen eine Unternehmensleistung – bspw. eine Postkorb-Funktionalität – verursacht, eingeführt und auf Werte im Intervall $[-1,1]$ normiert. $s_{l,t,i,j}$ resultiert für ein Kundensegment l zum Zeitpunkt t aus einer Unternehmensleistung i aus der Kategorie j mit $j \in \{BaF, LF, BeF\}$. Dabei wird -1 als absolute Unzufriedenheit, 1 als höchste Zufriedenheit und 0 als Bestätigung verstanden, bei der gerade das Konfirmationsniveau der Zufriedenheit erreicht ist. Die Zuordnung einer der drei Kano-Kategorien zu einer Leistung durch ein Kundensegment erfolgt unabhängig von der Zuordnung dieser Leistung durch andere Kundensegmente. Der ZB $s_{l,t,i,j}$ ergibt sich dann zeitabhängig als Funktion der Erwartungserfüllung $x_{l,t,i}$ und der Kategorie j :

$$(4) \quad s_{l,t,i,j} = f_j(x_{l,t,i}) = \begin{cases} f_{BaF}(x_{l,t,i}) & \forall j = BaF \\ f_{LF}(x_{l,t,i}) & \forall j = LF \\ f_{BeF}(x_{l,t,i}) & \forall j = BeF \end{cases}$$

Aus den bereits diskutierten Gründen müssen die modifizierten Funktionsverläufe zur Abbildung der Zusammenhänge von Erwartungserfüllung und ZB innerhalb der einzelnen Kategorien folgende Anforderungen erfüllen:

- (A 1) Leistungen innerhalb der BeF werden vom Kunden nicht gefordert. Daher gilt $e_{l,t,i} = 0$ bei den BeF.
- (A 2) Der ZB für Werte $x_{l,t,i} > 0$ in der Kategorie der BeF soll bis zum Maximalpunkt ($x = 1$) streng zunehmend wachsen, da angenommen wird, dass bei immer besserer Erwartungs(über)erfüllung bei den BeF auch die Werte des ZB zunehmend steigen.

Der Wertebereich für die ZB mit einer Unternehmensleistung aus den BeF ist somit auf $[0,1]$ festgelegt. Im Bereich nicht erfüllter Erwartungen ist die Funktion nicht definiert.

- (A 3) BaF führen auch bei einer Übererfüllung der Erwartungen ($x_{l,t,i} \geq 0$) nicht zu einem positiven ZB, sondern es wird maximal das Konfirmationsniveau der Zufriedenheit erreicht.
- (A 4) Eine Nichterfüllung der Erwartungen bei BaF führt sehr schnell, d. h. bereits bei schwacher negativer Diskonfirmation der Erwartungen, zu absoluter Unzufriedenheit mit dieser Leistung. [CoPL04]

Um diese Anforderungen zu erfüllen, setzen wir innerhalb der BaF die ZB für Werte $x_{l,t,i} \geq 0$ auf 0. Der Funktionsverlauf bei geringen negativen x -Werten sinkt sehr rasch vom Wert 0 auf den Minimalwert -1, auf dem es für deutlich negative x -Werte verbleibt. Als Wertebereich für die ZB mit einer Unternehmensleistung aus den BaF legen wir somit $[-1,0]$ fest.

- (A 5) Für LF wird von einem proportionalen Zusammenhang zwischen der Erwartungserfüllung $x_{l,t,i}$ und des ZB $s_{l,t,i,LF}$ ausgegangen. Die ZB bei Nicht-Erfüllung eines LF müssen dabei stets mindestens so groß sein wie die der BaF, die ZB bei Übererfüllung der Erwartungen allerdings niemals so groß wie die der BeF. Auch an den Endpunkten der Abszisse sollen die LF nicht zu extremen ZB führen, insgesamt sollte die Funktion also stets zwischen den beiden Funktionen der BaF und BeF verlaufen.
- (A 6) Eine Nichterfüllung der Erwartungen wiegt auf Grund der Prospect Theorie schwerer als ein Übertreffen derselben. [KaTv79; MiRB98]

Um diesen Anforderungen gerecht zu werden, sollte eine abschnittsweise lineare Funktion gewählt werden. [HaJF93] Da diese stets zwischen den Funktionen der BeF und BaF liegen muss,

die ihre Maximalwerte bei 1 bzw. -1 haben, und eine Nichterfüllung von Erwartungen schwerer wiegen soll als ein Übertreffen derselben, kann als Wertebereich für den ZB mit einer Unternehmensleistung aus den LF der Bereich $[\underline{a}, \bar{a}]$ mit $|\underline{a}| > |\bar{a}|$ und $\underline{a} \in (-1,0)$ bzw. $\bar{a} \in (0,1)$ festgelegt werden.

(A 7) Die Funktionen $f_j(x_{l,t,i})$ zur Bestimmung der Zufriedenheitsbeiträge sollen stetig und monoton sein.

Empirisch konnten bislang weder Sprungstellen noch nicht-monotone Kurvenverläufe in den Bewertungsfunktionen identifiziert werden. Die Unterstellung von stetigen und monotonen Funktionen erscheint daher gerechtfertigt. Als Weiterentwicklung von Abb. 1 zeigt Abb. 2 beispielhaft die Funktionsverläufe des modifizierten, auf die Kundenzufriedenheit übertragenen Kano-Modells.⁴

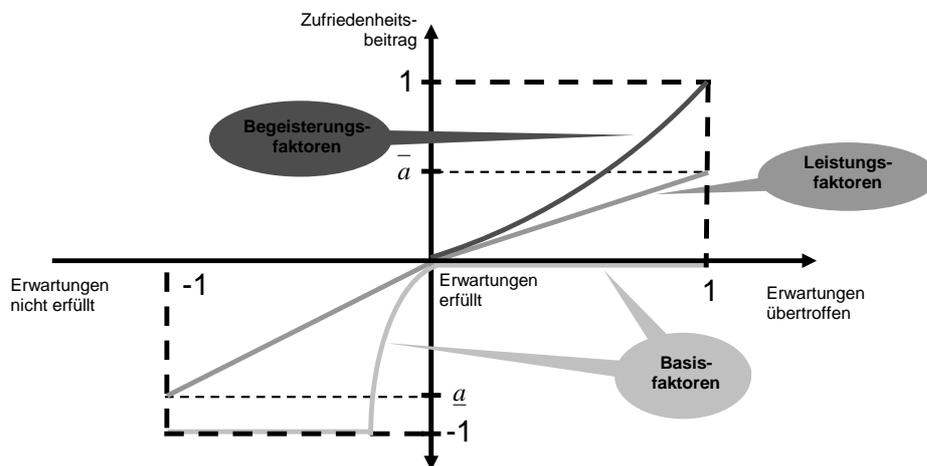


Abb. 2: Modifiziertes Kano-Modell der Kundenzufriedenheit

Die o. g. plausiblen Überlegungen an die Kurvenverläufe in den einzelnen Kategorien und den daraus formulierten Anforderungen (A 1) bis (A 7) schränken die Freiheitsgrade der Funktionen bzw. Funktionsklassen bereits erheblich ein. Auf Basis der Anforderungen und der nachfolgenden Spezifikation dürften sich Schätzfehler daher bei richtiger Kategorisierung in BaF, LF und BeF im Rahmen halten.⁵ Im Einklang mit den Anforderungen (A 1) bis (A 7) können die Funk-

⁴ Nicht angebotene BaF und LF gehen in die Berechnung mit -1 bzw. \underline{a} ein, da es sich um „Mussfaktoren“ bzw. vom Kunden verlangte Leistungen handelt. Es stellt sich bei den BaF und BeF also lediglich die Frage nach der (optimalen) Qualität der Umsetzung. Ansonsten könnten die aggregierten Zufriedenheitswerte für die BaF und LF einfach durch Weglassen potenziell gesteigert werden.

⁵ Zudem führt eine falsche Einteilung zwischen BaF und LF im Bereich der nicht erfüllten Erwartungen bzw. LF und BeF im Bereich der übererfüllten Erwartungen nur dann zu größeren Schätzfehlern, wenn zwischen Kundenerwartungs- und Leistungsniveau eine substantielle Differenz festzustellen ist (vgl. Abb. 2).

tionsverläufe für BeF, BaF und LF bspw. folgendermaßen dargestellt werden. Hierbei gehen wir zunächst von der Funktion der LF aus und bilden dann die beiden anderen Funktionen ab.

$$(5) \quad s_{l,t,i,LF} = \begin{cases} -\underline{a}x_{l,t,i} & \forall x_{l,t,i} \in [-1,0] \\ \bar{a}x_{l,t,i} & \forall x_{l,t,i} \in (0,1] \end{cases} \quad \text{und } \underline{a} \in (-1,0) \quad \text{und } \bar{a} \in (0,1) \quad \text{mit } |\underline{a}| > |\bar{a}|$$

$$(6) \quad s_{l,t,i,BeF} = \begin{cases} 0 & \forall x_{l,t,i} \in [-1,0] \\ \bar{a}x_{l,t,i} + b(x_{l,t,i})^2 & \forall x_{l,t,i} \in (0,1], \bar{a} \in (0,1) \text{ und } b \in (0,1-\bar{a}] \end{cases}$$

Damit der rechte Parabelast der Funktion der BeF nicht mit der Funktion der LF zusammenfällt (A 5), muss bei der linken Intervallgrenze von b die Null ausgeschlossen werden. Für die rechte Intervallgrenze gilt: Da die Funktion der BeF im Extremwert $x_{l,t,i} = 1$ den ZB $s_{l,t,i,BeF} = 1$ annehmen muss, gilt für die Parameter \bar{a} und b für $x_{l,t,i} > 0$ die folgende Bedingung:

$$(7) \quad b = 1 - \bar{a} \quad \forall x_{l,t,i} \in (0,1] \quad \text{und } j = BeF$$

Wie aus Formel (7) ersichtlich ist, ergibt sich für einen hohen Wert von \bar{a} ein niedriger Wert von b und umgekehrt. Bei hohem \bar{a} besitzt die lineare Funktion der LF eine hohe Steigung, gleichzeitig verläuft der rechte Parabel-Ast der Funktion der BeF mit geringer Steigung nahe oberhalb der linearen Funktion. Diese bedeutet, dass Kunden im Hinblick auf die ZB nahezu indifferent sind zwischen Unternehmensleistungen aus den LF oder den BeF.

Um für den Bereich der BaF die Anforderung (A 4) zu erfüllen, legen wir in dieser Kategorie einen Bereich zwischen 0 und $-d$ mit $d \in (0,1]$ fest, innerhalb dessen die ZB für leicht negative x -Werte auf den Minimalwert von -1 fallen. Links von $-d$ verbleibt der ZB dann auf dem Minimalwert. Berücksichtigt man auch die Anforderungen (A 3) und (A 7), so kann ein Funktionsverlauf für die BaF bspw. folgendermaßen dargestellt werden:

$$(8) \quad s_{l,t,i,BaF} = \begin{cases} -1 & \forall x_{l,t,i} < -d \\ -\underline{a}x_{l,t,i} - c(x_{l,t,i})^2 & \forall x_{l,t,i} \text{ für die gilt } -d \leq x_{l,t,i} < 0, \underline{a} \in (-1,0) \text{ und } c \in (0, \frac{1}{d^2} + \frac{\underline{a}}{d}] \\ 0 & \forall x_{l,t,i} \geq 0 \end{cases}$$

Damit auch hier die Funktion der BaF nicht mit der der LF zusammenfällt (A 5), muss die Null bei der linken Intervallgrenze von c ausgeschlossen sein. Die rechte Intervallgrenze ergibt sich aus der Überlegung, dass die Funktion an der Stelle $x_{l,t,i} = -d$ den ZB $s_{l,t,i,BaF} = -1$ annehmen soll. Für die Parameter \underline{a} , d und c gilt dann:

$$(9) \quad c = \frac{1}{d^2} + \frac{\underline{a}}{d} \quad \forall j = BaF$$

3.2 Aggregation der Zufriedenheitsbeiträge

Die mit den voranstehenden Funktionen ermittelbaren ZB für jede Unternehmensleistung sollen nun zu einer Gesamtzufriedenheit aggregiert werden. Diese Aggregation erfolgt in zwei Stufen: Zunächst werden die ZB innerhalb eines Kundensegments und einer Kategorie miteinander verknüpft und in einem weiteren Schritt die ZB der drei Kategorien zu einer Gesamtzufriedenheit pro Kundensegment verdichtet.

3.2.1 Aggregation der Zufriedenheitsbeiträge innerhalb einer Kategorie

Bei der Aggregation soll berücksichtigt werden, dass verschiedene Kundensegmente individuelle Präferenzen hinsichtlich ihrer Bedeutung bzgl. einzelner Unternehmensleistungen besitzen. Daher ergibt sich die folgende weitere Anforderung:

(A 8) Die vorangehend ermittelten ZB $s_{l,t,i,j}$ müssen gemäß ihrer Bedeutung für das Kundensegment gewichtet werden können. Die Bedeutung soll hierbei unabhängig von der Kategorie sein, der eine Leistung zugeordnet ist.

Dazu wird ein Gewichtungsfaktor $w_{l,t,i} \in [0,1]$ eingeführt, mit dessen Hilfe sich der gewichtete ZB $S_{l,t,i,j}$ ergibt:

$$(10) \quad S_{l,t,i,j} = s_{l,t,i,j} \cdot w_{l,t,i} \quad \text{mit} \quad \sum_i w_{l,t,i} = 1$$

Um die Ergebnisse bei Variationen von Gewichtungswerten vergleichbar machen zu können, besteht eine Notwendigkeit zur Normierung der Summe aller Werte $w_{l,t,i}$, bspw. auf den Wert 1. Durch diese Normierung wird sichergestellt, dass die Wertebereiche für die gewichteten ZB hierbei denen der ungewichteten ZB entsprechen.

Die gewichteten ZB einzelner Unternehmensleistungen müssen nun zu einem ZB pro Kategorie aggregiert werden. Wie diese Aggregation erfolgen soll, wird überraschender Weise in der Kundenzufriedenheitsforschung bisher kaum thematisiert. Vereinzelt wird zwar angenommen, dass die Kunden ihren Gesamteindruck aus den Teilzufriedenheiten mit einzelnen (Teil-) Leistungen zusammensetzen und diese nach ihrer Bedeutung gewichten, [Schü92, S. 175] verwendet wird jedoch stets nur ein einfaches, additives Modell. Dies ist bspw. in Kanos Originalmodell [BHMS96, S. 123] oder bei der Berechnung von Kundenzufriedenheitsindizes [Töpf99, S. 326ff.] der Fall. Dabei werden jedoch u. E. wichtige Aspekte bei der Untersuchung der Zufriedenheitsentstehung vernachlässigt, da ein additives Modell dadurch gekennzeichnet ist, dass Unzufriedenheit in einem Teilbereich durch positive Aspekte in anderen Bereichen ausgegli-

chen werden kann. Dies ist jedoch nicht immer realistisch, da extreme Unzufriedenheitswerte i. d. R. kaum ausgeglichen werden können: In einem solchen Fall ist eine kumulative Aggregation angemessen und eine Durchschnittsbildung inadäquat. Für eine Aggregation der Einzelwerte innerhalb einer Kategorie ergeben sich daher die folgenden Anforderungen:

- (A 9) Die ZB in den BaF und BeF müssen *akkumulativ* sein. Bei Hinzunahme eines positiven (bzw. negativen) ZB innerhalb der BeF (bzw. der BaF) müssen sich also auch die aggregierten Werte erhöhen (bzw. verringern).⁶
- (A 10) Innerhalb der LF soll ein Ausgleich von negativen und positiven Werten möglich sein, da hier einerseits keine extremen ZB zu finden sind, die nachhaltigen Einfluss auf den aggregierten ZB haben, und andererseits der Kunde sämtliche Produkte oder Dienstleistungen der LF erwartet und somit eine Nivellierung von negativen und positiven Werten plausibel erscheint. ZB von LF müssen sich daher aus einer Mittelwertbildung ergeben.

Einen möglichen Lösungsansatz für die akkumulative Aggregation der ZB innerhalb der BaF und BeF gemäß Anforderung (A 9) liefert der aus der Welt der wissensbasierten Systeme bekannte EMYCIN-Ansatz der probabilistischen Output-Aggregation. [BuWe92] Als Certainty-Ansatz wird er ursprünglich bei der Ermittlung von so genannten Certainty Factors für den Eintritt zweier (Regel-Output-)Fakten mit dem Ergebnis verwendet, dass die aggregierten Werte stets betragsmäßig höher sind als die Einzelwerte. Somit stellt er einen möglichen Lösungsansatz für die Aggregation von Werten aus den Extrembereichen der BaF und der BeF dar. In nicht-inkrementeller Schreibweise (im Gegensatz zu oben genannten Arbeiten) kann das Gesamtergebnis der aggregierten ZB pro Kategorie der BeF bzw. BaF als $\hat{S}_{l,t,BeF}$ bzw. $\hat{S}_{l,t,BaF}$ folgendermaßen formuliert werden:

$$(11) \quad \hat{S}_{l,t,BeF} = 1 - \prod_{i:j=BeF} (1 - S_{l,t,i,j})$$

$$(12) \quad \hat{S}_{l,t,BaF} = \prod_{i:j=BaF} (1 + S_{l,t,i,j}) - 1$$

$\hat{S}_{l,t,BeF}$ ist damit auch nach beliebig vielen Aggregationsschritten auf das Intervall [0;1] normiert, $\hat{S}_{l,t,BaF}$ auf [-1;0]. Zusätzlich ist die Aggregation streng monoton in dem Sinne, dass sich der aggregierte Wert $\hat{S}_{l,t,BeF}$ (bzw. $\hat{S}_{l,t,BaF}$) bei Hinzunahme eines positiven (bzw. negativen) ZB erhöht (bzw. erniedrigt), sodass die Anforderungen aus (A 9) erfüllt werden.

⁶ Beispiel zur Verdeutlichung: Ausgehend von einem bereits aggregierten ZB für die Kategorie der BeF von 0,8 soll bei Hinzunahme eines neuen, weiteren BeF mit einem ZB von 0,5 der aggregierte ZB steigen und nicht - wie bei Verwendung des Mittelwerts - sinken.

Anders verhält es sich für ZB innerhalb der Kategorie der LF. Hier sollen die einzelnen Leistungen durch eine Mittelwertbildung aggregiert werden. Hält man sich noch einmal vor Augen, dass der Kunde sämtliche zur Kategorie der LF gehörenden Produkte oder Dienstleistungen erwartet, erscheint die Verwendung des arithmetischen Mittels am intuitivsten.

$$(13) \quad \hat{S}_{l,t,LF} = \frac{1}{|LF|} \sum_{i:j=LF} S_{l,t,i,LF}$$

Dabei meint $|LF|$ die Kardinalität der Menge der Leistungsfaktoren mit $|LF| = \sum_{i:j=LF} 1$. Auch bei

dieser Aggregationsform bleibt auf Grund der Mittelwerteigenschaft der ursprüngliche Wertebereich von $[\underline{a}, \bar{a}]$ für $\hat{S}_{l,t,LF}$ erhalten. Damit haben wir nun pro Kundensegment, Zeitpunkt und Kategorie einen aggregierten ZB $\hat{S}_{l,t,j}$.

Positiv hervorzuheben ist die Eigenschaft der Reihenfolgeunabhängigkeit für beide vorgeschlagenen Verknüpfungsregeln (für EMYCIN vgl. [BuMW91, S. 213ff.]), denn i. d. R. wird im Rahmen eines zur Verfügung stehenden Budgets über mehrere umzusetzende Funktionalitäten aus eine Vielzahl von Projektideen zu einem Zeitpunkt zu entscheiden sein. Die Entscheidung über ein solches Maßnahmenprogramm soll dabei unabhängig davon sein, in welcher Reihenfolge die Funktionalitäten bei der Bestimmung des ZB einer Kategorie Berücksichtigung finden. Gerade bei zum Entscheidungszeitpunkt noch unklaren möglichen *Projektabhängigkeiten* ist die Vermeidung von Pfadabhängigkeiten bei der *Bewertung* ein nicht zu unterschätzender Vorteil.

3.2.2 Aggregation zu einem Gesamtzufriedenheitswert

Nachdem die ZB innerhalb jeder Kategorie aggregiert wurden, muss nun aus diesen eine Gesamtzufriedenheit ermittelt werden. Folgende Anforderungen werden hierbei an diese Aggregation gestellt:

- (A 11) Basierend auf der Annahme, dass ein ohnehin schon zufriedener Kunde geneigt ist, Wahrnehmungen eher als positiv einzuordnen, um kognitive Dissonanzen zu vermeiden, [Schü92, S. 171] sollte in der gewählten Aggregationsregel eine Veränderung des ZB in einer Kategorie unterschiedliche Auswirkungen auf die Gesamtzufriedenheit haben können, je nachdem, wie hoch die ZB in den anderen Kategorien sind. Dies bedeutet, dass bei großen ZB in zwei Kategorien eine Erhöhung des ZB in der dritten Kategorie

zu einer stärkeren Erhöhung der Gesamtzufriedenheit führt, als es der Fall wäre, wenn in den beiden anderen Kategorien Unzufriedenheit vorherrschen würde.

(A 12) Der Wert der Kategorie der BaF muss ein k.o.-Kriterium darstellen können, d. h., wenn das Kundensegment in dieser Kategorie völlig unzufrieden ist ($\hat{S}_{l,t,BaF} = -1$), darf dies nicht durch gute Werte in den LF oder BeF ausgeglichen werden können, sondern muss insgesamt zu Unzufriedenheit führen. [CoPL04]

(A 13) Sofern $\hat{S}_{l,t,BaF} > -1$ sollen sich die aggregierten ZB der drei Kategorien ausgleichen können.

(A 14) Die ZB der einzelnen Kategorien sollen mit unterschiedlichen Gewichtungen in die Gesamtzufriedenheit eingehen können, sodass gewährleistet werden kann, dass bspw. BeF stärker gewichtet werden können als BaF.

Auf Grund dieser Anforderungen ist eine rein additive Verknüpfung (z. B. mit dem arithmetischen Mittel) wiederum nicht möglich, da hier weder die Anforderung (A 11) noch (A 12) erfüllt werden kann. Zur Entwicklung einer anforderungskonformen „Customer Satisfaction- (CS-)Funktion“ wird auf Grund ihrer Multiplikatивität und Konkavität von einer Cobb-Douglas-Funktion ausgegangen. Diese enthält die drei Argumente y_{BaF} , y_{LF} und y_{BeF} (für die drei Kategorien) sowie die (die Gewichtung der Argumente ermöglichenden) Exponenten α_{BaF} , α_{LF} und α_{BeF} und kann folgendermaßen dargestellt werden:

$$(14) \quad f(y_{BaF}, y_{LF}, y_{BeF}) = y_{BaF}^{\alpha_{BaF}} \cdot y_{LF}^{\alpha_{LF}} \cdot y_{BeF}^{\alpha_{BeF}}$$

mit $\alpha_{BaF}, \alpha_{LF}, \alpha_{BeF} \in (0,1)$ und $\alpha_{BaF} + \alpha_{LF} + \alpha_{BeF} = 1$

Bei der Transformation der Cobb-Douglas-Funktion auf die Kundenzufriedenheit und unser bisheriges Modell muss jedoch Folgendes beachtet werden: Zunächst stellen die drei ZB pro Kategorie die jeweiligen Argumente der Funktion dar. Da in eine Cobb-Douglas-Funktion jedoch keine negativen Argumente eingehen dürfen, müssen die ZB zunächst von ihrem Gesamtwertebereich $[-1,1]$ auf einen positiven Wertebereich transformiert werden. Normieren wir daher die einzelnen ZB $\hat{S}_{l,t,j}$ auf das Intervall $[0,1]$, so muss folgende lineare Transformationsregel angewandt werden.

$$(15) \quad \bar{S}_{l,t,BaF} = (\hat{S}_{l,t,BaF} + 1)/2, \quad \text{wobei } \hat{S}_{l,t,BaF} \in [-1,0] \Rightarrow \bar{S}_{l,t,BaF} \in [0,0,5]$$

$$(16) \quad \bar{S}_{l,t,LF} = (\hat{S}_{l,t,LF} + 1)/2, \quad \text{wobei } \hat{S}_{l,t,LF} \in [\underline{a}, \bar{a}] \Rightarrow \bar{S}_{l,t,LF} \in [(\underline{a} + 1)/2, (\bar{a} + 1)/2]$$

$$(17) \quad \bar{S}_{BeF} = (S_{BeF} + 1)/2, \quad \text{wobei } S_{BeF} \in [0,1] \Rightarrow \bar{S}_{BeF} \in [0,5,1]$$

Als transformierte Cobb-Douglas-Funktion ergibt sich damit zunächst:

$$(18) \quad f(\bar{S}_{l,t,BaF}, \bar{S}_{l,t,LF}, \bar{S}_{l,t,BeF}) = \bar{S}_{l,t,BaF}^{\alpha_{BaF}} \cdot \bar{S}_{l,t,LF}^{\alpha_{LF}} \cdot \bar{S}_{l,t,BeF}^{\alpha_{BeF}}$$

$$\text{mit } f(\bar{S}_{l,t,BaF}, \bar{S}_{l,t,LF}, \bar{S}_{l,t,BeF}) \in \left[0, 0,5^{\alpha_{BaF}} \cdot \left(\frac{\bar{a} + 1}{2}\right)^{\alpha_{LF}}\right]$$

Aus den Grundüberlegungen des qualitativen Modells nach Kano abgeleitet (vgl. auch (A 14)), sollen die BeF stets den größten Einfluss auf die Gesamtzufriedenheit haben. [Saue00, S. 179] Daneben gehen die LF mit einem höheren Wert in die Gesamtzufriedenheit ein als die BaF. Um dies gewährleisten zu können, muss der Exponent der BeF stets als größter, der Exponent der BaF als kleinster Wert gewählt werden.

$$(19) \quad 0 < \alpha_{BaF} < \alpha_{LF} < \alpha_{BeF} < 1$$

Damit die transformierte Cobb-Douglas-Funktion für eine mögliche Ermittlung der monetären Auswirkungen der Gesamtzufriedenheit berücksichtigt werden kann, sollte ihr Wertebereich nun ebenfalls auf [0;1] normiert werden. Dies geschieht, indem wir die Funktion durch ihren Höchstwert $\left(0,5^{\alpha_{BaF}} \cdot \left(\frac{\bar{a} + 1}{2}\right)^{\alpha_{LF}}\right)$ dividieren. Es ergibt sich somit für die Gesamtzufriedenheitsfunktion CS für ein homogenes Kundensegment:

$$(20) \quad CS_{l,t} = \frac{\bar{S}_{l,t,BaF}^{\alpha_{BaF}} \cdot \bar{S}_{l,t,LF}^{\alpha_{LF}} \cdot \bar{S}_{l,t,BeF}^{\alpha_{BeF}}}{0,5^{\alpha_{BaF}} \cdot \left(\frac{\bar{a} + 1}{2}\right)^{\alpha_{LF}}} \Rightarrow CS_{l,t} \in [0,1]$$

Man beachte, dass diese Funktion insbesondere auch die k.o.-Eigenschaft (A 12) für die Kategorie der BaF erfüllt, diese aber nicht für die beiden anderen Kategorien gelten. Mit Hilfe verschiedener, den jeweiligen Anforderungen gerecht werdender Aggregationsformen wurden also die ZB bzgl. einzelner Unternehmensleistungen zu ZB der jeweiligen Kategorien zusammengefasst, die wiederum im Anschluss daran zu einer Gesamtzufriedenheit aggregiert wurden. Der auf diesem Wege ermittelte Wert für die Gesamtzufriedenheit bzgl. aller Unternehmensleistungen je Kundensegment kann nun als Basis sowohl für eine empirische Überprüfung als auch für weiterführende Aggregationen verwendet werden, bspw. dann wenn segmentübergreifende Investitionen analysiert werden sollen.

Kritisch sei angemerkt, dass nach unserer Projekterfahrungen Investitionen in die Kundenzufriedenheit nur dann wirtschaftlich erfolgreich sind, wenn eine hohe bis sehr hohe Kundenzufriedenheitssteigerung erzielt werden kann. Denn auf Grund der Indifferenzzone der Kundenzu-

friedenheit führt nur ein starker Anstieg der Kundenzufriedenheit zu wahrnehmbaren Verhaltensänderungen (konsistent mit [HeJo99, S. 595]).

4 Fazit und Ausblick

Das in diesem Beitrag vorgestellte Modell wurde entwickelt, um in der Praxis die Kundenzufriedenheit besser und vor allem konsistenter quantifizieren zu können. Den Herausforderungen einer Analyse von Kundenzufriedenheitsquellen und deren Modellierung auf Basis des Kano-Modells ist bisher in der Literatur nach unserer Kenntnis noch nicht hinreichend Rechnung getragen worden. In der praktischen Anwendung erlaubt das Modell die konsistente Berücksichtigung von Faktoren, die bislang als schwierig quantifizierbar galten und bei Entscheidungen daher oft vernachlässigt wurden. Mit den in diesem Beitrag begründeten Funktionsklassen kann das vorgestellte Modell dem Anwender daher eine gewisse Hilfestellung geben, um die dargestellten Zusammenhänge quantitativ zu erfassen und zumindest grob abzuschätzen.

Für die Zukunft besteht weiterer Forschungsbedarf, das dargestellte Modell insgesamt zu erweitern und zu verfeinern. Neben der Überprüfung der Validität und Reliabilität der vorgeschlagenen Methode, sollte das Augenmerk auf die Schätzung von (zumindest) branchenspezifischen Transformationsfunktionen von Kundenzufriedenheit (und weiteren Einflussfaktoren) in Loyalität und Unternehmenserfolg gelegt werden. Die Entwicklung und Implementierung von IT-Controlling-Systemen stellen weitere wichtige Instrumente zur gezielten Informationsbeschaffung über die Verhaltensweisen des Kunden dar. Durch den Einsatz derartiger Anwendungen lässt sich z. B. die Nutzung von verschiedenen Funktionalitäten einer Webseite oder eines mobilen Portals durch die Kunden messen, im Zeitablauf vergleichbar machen und somit das Kundenverhalten empirisch überprüfen. Im Rahmen der Bewertung von Maßnahmenprogrammen ist insbesondere die Bestimmung eines (in langfristiger Perspektive) kapitalwertoptimalen Niveaus der Kundenzufriedenheit. Auch eine allgemeine Aussage darüber, ob eine Investition in BaF, LF oder BeF am aussichtsreichsten ist bzw. in welcher Reihenfolge die entsprechenden Funktionalitäten generell zur Verfügung gestellt werden sollten, stellen theorie- und praxisrelevante Themen dar, denen sich die zukünftige Forschung widmen sollte.

Literatur

- [BHMS96] *Bailom, F.; Hinterhuber, H.J.; Matzler, K.; Sauerwein, E.*: Das Kano-Modell der Kundenzufriedenheit. In: *Marketing ZFP* 18 (1996) 2, S. 117-126.
- [BuMW91] *Buhl, H.U.; Massler, T.; Weinhardt, C.*: EAG: Ein Erweiterungsansatz zur Darstellung und Verarbeitung unsicheren Wissens in wissensbasierten Systemen. In: *Wirtschaftsinformatik* 33 (1991) 3, S. 213-218.
- [BuWe92] *Buhl, H.U.; Weinhardt, C.*: EAG: Ein Verfahren zur Gewissheitsverarbeitung in wissensbasierten Systemen. In: *Informationstechnik it* (1992) 5, S. 296-306.
- [CoPL04] *Conklin, M.; Powaga, K.; Lipovetsky, S.*: Customer satisfaction analysis: Identification of key drivers. In: *EJOR* 154 (2004) 3, S. 819-827.
- [HaJF93] *Hardie, B.; Johnson, E.; Fader, P.*: Modeling Loss Aversion and Reference Dependence Effects on Brand Choice. In: *Marketing Science* 12 (1993) 4, S. 278-294.
- [HvNH98] *Herrmann, A.; von Nitzsch, R.; Huber, F.*: Referenzpunktbezogenheit, Verlustaversion und abnehmende Sensitivität bei Kundenzufriedenheitsurteilen. In: *ZfB* 68 (1998) 11, S. 1225-1244.
- [HeJo99] *Herrmann, A.; Johnson, M.*: Die Kundenzufriedenheit als Bestimmungsfaktor der Kundenbindung. In: *ZfbF* 51 (1999) 6, S. 579-598.
- [HoSt01] *Homburg, C.; Stock, R.*: Theoretische Perspektiven zur Kundenzufriedenheit. In: *Homburg, C. (Hrsg.): Kundenzufriedenheit – Konzepte, Methoden, Erfahrungen*, 4. Aufl., Wiesbaden, 2001, S. 19-50.
- [KaTv79] *Kahneman, D.; Tversky, A.*: Prospect Theory. In: *Econometrica* 24 (1979) 2, S. 263-291.
- [KTST84] *Kano, N.; Tsuji, S.; Serak, N.; Takahashi, F.*: Attractive Quality and Must-be Quality. In: *The Journal of the Japanese Society for Quality Control* (1984) 4, S. 39-48.

- [KaHu01] *Kaapke, A.; Hudetz, K.:* Der Einsatz des Kano-Modells zur Ermittlung von Indikatoren der Kundenzufriedenheit. In: *Müller-Hagedorn, L.* (Hrsg.): Kundenbindung im Handel, 2. Aufl., Frankfurt a.M., 2001, S. 123-146.
- [Kraf99] *Krafft, M.:* Der Kunde im Fokus: Kundennähe, Kundenzufriedenheit, Kundenbindung – und Kundenwert?. In: *DBW 59* (1999) 4, S. 511-530.
- [MaSt00] *Matzler, K.; Stahl, H.:* Kundenzufriedenheit und Unternehmenswertsteigerung. In: *DBW 60* (2000) 5, S. 626-639.
- [MiRB98] *Mittal, V.; Ross, W.; Baldasare, P.:* The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. In: *Journal of Marketing*, 62 (1998) 1, S. 33-48.
- [Oliv97] *Oliver, R.:* Satisfaction: A Behavioral Perspective on the Consumer. New York, 1997.
- [Saue09] *Sauerwein, E.:* Das Kano-Modell der Kundenzufriedenheit: Reliabilität und Validität einer Methode zur Klassifizierung von Produkteigenschaften, Wiesbaden, 2000.
- [Schü92] *Schütze, R.:* Kundenzufriedenheit – After Sales Marketing auf industriellen Märkten, Wiesbaden, 1992.
- [Tont00] *Tontini, G.:* Identification of customer attractive and must-be requirements using a modified Kano's method: Guidelines and case study. In: *Annual Quality Congress Proceedings*, 2000, S. 728-734.
- [Töpf99] *Töpfer, A.:* Kundenzufriedenheit messen und steigern, 2. Aufl., Neuwied, 1999.
- [WiLe03] *Wirtz, J.; Lee, M.C.:* An Examination of the Quality and Context-Specific Applicability of Commonly Used Customer Satisfaction Measures. In: *Journal of Service Research* 5 (2003) 4, S. 345-355.

Einführung in den Track

eHealth / Informationsmanagement im Gesundheitswesen

Prof. Dr. Hans Czap

Universität Trier

Prof. Dr. Dr. Detlef Ruland

McKinsey & Company, Inc.

Dr. Jörg Haas

HWB AG

The Health-Care Domain shows still features that are typical for the information systems (IS) domain in industry of the 1970ies. Also, there are standards, like HL7, hospital IS in general are fragmented lacking integration. Since a couple of years process-oriented views, like Clinical pathways and evidence based medicine are widely discussed. Nevertheless, their deployment still remains in the stage of requirement definition and development. Similar deficits are reported for planning and control on the operational level, not to mention planning and control on the more strategic level.

Besides the old problems, which are widely known, some new problems arose stemming from recent developments. For example, the integration of different actors of the health care domain establishes a major issue of future IS. Electronic Patient records and their exchange, the introduction of Patient Data Cards, and the problem how models of integrated care and disease management should look like are some of the topics discussed intensively. Mobile devices have the potential to overcome traditional organizational structures in hospitals and in the healthcare sector. Their integration in connection with autonomous subsystems, like intelligent agents, will allow for better care and medical attendance in standard situations as well as in emergency cases.

This track aims to promote research and scientific exchange related to still unsolved problems of Health Information Systems and to bring together practitioners, scientists, and researchers.

Programmkomitee:

Prof. Dr. Detlev Ruland, McKinsey & Company, Inc.

Florian Schaudel, McKinsey & Company, Inc.

Prof. Dr. Hans Czap, Universität Trier

Prof. Dr. Helmut Krcmar, Technische Universität München

Dr. Ingo Schellhammer, McKinsey & Company, Inc.

Dr. Jörg Haas, HWB AG

Maren Rowold, McKinsey & Company, Inc.

Matthias Behrens, navisco

Dr. Michael Dahlweid, GWI

Prof. Dr. Ralph Bergmann, Universität Trier

Prof. Dr. Rainer Unland, Universität Duisburg-Essen

Elektronische Intermediation im Gesundheitswesen – das Beispiel Schweiz

Konrad Walser; Thomas Myrach

Institut für Wirtschaftsinformatik, Abteilung Information Management
der Universität Bern
Engehaldenstr. 8
3012 Bern/Schweiz
{konrad.walser | thomas.myrach}@iwi.unibe.ch

Abstract

In diesem Beitrag werden Geschäftsmodelle von elektronischen Intermediären analysiert, die im Gesundheitswesen der Schweiz tätig sind. Dabei steht der Rechnungsdatenaustausch zwischen Leistungserbringern und Kostenträgern im Vordergrund. Anhand der Fallbeispiele wird einerseits die Motivation zur Intermediation ersichtlich, andererseits wird der Einfluss unterschiedlicher Trägerschaften auf die Einstellung zur Intermediation deutlich. Sowohl aus Sicht der Leistungserbringer als auch der Kostenträger werden unterschiedliche Nutzungsstrategien bezüglich der Intermediäre dargestellt. Über die Darstellung der Spezifika der elektronischen Intermediation im Gesundheitswesen hinaus werden dadurch auch generelle Einsichten in die Positionierung von Intermediären möglich.

1 Einleitung

Intermediation wird als Vermittlungsdienst zwischen Marktpartnern definiert. Bei vielen Geschäftstransaktionen kann die Abwicklung von Leistungsaustauschen zwischen Marktakteuren von Intermediären unterstützt werden. Intermediäre können vielfältige betriebswirtschaftliche und technische Funktionalitäten bzw. Leistungen bereitstellen. Im E-Business werden Mittler zwischen Marktpartnern auf elektronischer Ebene auch als Cybermediäre [SBS95] oder Hypermediäre [Har02] bezeichnet. Diese üben ihre Vermittlungsfunktion über elektronische Plattformen und Medien aus. Elektronische Intermediäre treten in verschiedenen Branchen und Geschäftsbereichen auf. Teilweise stellen

sie ein Pendant zu etablierten Offline-Intermediären dar. Daneben werden auch neue Intermediationsformen realisiert, die so in der Offline-Welt nicht existieren. Ein Beispiel dafür sind Value Added Networks (VAN [Har02]), über die der elektronische Datenaustausch (EDI) zwischen Geschäftspartnern ermöglicht wird [HaNe05: 736].

Bei der Transformation des Gesundheitswesens richtung eHealth können Intermediäre eine wesentliche Rolle spielen [WE05]. Geschäftsmodelle von Intermediären im Gesundheitswesen existieren in der Schweiz wie im Ausland teilweise schon länger. Jedoch sind aufgrund neuer technologischer Entwicklungen (Internetverbreitung, EAI-Technologie und B2B-Integration) neue Geschäftsmodelle von Interesse. Zudem können je nach (staatlicher oder privater) Organisation des Gesundheitswesens integrative Infrastrukturen erforderlich sein, da sich Partner im Gesundheitswesen auf ihre Kernkompetenzen konzentrieren wollen.

Der vorliegende Beitrag erläutert die Rahmenbedingungen für die Integration anhand des schweizerischen Gesundheitswesens. Weiter werden Kosten-Nutzen-Überlegungen zur Intermediation angestellt. Die Situation im Schweizer Gesundheitswesen wird anhand einer Geschäftsmodellanalyse der jeweiligen Intermediäre und entsprechender Intermediärsstrategien zwischen Leistungserbringern und Kostenträgern illustriert.

2 Integration im Gesundheitswesen

Die entwickelten Industrieländer sind trotz teilweise sehr unterschiedlicher Ausprägungen ihres Gesundheitswesens mit grundsätzlich ähnlichen Problemen konfrontiert. Steigende Anforderungen an medizinische Leistungen und ein massives Anwachsen entsprechender Leistungskataloge, eine ungünstige Altersstruktur der Population, aber auch die Überlagerung von staatlichen Reglementierungen und Marktmechanismen tragen zu einer überproportionalen Steigerung der Ausgaben im Gesundheitswesen bei, so auch in der Schweiz.

Angesichts dieser Entwicklung, die insbesondere das Funktionieren der kollektiven Sicherungssysteme bedroht, stehen Politikern und anderen Akteuren im Gesundheitswesen grundsätzlich drei Maßnahmen zur Verfügung: Zum einen können die Beiträge zu den kollektiven Sicherungssystemen entsprechend des steigenden Finanzierungsbedarfs angehoben werden; weiterhin besteht die Möglichkeit, zunehmend Leistungen ganz oder teilweise aus dem Leistungsumfang der Versicherungen auszugliedern und sie der privaten Finanzierung anheim zu stellen; schließlich kann mit verschiedenen Maßnahmen versucht werden, die Kosten-

steigerung im Gesundheitswesen einzudämmen. Eine der Maßnahmen, um Einsparungen im Gesundheitswesen zu erreichen, besteht in der Verbesserung der inner- und zwischenbetrieblichen Prozesse insbesondere im administrativen Bereich. So verspricht man sich durch besser integrierte Informationsaustausche zwischen den Beteiligten der Gesundheitswertschöpfungskette Sparpotentiale. Der elektronische Datenaustausch wurde mit dem Tarif Médical (TarMed) per 1.1.2006 vereinbart [Ae04]; dies gemäß Artikel 5, Absatz 5 des Rahmenvertrags zum Krankenversicherungsgesetz (KVG) und zum Unfallversicherungsgesetz (UVG). Er ist jedoch aus verschiedenen Gründen heute noch nicht vollständig umgesetzt. Beim TarMed handelt es sich um den gesamtschweizerischer Tarif für die Abrechnung von ambulanten Arztrechnungen, der zwischen ambulanten Leistungserbringern oder Ärzten und Versicherungen vereinbart wurde. Der TarMed gilt entsprechend dem UVG ab dem 1.5.2003 und entsprechend dem KVG ab dem 1.1.2004.

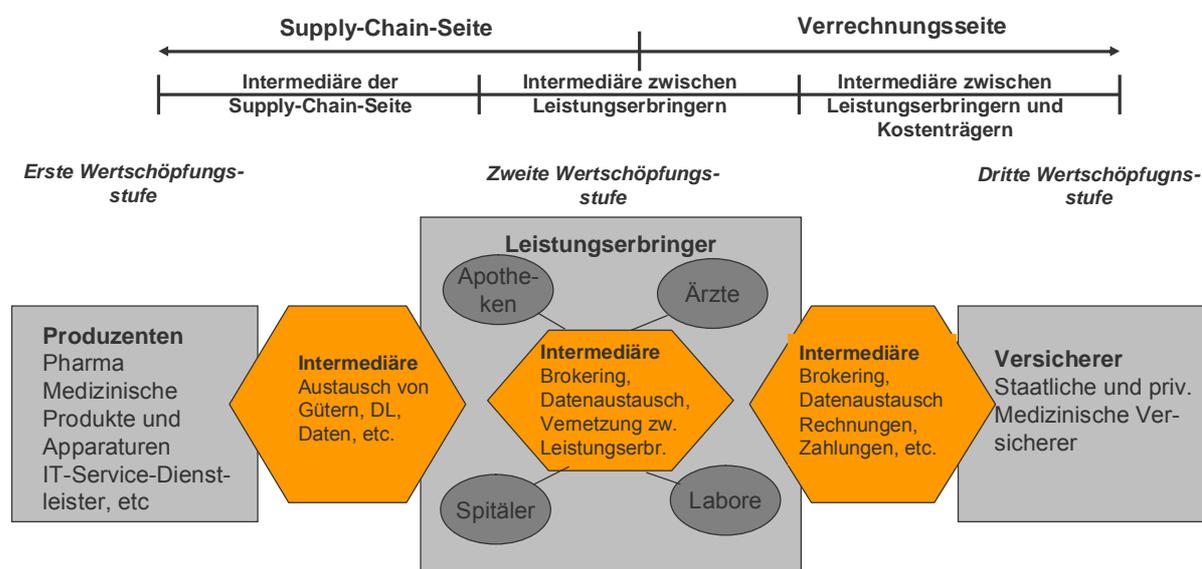


Abbildung 1: Wertschöpfungskette des Gesundheitswesens.

In einer groben Betrachtung können drei Akteursgruppen oder Wertschöpfungsstufen in der Gesundheitswertkette unterschieden werden (Abbildung 1) [WE05]. Auf einer ersten Wertschöpfungsstufe sind Produzenten und Dienstleister von Vorleistungen für die eigentliche Leistungserstellung angesiedelt. Sie gehören den Bereichen Pharma, medizinisches Zubehör, medizinische Apparate und Einrichtungen an. Auf einer zweiten Stufe sind die eigentlichen Leistungserbringer medizinischer Leistungen zu positionieren. Dies sind Ärzte, Spitäler, Kliniken, Apotheken, Labore, etc. Auf einer dritten Stufe können je nach Betrachtungsfokus Patienten oder Kunden als Leistungsnehmer oder Krankenkassen als Kostenträger aufgeführt

werden. Zu den Kostenträgern gehören (private und staatliche) Unfall- und Krankenversicherer, d.h. kollektive Absicherungssysteme, über welche eine direkte oder indirekte finanzielle Abgeltung der durch Patienten in Anspruch genommenen Leistungen erfolgt. Auf Basis der Wertschöpfungsstufen können unterschiedliche Bereiche der Intermediation und verschiedene Intermediärstypen unterschieden werden: Intermediäre der Supply-Chain-Seite, Intermediäre zwischen den Leistungserbringern sowie Intermediäre zwischen Leistungserbringern und Kostenträgern [WE05]. Auf Letztere wird in diesem Beitrag fokussiert.

Der Gesundheitswesen der Schweiz ist geprägt durch eine große Zahl verschiedener Leistungserbringer. So existieren ca. 700 Spitäler und ähnliche Institutionen, ca. 19'000 Ärzte, ca. 1'900 Apotheken sowie ca. 300 Labore. Bei den Kostenträgern existieren ca. 100 private Versicherer und die Schweizerische Unfall-Versicherungsanstalt (SUVA) als staatlicher Versicherer. Zudem kommt erschwerend hinzu, dass seitens der Leistungserbringer (Spitäler, Ärzte, Labore, Apotheken) und der Kostenträger jeweils eine Reihe von Anbietern ERP-ähnlicher Lösungen vorhanden ist. Nebst der Anzahl an Leistungserbringern und Kostenträgern erhöht das einerseits die Integrationskomplexität und andererseits den potentiellen Nutzen der Intermediation.

Bei den genannten Zahlen an Leistungserbringern und Kostenträgern ergibt sich eine sehr hohe Zahl an (potentiellen) Kommunikations- und Transaktionsbeziehungen und daraus abgeleiteten technischen Schnittstellen und Prozessintegrationen, die ohne Intermediäre durch die Beteiligten selbst initialisiert, gepflegt und administrativ verwaltet werden müssten. Zudem ist eine zwischenbetriebliche Integration zwischen Kostenträgern und Leistungserbringern (mit oder ohne Intermediär) im Gesundheitswesen grundsätzlich dann lohnend, wenn dadurch eine Kostenreduktion für administrative oder (redundanter) medizinische Prozesse oder eine Wirkungs- oder Nutzensteigerung für die beteiligten Partner erreicht wird. Durch die erforderliche Integration zwischen den verschiedensten Partnern (mit oder ohne Intermediär) resultieren die folgenden ökonomischen Konsequenzen. Auf der Kostenseite resultieren hohe Transaktionskosten aufgrund der Verhandlungen mit verschiedensten Partnern (Versicherer, Lieferanten, Apotheken). Kostenseitig wirken sich zudem die Aufwände für die betriebliche und technische Integration mit den verschiedensten Partnern aus, welche ohne Intermediär mit den relevanten Geschäftspartnern je separat aufzubauen sind. Bei den Partnern sind dabei unterschiedliche Daten- und Prozessstypen sowie unterschiedliche Informationssysteme vorhanden. In Abhängigkeit von den eingesetzten Informationssystemen bestehen auch

unterschiedliche Grade der Rechnungsdatenintegration. Weiter fallen Kosten für spezifische Soft- und Hardware an, welche eine einfache Adaption und Integration verschiedenster Partner im Gesundheitswesen erlaubt. Zudem führt der Einsatz unterschiedlicher Soft- und Hardware seitens der Nutzer der Intermediärsdienstleistungen zu Zusatzaufwänden für Schnittstellen sowie Soft- und Hardware-Anpassungen, etc. Dies kann gar Auswirkungen auf die Integration weiterer (interner) Informationssysteme bei den Partnern haben.

3 Allgemeine Überlegungen zur Intermediation im Gesundheitswesen

Für die Etablierung einer zwischenbetrieblichen Integration, wie sie etwa für den elektronischen Austausch von Rechnungsdaten erforderlich ist, müssen unter Umständen erhebliche Investitionen in die technische Infrastruktur getätigt werden. Ein Kostentreiber ist dabei die Anzahl der Partnerbeziehungen, die dafür aufzubauen sind. Es ist damit zu rechnen, dass für jede aufzubauende Beziehung spezifische Investitionskosten anfallen. Zur Reduktion der Aufwände, die mit der technischen und organisatorischen Implementierung zwischenbetrieblicher Integration anfallen, sind hauptsächlich zwei Hebel denkbar: Zum einen die Standardisierung des Datenaustausches und der betriebswirtschaftlichen und technischen Schnittstellen. Dies senkt die Initialkosten für den Aufbau der technischen Infrastruktur und die wiederkehrenden Kosten für den Verbindungsaufbau zu Partnern. Zum anderen kann durch das Einschalten eines Intermediärs, z.B. auf Basis eines elektronischen Hubs, die Zahl der aufzubauenden Partnerbeziehungen reduziert werden.

Im schweizerischen Gesundheitswesen wurde vom www.forum-datenaustausch.ch eine Standardisierung des Datenaustauschs zwischen Leistungserbringern und Kostenträgern initiiert. Dieses stellt für das schweizerische Gesundheitswesen auf Basis des XML-Standards *Templates* für unterschiedliche Transaktionstypen des elektronischen Datenaustausches bereit. Das Ziel dieser Bemühungen ist es, die Kosten des elektronischen Datenaustausches zu senken, weil ein entsprechender Standard die individuelle Anpassung der Datenformate für eine Integration in andere Informationssysteme aufhebt oder nicht mehr notwendig macht. Jedoch ist es innerhalb des gesamten Gesundheitswesens schwierig, entsprechende Standards durchzusetzen, an die sich alle Beteiligten halten, weil dadurch unter Umständen kostenintensive Anpassungen der Informationssysteme der Beteiligten erforderlich sind. Damit würden die Transaktionskosten ohne Intermediär hoch bleiben. Ausgehend von den Schwierigkeiten der

Standardisierung des Datenaustausches werden zudem Geschäftsmodelle von Intermediären interessant, welche Datentransformationen unterschiedlicher Art übernehmen und unter Umständen auch innerbetriebliche Datentransformationen (etwa in Spitälern) lösen können.

Unter der Annahme, dass Beziehungsaufbau und -pflege mit entsprechenden Transaktionskosten belastet sind [PiWi01], wird das Kosteneinsparungspotential durch Intermediation plausibel, wobei die potentiellen Einsparungen von der Größe des Netzwerks abhängen. Während die Anzahl der Beziehungen ohne Intermediation mit der Zahl der Partner geometrisch wächst, erfolgt dieses Wachstum mit Intermediation arithmetisch. Die auf relativ einfachen Modellannahmen beruhende Analyse eines Beziehungsnetzwerks für die zwischenbetriebliche Integration zeigt, dass eine Intermediation in Anlehnung an *Metcalf's Law* prinzipiell umso nützlicher ist, je mehr Partner am Netzwerk partizipieren [SV98: 184]. Dabei wird allerdings eine übergeordnete und auf das gesamte Netzwerk bezogene Kosten-Nutzen-Betrachtung vorgenommen. Bei den Akteuren im Netzwerk handelt es sich jedoch typischerweise um eigenständige Einheiten, die opportunistische Interessen verfolgen. Dies gilt prinzipiell auch für einen eingeschalteten Intermediär, der sich seine Wertschöpfung von den angeschlossenen Partnern entgelten lassen wird. Dabei hängt es vom Erlösmodell ab, welche (Arten von) Belastungen die jeweiligen Akteure tragen müssen. Durch das Entgelten der Leistungen wird den Partnern zumindest ein Teil der wirtschaftlichen Einsparungen durch die Intermediation wieder genommen. Darauf basiert das finanzielle Geschäftsmodell eines Intermediärs [PiWi01] [Zbo96] [Har02].

Eine besondere Qualität erhält ein Beziehungsgeflecht unter Einbezug eines Intermediärs durch die dabei entstehenden Abhängigkeiten. Diese können unter Umständen durch einen opportunistisch handelnden Partner ausgenutzt werden. Allein die Furcht vor einem derartigen Verhalten kann eine erfolgreiche Intermediation verhindern. Deshalb spielen Vertrauen, vertrauensbildende Maßnahmen und Absicherungen gegen Missbrauch aus theoretischer und praktischer Sicht eine nicht unerhebliche Rolle bei der Intermediärsetablierung. In diesem Zusammenhang dürfte auch das Betreibermodell oder die Trägerschaft des Intermediärs eine wesentliche Rolle spielen. Ein unabhängiger Intermediär ist möglicherweise in einer anderen Situation als abhängige Intermediäre, welche einer Interessensgruppe im Netzwerk nahe stehen.

4 Geschäftsmodellanalyse dreier Intermediäre

Der Markt für Intermediäre im schweizerischen Gesundheitswesen ist je nach Intermediärskategorie verhältnismäßig jung, bezüglich Preisgestaltung und Wettbewerbssituation stark in Bewegung und teilweise restriktiv in der Herausgabe von Preis-, Kosten- und Umsatzinformationen. Der entsprechende Markt ist determiniert durch das Wechselspiel von freiem Markt, gesetzlichen Regelungen und politischen Interessen der beteiligten Parteien der Wertschöpfungskette. Der Markt für Intermediäre ist weiter heterogen aufgrund der unterschiedlichen Dienstleistungen sowie Leistungsbündel und -abgeltungen, aber auch aufgrund der verschiedenen Eignermodelle. Es existieren, wie zu zeigen ist, unterschiedliche Geschäftsmodelle der Intermediation.

4.1 Überblick über Geschäftsmodelle

Nachfolgend sollen drei im Schweizerischen Gesundheitsmarkt operierende Intermediäre miteinander verglichen werden, die ausschließlich oder teilweise zwischen den Wertschöpfungsstufen Leistungserbringer und Kostenträger operieren. Bei den vorgestellten Intermediären handelt es sich um Trust Center, H-Net und Medidata, die sich bezüglich Trägerschaft und Geschäftsmodellen deutlich voneinander unterscheiden (Abbildung 2). Bemerkenswert ist dabei, dass alle drei denkbaren Intermediationsformen zwischen zwei Parteien auftreten: Verbindung einerseits zu den Leistungserbringern und andererseits zu den Kostenträgern sowie Unabhängigkeit bezüglich beider Parteien.

	Trust Center	H-Net	Medidata
Eignerschaft	Ärzteschaft, Kantonale Ärztesellschaften, FMH. Diese drei Eigenergruppen besitzen mehr als 85% der finanziellen Anteile.	Keine Beteiligungen seitens Leistungserbringer oder Kostenträger. 100% im Besitz der Mitarbeiter und des Managements.	Mehrheitsbeteiligung > 70% durch fünf große Krankenversicherer (Kostenträger).
Dienstleistungen	Plattform Mediport; Datentransport und Haltung für ein oder mehrere Kantone sowie Aggregationsmöglichkeit für ganze Schweiz; Rechnungsdaten können von Kostenträgern über Code (Token) abgerufen werden.	Produkte H-Net eFaktura und H-Net Business Connector; Einsatz für Meldungsvermittlung und Datenaustausch unterschiedlichster Formate; Rechnungsvermittlung und Anbindungen zu EBPP- und Logistikanietern.	Datenaustausch und -haltung für Rechnungsdaten.
Verbundene Partner	Ärzte und Ärztesellschaften, teilweise Supplier für Ärzte, Versicherer über Token.	Ärzte, Spitäler, Labore, Post, Paynet, PostFinance, GHX, Krankenversicherer, etc.	Versicherer, Spitäler, Labore, Ärzte, Ärztekassen.
Schwerpunkt der Intermediärstätigkeit	Ärzte; Rechnungsspeicherung zur Aggregation und Weitergabe.	Weitergabe von Daten, keine eigene Datenhaltung; Ziel ist es, möglichst viele verschiedene Kunden in verschiedensten Bereichen des Gesundheitswesens zu erreichen.	Versicherer; Rechnungsabholung und -speicherung. Zudem Angebote von Katalogen verschiedenster Bereiche für Kostenträger.
Betriebsgesellschaft(en)	Verschiedene Betriebsgesellschaften der Trust Center kantonal oder überkantonal, z.B. Argomed.	H-Net.	Mediport.
Plattformbetreiber/-entwickler	<i>Technisch:</i> HIN (Health Info Net) für die gesamte Schweiz.	Integic.	Medidata.

Abbildung 2: Vergleich dreier schweizerischer Verrechnungsintermediäre.

Das Gesamtvolumen der jährlich 50 Mio. auszutauschenden Rechnungen verteilt sich zwischen den Leistungserbringern etwa wie in der folgenden Abbildung 3 angegeben [WEKO06].

Gültig für das Jahr 2005	Spitäler	Ärzte	Labore	Apotheken
Geschätzte Anzahl ambulanter Rechnungen	5 Mio.	30 Mio.	3 - 6 Mio.	15 - 19 Mio.
Davon in elektronischer Form (Tendenz steigend)	15 - 20 % ca. 1 Mio.	5 - 10 % ca. 3 Mio.	10 - 20 % ca. 0.45 Mio.	70 - 90 % ca. 13.6 Mio.
Davon in Papierform	80 - 85 % ca. 4 Mio.	90 - 95 % ca. 27 Mio.	80 - 90 % ca. 4.05 Mio.	10 - 30 % ca. 3.4 Mio.

Abbildung 3: Verteilung der Rechnungen in elektronischer Form und in Papierform je Leistungserbringerkategorie.

Zur Anzahl und zu den Anteilen an elektronischen und in Papierform empfangenen Rechnungen seitens der Kostenträger sind keine detaillierten Zahlen bekannt. Ebenfalls ist es schwierig, Marktanteile der erwähnten Intermediäre zu nennen. Die Mehrheit der

Apothekenrechnungen werden über den nicht weiter charakterisierten Intermediär der OFAC (Berufsgenossenschaft der Schweizer Apotheker) übermittelt. Die Mehrheit der Rechnungen der ambulanten Ärzte wird über die Trust Center abgewickelt und die Mehrheit der Spitalabrechnungen sowie der Laborabrechnungen erfolgt über die H-Net.

Ausgehend von den drei in Abbildung 2 erwähnten Finanzierungsarten und Differenzierungskriterien lassen sich unterschiedliche Kosten-Nutzen-Rechnungen oder Erlösmodelle der Intermediäre unterscheiden ([Sch05], [Sc06]). Auf Basis der Geschäftsmodell-Charakterisierung können folgende Macht- und Interessensverhältnisse festgestellt werden.

Die *Trust Center* sind, da letztlich von den Ärztesellschaften und der FMH (Foederatio Medicorum Helveticorum) kontrolliert, als Interessensvertreter der Ärzte anzusehen. Sie haben neben der reinen Datenaustauschfunktion zuhanden der Kostenträger auch die Funktion, in der Verhandlung um die Höhe der TarMed-Taxpunktweite die sogenannte Datenparität [Sto03] sicherzustellen. Dies umfasst die (anonymisierte) Aggregation der Verrechnungsdaten auf kantonaler und nationaler Ebene zur Taxpunktwertverhandlung mit den Versicherern oder Kostenträgern. Der „politische Gegner“, d.h. die Kostenträger, möchten tendenziell Taxpunktweite senken. Die Geschäftsstrategie der Trust Center ist hauptsächlich auf die Ärzte ausgerichtet. Sie berücksichtigt keine weiteren Leistungserbringer des Gesundheitssystems. Vorteile der Plattform liegen aufgrund der großen Anzahl Ärzte und Rechnungen bei den Economies of Scale.

H-Net als unabhängiger und politisch nicht gebundener Intermediär bietet aufgrund von deren Transaction Delivery Network (TDN) auf Basis der Software von Sun-Seebeyond umfassende zwischenbetriebliche Integrationsfunktionalität an, wie sie auch in Enterprise-Application-Integration-Lösungen abgedeckt werden. Dazu gehört unter vielen anderen Funktionen Transformations- und Routing-Funktionalität für Nachrichten (die u.a. Rechnungsdaten enthalten) im elektronischen Datenaustausch. Über den entsprechenden Hub werden unterschiedliche betriebswirtschaftliche und technische Geschäftsvorfälle zwischen Leistungserbringern, Kostenträgern und Lieferanten der Leistungserbringer ermöglicht. Für H-Net ist ausgehend von ihrem Geschäftsmodell und ihrer Plattform die Standardisierung von elektronischen Dokumenten etwa für die Verrechnung somit nicht von so zentraler Bedeutung wie für die anderen Intermediäre. Vorteile von H-Net liegen bei der Skalierbarkeit und den technischen Möglichkeiten der Plattform. Ferner ist die offene Konzeption der Plattform gegenüber Kostenträgern, verschiedenen Leistungserbringern und Lieferanten der

Leistungserbringer ein Vermarktungsvorteil. Die H-Net-Plattform erlaubt zudem aufgrund des Funktionsumfangs eine starke Individualisierbarkeit der Dienstleistungspalette (Economies of Scope). Dies kann zugleich eine Ausstiegshürde für die Partner beidseits des Intermediärs aus der Beziehung darstellen.

Medidata als dritte Kraft setzt sich zum Ziel, dass die Kostenträger (primär die fünf großen Versicherer SUVA, Winterthur, Zürich, CSS sowie Helsana als Mehrheitseigner) ein für sie möglichst umfassendes Dienstleistungsportfolio erhalten. Dazu gehören die Verfügbarkeit von Leistungskatalogen sowie die weitgehende Integration des Rechnungsdatenaustausches von möglichst vielen Leistungserbringern (Skaleneffekte). Dabei steht aufgrund der komplexen Datenintegration das Interesse an einer Standardisierung verschiedenster Meldungstypen und Belege auf Basis von XML im Vordergrund. Dies ist einer der Gründe dafür, dass das Forum Datenaustausch als Standardisierungsgremium für den Datenaustausch im schweizerischen Gesundheitswesen früher der Medidata nahe stand. Heute ist dieses Forum unabhängig und es sind heute alle Beteiligten des Gesundheitswesens darin vertreten, auch die Intermediäre.

Strategien der Intermediäre können nicht ohne Erwähnungen von verschiedenen Diskursebenen diskutiert werden. Ein Diskussionspunkt ist die Übernahme einmaliger und wiederkehrender Kosten für die Datenintegration. Hier vertreten die Leistungserbringer die Position, dass die Kostenträger durch die integrierte Leistungsverrechnung größere Vorteile haben als sie selber (Datenerfassung, IT-Integration, etc. verursachen Zusatzkosten), weshalb sie fordern, dass die Kostenträger die Kosten auch für die Integration auf Leistungserbringerseite übernehmen sollen. Dies sehen die Kostenträger nicht so. Eine weitere Ebene ist die der Macht und des Vertrauens, was mit der Datenerfassung und -verwertung einhergeht, insbesondere weil der TarMed eigentlich vorschreibt, dass Diagnosen der Leistungserbringer mit den Rechnungsdaten mitzuliefern sind. Dadurch würde eine Transparenz auf Seiten der Leistungsersteller entstehen, die diese gegenüber den Kostenträgern nicht unbedingt befürworten. Zu Beginn argumentierten die Ärzte denn auch so und es gab und gibt noch immer Versuche, die TarMed-Diagnose-Codes beim Datenaustausch zu unterdrücken. Dieses Vertrauensproblem führte u.a. auch zur Gründung der Trust Center.

4.2 Preismodelle

Ausgehend von den obigen Äußerungen sollen für die drei charakterisierten Intermediäre die entsprechenden Preismodelle vertieft werden. Dafür ist eingangs kurz auf zwei unterschiedliche Abrechnungsverfahren zwischen Leistungserbringern und Kostenträgern

einzuweisen. Zentral für den Austausch von Daten zwischen Leistungserbringern und Kostenträgern sind die in der Schweiz verbreiteten unterschiedlichen Abrechnungsarten *Tiers Garant* und *Tiers Payant*, wie sie bei [DGL01] und [Ja01] beschrieben werden. Der *Tiers Garant*, der Normalfall in der ambulanten Behandlung, bezeichnet das Kostenerstattungsprinzip. Honorarschuldner ist der Patient, der die Rechnung bezahlt. Er erhält im Rückgriff auf seine Krankenversicherung die Kosten vergütet. Mit dem *Tiers Payant* wird das Sachleistungs- oder Naturalleistungsprinzip bezeichnet. Honorarschuldner ist die Versicherung, welche die Rechnungen dem Arzt/Spital direkt bezahlt. Dies ist der Normalfall bei stationären Behandlungen (z.B. im Spital) und bei Unfall- und Invalidenversicherungsfällen.

Trust Center: Kostenträger, welche die Rechnungen bei den ärztseitigen Trust Centern (betriebswirtschaftliche Abwicklung der Verrechnung) abholen, zahlen im *Tiers Garant* eine Transaktionsgebühr von CHF 1. Die Trust Center verlangen für die mittels *Token* abgeholten Rechnungen offiziell CHF 1.40 zu denen zusätzlich Kosten von CHF 0.02 bis 0.03 für die Übermittlung über die Plattform HIN anfallen sowie eine monatlich an HIN zu entrichtende Gebühr von CHF 50.- pro genutztes Trust Center (in der Schweiz maximal elf). Zudem resultieren für die Trust Center seitens der Leistungserbringer pauschale Einnahmen von jährlich zwischen CHF 160 bis CHF 400 für Mitglieder der Ärztesellschaften. Für Nichtmitglieder entstehen weitere Kosten von bis zu CHF 300.

H-Net: Für den H-Net Business Connector mit dem größten Dienstleistungsumfang [Bo05] fallen jährlich vier- bis fünfstelligen Kosten (in CHF) an. Für die Übermittlung von elektronischen Rechnungen (eFaktura), was ein Teil des umfassenden Leistungsportfolios von H-Net ist, fallen Verursacher-orientierte Sendekosten von CHF 0.50 pro Rechnung/Meldung an. An H-Net angeschlossen sind mit Stand September 2005 laut [Bo05] etwa 112 Spitäler, 43 Versicherer direkt, 14 Versicherer über das Roaming mit Medidata, 3 Trust Center, 8 Radiologiestationen, ca. 700 Ärzte und diverse Labore.

Medidata: Das Preismodell von Medidata sieht für die Leistungserbringer keine Kosten für das Einlesen und Versenden der elektronischen Rechnungen an Medidata vor. Diese bezahlen die Kostenträger (Empfänger) pauschal (sechs- bis siebenstelligen CHF-Beträge pro Jahr für die fünf großen Kostenträger SUVA, Winterthur, Zürich, CSS sowie Helsana) oder mit Einzelabrechnungsbeträgen pro empfangene Rechnung (alle weiteren Versicherer, die nicht Eigner von Medidata sind). Mit dem Pauschalbetrag werden alle Dienstleistungsabgeltungen – z.B. auch Katalogdatenzugriffe – abgegolten [WE05]. Die Kosten für die Einzelabrechnung betragen

CHF 0.30 bei Intermediären, welche z.B. Apothekerrechnungen bündeln (Konzentratoren), im *Tiers Payant* (z.B. OFAC, der Verrechnungsdienstleister für die Apotheken in der Schweiz) und CHF 0.90 bei Kostenträgern im *Tiers Garant*.

Zu den Einsparpotenzialen mittels Intermediäreinsatz im elektronischen Geschäftsverkehr im Gesundheitswesen werden unterschiedliche Aussagen gemacht. [Sc06] gibt Einsparpotenziale von mindestens CHF 3.50 für den Vergleich der elektronisch verarbeiteten Rechnung gegenüber der herkömmlichen Papierrechnung an. Die geringeren Einsparpotenziale ergeben sich bei den Leistungserbringern, welche Druck- und Versandkosten der Papierrechnung vermindern können. Dies ist unterschiedlich, je nachdem, ob *Tiers Payant* oder *Tiers Garant* als Abrechnungsverfahren zum Einsatz gelangen. Die größeren Einsparpotenziale liegen bei den Kostenträgern, welche, sofern die Abrechnungen elektronisch eingelesen werden können, geringere Prozesskosten für die Rechnungsverarbeitung tragen. Das gesamte Einsparpotenzial für die elektronische Verrechnung zwischen Leistungserbringern und Kostenträgern beträgt laut [SCH05] für beide Wertschöpfungsstufen zusammen ca. CHF 200 bis CHF 300 Mio.

4.3 Intermediationsstrategien von Leistungserbringern und Kostenträgern

Ausgehend von den obigen Aussagen kann auf der Verrechnungsseite der Gesundheits-Wertschöpfungskette von einem unterschiedlichen Vorgehen der Partner bezüglich der Intermediärsnutzung ausgegangen werden. Die Abbildung 4 (in Anlehnung an [Sc06]) bietet einen Überblick über Intermediationsstrategien zwischen Leistungserbringern und Kostenträgern für den Austausch elektronischer Rechnungsdaten. Kostenträger können die Strategie Single-Channel-, Multi-Channel- und „Kostenloser Rechnungsbezug“-Intermediation haben. Als Vermittlungspartner (mittlere Schicht) kommen die in diesem Beitrag charakterisierten Intermediäre Trust Center (mit HIN als technischer Plattform), Medidata, H-Net und die nicht weiter thematisierte EDS in Frage.

Im Interesse der (fünf) großen *Kostenträger mit Single-Channel-Strategie*, denen Medidata mehrheitlich gehört, steht der Versuch, durch ihre Größe Rechnungsvolumen zu bündeln, um durch die Menge oder Größe der Anzahl Rechnungen und die Anzahl Partner, mit denen zusammengearbeitet wird, Skaleneffekte zu generieren, da die Rechnungsempfänger zahlen. Die fünf großen Kostenträger erlassen (d.h. deren Intermediär Medidata) den Leistungserbringern entsprechend die Kosten für den Rechnungsdatenaustausch (entgegen dem TarMed, welcher ausdrücklich eine Verrechnung beim Leistungserbringer vorsieht).

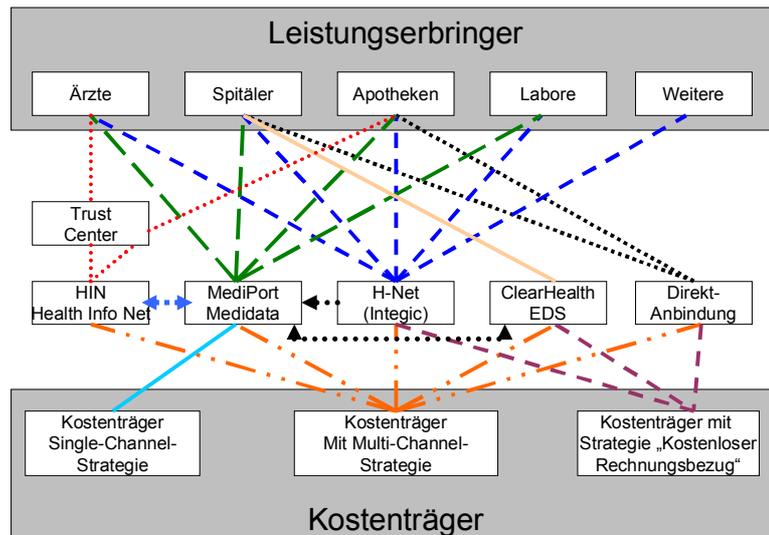


Abbildung 4: Intermediationsstrategien von Leistungserbringern und Kostenträgern.

Ausgehend von den Aussagen zweier mittelgroßer Kostenträger diskutieren auch diese zunehmend Single-Channel-Strategien der Intermediation in Richtung Leistungserbringer. Dies erfolgt hauptsächlich da, wo eine entsprechende Menge an zu übermittelnden Daten Skaleneffekte auf der Intermediationskostenseite ergeben. Zudem können über eigene Single-Channel-Schnittstellen (d.h. separate Verbindungen zu den einzelnen Versicherern) u.a. (hohe) Intermediärsgebühren vermieden werden. Diese Maßnahme kann als gegen Medidata oder H-Net gerichtet verstanden werden (Frage des Intermediations- oder Disintermediations-Nutzens). Die Interessen der *Kostenträger mit Multi-Channel-Strategie* liegen beim günstigsten Preis oder der Erreichbarkeit erforderlicher Partner. Es existieren nicht nur die fünf großen Kostenträger sondern auch kleinere, welche mit den teilweise höheren Preisen der Medidata nicht einverstanden sind und daher günstigere Varianten suchen, etwa über H-Net oder über die Trust Center. Einige wenige Versicherer profitieren hier von einem kostenlosen Rechnungsbezug über die EDS, welche einen Roaming-Vertrag mit der Medidata hat, der den kostenlosen Rechnungsbezug ermöglicht. Bei den Versicherern, die eine günstigere Intermediations-Variante suchen, ist die Masse der (bezogenen) Rechnungen nicht so groß, womit andere Intermediäre interessanter sind als diejenigen für die Single-Channel-Strategie. Es kommen (zudem), in Abhängigkeit von der Rechnungs(daten)menge oder des Leistungsumfangs des Intermediärs, pauschale oder transaktionsorientierte Abgeltungsmodelle zum Einsatz. Zudem ergeben sich über die unterschiedlichen Assoziierungen bei den Intermediären (auch aufgrund der Eignerschaft) unterschiedliche Assoziierungen von Leistungserbringern, so sind beispielsweise die Ärzte mit den „eigenen“ Trust Centern verbunden und die Spitäler

traditionell eher mit der unabhängigen H-Net. Medidata hat keine spezifischere Bindung zu bestimmten Leistungserbringern.

Die Direktanbindung von Kostenträgern und Leistungserbringern spielt vor allem dort eine Rolle, wo aufgrund der Mengen- und Kostenverhältnisse eine Intermediation gegenüber einer selbst erstellten Lösung (aus Leistungserbringersicht, hier vor allem die OFAC betreffend) weniger interessant ist. In der Tat vermittelte die OFAC als Intermediärin oder Konzentratorin der Apotheken 2004 für drei Viertel aller schweizerischen Apotheken (1'300) 11. 6 Mio. Rechnungen über ihr Netzwerk zuhanden der Kostenträger. Dies erfolgte über Telefon, Post und elektronisch. Der Gesamtbetrag der Rechnungsstellung betrug mehr als CHF 2 Mia. [OFAC06]. OFAC bündelt die Apothekenbeziehungen (elektronischer Austausch der Rechnungsdaten, Kostengutsprache) über ihr OVAN-Netzwerk (OFAC Value Added Network; basierend auf ADSL).

Zusätzliche Komplexität entsteht im geschilderten Modell in Abbildung 4 durch die verschiedenen Roaming-Abkommen zwischen den Intermediären, über die einerseits Wettbewerbspolitik betrieben wird und andererseits ein zusätzlicher Nutzen entsteht. Letzterer kann einerseits darin gesehen werden, dass: erstens der Teilnehmerkreis/Netzwerk des Intermediärs aus der Sicht des Dienstleistungsportfolios größer wird; zweitens Konkurrenten durch die machtorientierte Aufoktruierung des eigenen Preismodells in Bedrängnis gebracht werden können, was dann der Fall ist, wenn durch ein Nicht-Eingehen des Roamings Mehrkosten beim Intermediär z.B. durch Papiausdrucke der Rechnungen entstehen; drittens über ein erweitertes Roaming (Roaming zwischen eigenem Netzwerk und weiteren Added-Value-Anbietern, etwa im Bereich der elektronischen Rechnungsstellung und -zahlung (EBPP)), wodurch die Attraktivität des Dienstleistungsportfolios des Intermediärs erhöht wird. Roaming-Verträge existieren (per Ende 2005) zwischen EDS und Medidata sowie zwischen H-Net und Medidata. Ein Roaming zwischen Medidata und den Trust-Centern ist technisch realisiert worden, scheiterte aber an den (geschilderten) politisch unterschiedlichen Interessen.

Zu den Wettbewerbsstrategien der drei Intermediäre kann zusammenfassend wie folgt argumentiert werden: Medidata hat aufgrund der bereit stehenden und eingeschossenen finanziellen Mitteln eine starke Stellung. Dies ist durch die Macht der fünf großen Krankenversicherer als Eigner der Medidata gegeben. Medidata bietet eine kleine bis mittelgroße Dienstleistungspalette an, neben der eFaktura auch Katalogdatenzugriffe, etc. Die Dienstleistungspalette ist analog zu den Trust Centern mehrheitlich auf die Rechnungs-

datenvermittlung zwischen Kostenträgern und Leistungserbringern ausgerichtet. Die Trust Center haben durch die Anzahl Ärzte und die dadurch generierte Anzahl Rechnungen eine dominante Machtstellung inne. Sie stellen vom Rechnungsvolumen her den größten Intermediär dar (wobei derzeit nur ein geringer Teil elektronisch vermittelt wird, Tendenz jedoch steigend). Die Dienstleistungspalette ist jedoch eher schmal. Sie beinhaltet zusätzlich zum elektronischen Datenaustausch das Datenanalysewerkzeug Praxisspiegel für Ärzte. Bei der H-Net, dem (volumenmäßig) kleinsten Intermediär zwischen Kostenträgern und Leistungserbringern (nicht aber, was das Dienstleistungsangebot über die gesamte Gesundheitswertschöpfungskette betrifft) gehen Roamingverträge und Dienstleistungspalette weit über den Umfang der anderen beiden Intermediäre hinaus.

5 Zusammenfassung und Ausblick

Im vorliegenden Beitrag wurde anhand einer explorativen Studie im schweizerischen Gesundheitswesen untersucht, inwieweit die Anforderungen nach besser integrierten Informationsaustauschen zwischen den Beteiligten der Gesundheits-Wertschöpfungskette zum Auftreten von spezifischen Intermediären führen. Ausgehend von allgemeinen Überlegungen zu Vorteilen und Nutzen der Intermediation in diesem Bereich ergibt eine grobe Analyse der Struktur des Gesundheitswesens und der Komplexität der technischen Integration zwischen den einzelnen Akteuren, dass das Einschalten von Intermediären grundsätzlich eine sinnvolle Option sein kann, um Skaleneffekte (oder Economies of Scope) zu realisieren. Tatsächlich zeigt sich anhand des Beispiels der Schweiz, dass sich ein Markt für Intermediation gebildet hat, auf dem verschiedene Intermediäre tätig sind. Dieser Markt befindet sich allerdings erst in einem frühen und daher noch dynamischen Stadium, weshalb in den nächsten Jahren mit Veränderungen und Konsolidierungen seitens der Leistungserbringer, Intermediäre und Kostenträger gerechnet werden muss.

Eine besonders interessante Facette der hier vorgestellten Fallstudie(n) ist, dass die zu beobachtenden Intermediäre zwischen den Leistungserbringern und Kostenträgern bezüglich Trägerschaft und Naheverhältnis den drei denkbaren Idealtypen der Intermediation entsprechen: Verbindung zur Leistungserbringungsseite, Verbindung zur Kostenträgerseite und Neutralität zwischen beiden Parteien. Ausgehend von diesen Konfigurationen wurde anhand einer Analyse des jeweiligen Geschäftsmodells der drei Intermediäre untersucht, inwieweit sich die

spezifischen Interessenslagen niederschlagen. Die Untersuchung zeigt deutlich auf, dass opportunistische Interessen in der Tat einen erheblichen Einfluss auf das Geschäftsgebahren von Intermediären haben und dass dies die Position des Intermediärs zu den jeweiligen Interessensparteien beeinflusst. Dies wiederum schlägt sich nieder in den jeweiligen Intermediationsstrategien der Akteure auf Seiten der Kostenträger und Leistungserbringer.

Somit wird durch die Fallstudie(n) illustriert, dass für die Beurteilung von Intermediation nicht nur verhältnismäßig einfache und globale Betrachtungen auf der Basis der Netzwerkökonomie angestellt werden dürfen, sondern jeweils die Position und opportunistischen Interessen aller beteiligten Parteien berücksichtigt werden muss. Hier spielen neben vordergründig rationalen, ökonomischen Argumenten auch verhaltensorientierte Faktoren wie Macht und Vertrauen bzw. Misstrauen eine nicht unerhebliche Rolle [PiWi01].

Ausgehend von der hier geschilderten explorativen Untersuchung lässt sich weiterer Forschungsbedarf in verschiedene Richtungen ableiten. Hier sollen zwei Stoßrichtungen skizziert werden: Zum einen muss es darum gehen, die hier weitgehend qualitativ vorgenommenen Betrachtungen zu konkretisieren und die Kosten-Nutzen-Relationen der Intermediation im Gesundheitswesen sowie die Transaktionskosten bei verschiedenen Intermediärsmodellen genauer zu bestimmen; zum anderen stellt sich die Frage, inwieweit die hier für die Schweiz beleuchteten Phänomene sich in ähnlicher Form auch im internationalen Bereich finden lassen und ob allenfalls unterschiedliche Gesundheits-Netzwerk- und -Wertschöpfungs-Topologien die Ausprägung von Intermediation beeinflussen.

Literaturverzeichnis

- [Ae04] Aerosana (2004): Aerosana (2004): Was ist der TarMed, auf URL: http://www.aerosana.ch/pdf/de_bulletin_1_04.pdf (Aufruf per 2005-01-04; erstellt per 2004 Frühjahr).
- [Bo05] Böni, S. (2005): H-Net – Die Integrationsplattform für das Schweizer Gesundheitswesen, Präsentation vom 2005-09-29 vor der Swiss ICT Fachgruppe Investment, auf URL: <http://www.pascal-sieber.net/files/cno/cno-05/fg12-stefan-boeni.pdf> (Aufruf per 2006-06-14, erstellt per 2005-09-29).

- [DGL01] Darioli, R.; Gutzwiller, F.; Ludwig, Ch.; Pernegger, T.; Ramseier, E.W.; Schütz, R.; Westkämper, R. (2001): Texte zur Versicherungsmedizin, Schweizerisches Skript zur Versicherungsmedizin zur medizinischen Ausbildung, auf URL: <http://www.staatsexamen.ch/pdf/Versicherungsmedizin.pdf> (Aufruf per 2004-12-20; erstellt per 2001-12).
- [HaNe05] Hansen, H.R.; Neumann, G. (2005): Wirtschaftsinformatik I, Lucius&Lucius – UTB, Stuttgart.
- [Har02] Hartert, D. (2002): Global Virtual Market – Konzept für einen dezentral organisierten elektronischen Markt, auf URL: http://bibserv7.bib.uni-mannheim.de/madoc/volltexte/2003/59/pdf/59_1.pdf (Aufruf per 2006-06-12; erstellt per 2002).
- [Ja01] Jakob, R. (2001): Vom Tiers Garant zum Tiers Payant, auf URL: http://www.medidata.ch/news/_pdf/Schweizer_Versicherung_2001-11-de.pdf sowie in: Schweizer Versicherung (2001) 11, S. 71-73.
- [OFAC06] OFAC (2006): Porträt – Wer sind wir?, auf URL: http://www.ofac.ch/all/societe/text_portrait.htm (Aufruf per 2006-06-24).
- [PiWi01] Picot, A.; Reichwald, R.; Wigand, R.T. (2001): Die Grenzenlose Unternehmung, Gabler, Wiesbaden.
- [SBS95] Sarkar, M. B.; Butler, B.; Steinfield, C. (1995): Intermediaries and cybermediaries: A continuing role for mediating players in the electronic marketplace. Journal of Computer Mediated Communication [Online], 1 (3) sowie unter URL: <http://www.ascusc.org/jcmc/vol1/issue3/sarkar.html> (Aufruf per 2006-02-27).
- [Sc06] Schellenberg, R. (2006): Ambulante Rechnungen in der Schweiz: Was darf eine elektronische Rechnung kosten?, auf URL: <http://www.ebppinfo.de/portal/forum/schellenberg> (Aufruf per 2006-03-01; erstellt per 2006-02-06).

- [Sch05] Schellenberg, R. (2005): Was darf eine elektronische Rechnung kosten? Auf URL: http://www.portx.ch/de/media/files/klinik_und_heim_5-05_paper-sel.pdf (Aufruf per 2006-03-01; erstellt per 2005-05).
- [Sto03] Stoffel, U. (2003): Was ist uns Datenparität wert? – Trust-Center versus „Mediport“: Verwirrung um den elektronischen Datenaustausch im TARMED, in: Schweizerische Ärztezeitung 84 (2003) 24, S. 1510-1512.
- [SV98] Shapiro, C.; Varian, H.R. (1998): Information Rules – A Strategic Guide to the Network Economy, Harvard Business School Press, Boston.
- [WE05] Walser, K.; Egle, U. (2005): Intermediation im schweizerischen Gesundheitswesen, auf URL: http://www.im.iwi.unibe.ch/publikationen/pdfs/Interm_eHealth_Arb_Ber.pdf (Aufruf per 2006-02-27; erstellt per 2005).
- [WEKO06] Wettbewerbskommission (2006): Nicht veröffentlichter Schlussbericht vom 26. September 2006 zur elektronische Abrechnung im Gesundheitswesen, Bern.
- [Zbo96] Zbornik, S. (1996): Elektronische Märkte, elektronische Hierarchien und elektronische Netzwerke: Koordination des wirtschaftlichen Leistungsaustausches durch Mehrwertdienste auf der Basis von EDI und offenen Kommunikationssystemen, Universitätsverlag Konstanz, Konstanz.

Process-based Performance Measurement in Healthcare Networks

Günter Schicker, Jörg Purucker, Freimut Bodendorf

Department of Information Systems
University of Erlangen-Nuremberg
Lange Gasse 20
90403 Nuremberg, Germany
bodendorf@wiso.uni-erlangen.de

Abstract

Coordination and controlling in healthcare networks becomes increasingly important to enable integrated care scenarios, to enhance patient satisfaction and to reduce costs of the treatment processes. Based on the balanced scorecard a process-oriented approach for performance measurement in healthcare networks is introduced. The underlying systems architecture is presented. Integrating data from different sources and providers enables the calculation and visualization of key performance indicators in a network performance cockpit. Compliance scorecards are used to implement the network strategy and to ensure the achievement of goals. Real-time process data is obtained from a component that controls the flow of interorganizational treatment processes by web service technology. This component also supports treatment processes by process oriented e-services.

1 Introduction

The healthcare industry is one of the most important economic sectors in Germany causing annual expenses of about 230 billion euros (over 10 percent of the gross domestic product of Germany) and employing more than 4.2 million people. The German healthcare system is facing massive challenges due to the demographic and economic development as well as the increasing costs for medical innovation. Furthermore the quality is judged to be not better than mediocre [Ramm04, 147]. To improve patient satisfaction and to reduce costs of treatment processes by enhancing the cooperation between the healthcare providers a lot of different healthcare networks have been founded. In Germany the law “GKV-Modernisierungsgesetz”

enacted in 2003 improved the possibilities to realize integrated care mechanisms especially establishing cross-sectoral healthcare networks. An empirical study reveals that 81 percent of the respondents expect that networking in the healthcare industry will increase in the next three to five years [ScKB06, 17]. Moreover 88 percent of the survey participants agree that the demand for coordination and IT-support in healthcare networks is going to rise in the future. The study initiated at the Department of Information Systems II at the University Erlangen-Nuremberg addressed german and suisse ambulant healthcare networks (healthcare network managers as well as physicians). The survey investigated the maturity of healthcare network organizations regarding strategy, processes, and information technology. Only five of 90 networks show good results in overall maturity. Especially in regard of network controlling there are still a lot of challenges to cope with. Whereas three of four participants of the above mentioned survey agree that goals for the network are clearly defined only 17 % have a structured controlling system in place. Moreover just a small minority measures key data of the network to realize performance gaps. To evaluate the achievement of objectives and to improve performance more transparency by introducing an IT-supported controlling system is needed. Otherwise the advantages of healthcare networks regarding quality, efficiency and patient sovereignty can not be proved and as a result the existence of network organizations cannot be assured.

2 Research Project

The research project focuses on the IT-driven management of healthcare networks. Whereas many research projects deal with the integration of health data (e.g., electronic health records [ScKB06, 45]) this project focuses on coordination and control of interorganizational processes. Goal of the project is to support coordination and control of healthcare network processes by providing healthcare suppliers and network managers with a customized set of electronic services. Based on a balanced scorecard approach a healthcare performance cockpit delivers information for healthcare network managers and service providers. Process portals enable the interaction between users (e.g., patients, physicians) and the use of e-services provided by the system [for details see ScBo05, 7]. A process integration platform is realized enabling the design and runtime execution of a process-based e-service logistics. To analyze the requirements of network controlling the research team cooperates with the healthcare network

“Qualitäts- und Effizienzgemeinschaft Nürnberg-Nord (QuE)” which is organized as a gatekeeper system [WaLF05, 13]. The integrated care contract spans ambulant, clinical and home care service providers and is financed by a full capitation model.

3 Process-based E-Service Logistics

The concept of process-based e-service logistics is based on the interdisciplinary coordination theory. „Coordination is managing dependencies between activities performed to achieve a goal” [MaCr90, 361]. Whereas this definition is widely accepted coordination theory deals with many different means of coordination (e.g., based on forms, conversation structure or information sharing). This project argues for a process-oriented approach of coordination supported by process models as a special kind of plan in terms of coordination theory. To transfer the general tasks and principles of coordination to the healthcare domain it has to be analyzed who is cooperating and which processes and coordination tasks exist within healthcare networks.

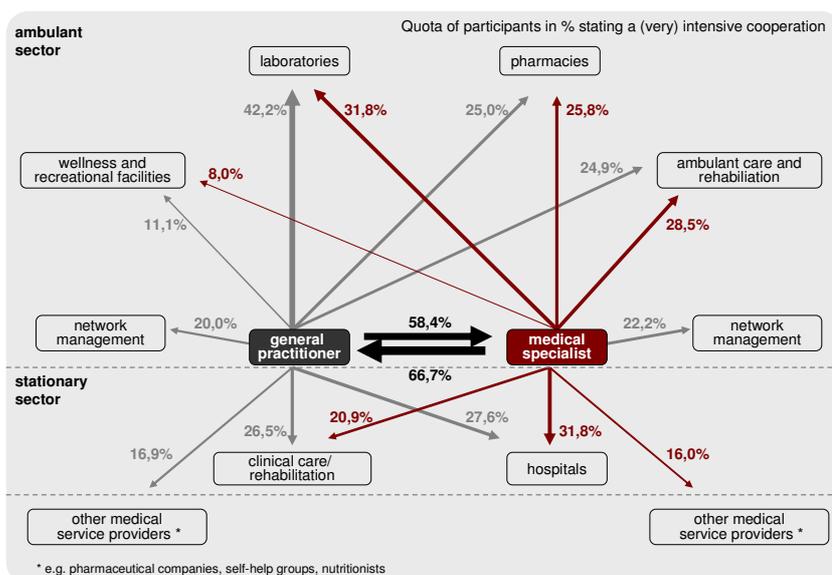


Figure 1: Intensity of cooperation

Figure 1 shows actors within integrated healthcare networks and the intensity of cooperation [ScKB06, 21]. The results show that intensive cooperation is taking place across sectoral borders resulting in numerous coordination tasks along interorganizational treatment processes. Table 1 shows some examples of processes, tasks and supporting e-services.

The research project focuses on the treatment process from a cooperative view regarding the patient’s way throughout the whole healthcare network. The individual characteristics of each

patient, the high degree of volatility of each process and its complexity have to be considered when supporting the execution of individual processes by information technology (e.g., individualization, adaptability, flexibility) [ScBo06].

healthcare network processes	coordination tasks	e-services
management processes: needed to control the healthcare network	goal adjustment, network monitoring and reporting, planning, guideline implementation	healthcare performance cockpit (balanced scorecard, stakeholder-specific reports)
medical treatment processes¹: adding value to patients and resulting in revenue for healthcare providers	controlling health status of patients, exchange and adjustment of medical reports, discharge letters or prescriptions	patient monitoring service (e.g. bluetooth scale), electronic prescriptions, electronic discharge notes
support processes: enabling processes laying the foundation to run the business	absorption of costs, accounting, billing and payment, master data management	web service orchestrated workflows for cost absorption, e-billing, patient master index

Table 1: Processes, tasks, e-services

The concept of process-based e-service logistics aims to support the coordination of healthcare network processes by providing patients and healthcare suppliers with a customized set of electronic services. Electronic services are software components which encapsulate functions (e.g., logic or data centric services) in a coarse-grained manner, e.g. using web services as technical representation [KrBS04, 70ff]. The e-service requirements regarding information and coordination in healthcare networks are derived from customized process models. They result in a process-based e-service logistics model executed by a process management platform (Individual Value Web System IVWS) supporting the coordination of individual treatment processes by providing network participants with e-services. At the level of individual patient instances treatment processes and the flow of activities throughout the network are coordinated by a gatekeeper model. The gatekeeper system aims to improve the quality of care and to realize synergies during the treatment process e.g. by avoiding unnecessary medical examinations. One member of the healthcare network is the contact person (gatekeeper) collecting all information about the patient and coordinating his treatment. The system architecture has to support this gatekeeper concept which defines the business architecture within the healthcare network [AiDo05, 614]. Hence, the central execution of web service-based workflows [BGHS03, 61] is the basic technical principle ensuring a high degree of structural analogy of business and systems architecture. To achieve this, the research project uses web service technology and the concept of service oriented architecture as technical basis. The process and e-service scheme instantiated at the first stop of the patient in the healthcare network is executed by the IVWS (for details see [ScBo06]). Figure 2 shows the architecture of the IVWS. The Meta-

¹ In literature several terms are used for medical treatment processes (e.g. guidelines, clinical/ critical pathways, interdisciplinary care paths) pointing out different origins, goals and perspectives [GISS04, 19; GrMW03, 22ff; John02, 13].

Orchestration-Server (MOS) enables the execution of individual processes and e-services that can be customized at run time.

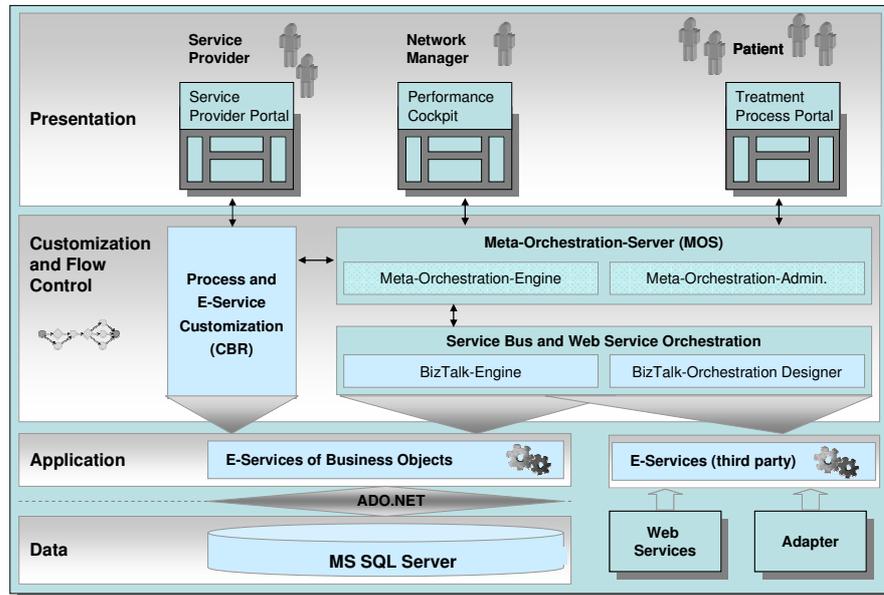


Figure 2: Architecture – Individual Value Web System (IVWS)

The system distributes e-services to roles across the network and informs the gatekeeper about the patient status. Thus the gatekeeper gets transparency concerning the treatment process giving him the possibility to intervene if necessary. The research work is based on the concept of a process-enabled service-oriented architecture (SOA). It enables “lightweight” application frontends which are only responsible for interacting with system users (dialog control). Moreover the concept argues for the encapsulation of processes within process centric (web) services. The complexity of backend systems is encapsulated within intermediary services. As a result the separation of process logic (within a process layer) and business logic (within a basic services layer) is assured [KrBS04, 79].

4 Performance Measurement in Healthcare Networks

Performance measurement is a controlling approach that focuses on the assessment of effectiveness and efficiency in companies by especially considering strategic relevant aspects including non-monetary measures [Glei02, 447]. The process-based approach for performance measurement which is described in this section relies on process data. Therefore the central process management platform that was described in section 3 is one important source of data.

4.1 Requirements

Table 2 shows crucial requirements for performance measurement systems in healthcare networks. Domain neutral requirements are relevant for network performance measurement systems not regarding the specific domain, whereas domain specific requirements are aligned to healthcare networks. Because of the given autonomy of actors within a network the network strategy must be developed and operationalized in a cooperative manner [Cors00, 24]. A big challenge for performance measurement is the extraction and integration of data from heterogeneous network actors by taking care of data privacy [LiSS04, 108; Wenn03, 62; PaBr01, 167]. Another challenging task is the definition of measures that reflect the network compliance of healthcare suppliers and that can be compared across the network and to external suppliers [Wenn03, 57ff]. Measures that reflect the network compliance must be influenceable by the suppliers in question. Process related measures can be retrieved by process data but decrease the freedom to act and therefore might badly influence the motivation. On the other hand outcome related measures are hard to calculate and mostly depend on factors that can hardly be influenced by healthcare suppliers (e. g., existence of multiple diseases, occurring of complications, patient and supplier cooperation) [PiRW03, 538ff; AQUA02, 6].

Domain neutral	Domain specific
<ul style="list-style-type: none"> • alignment to network goals • cooperative development and operationalization of network strategy • integration of heterogeneous IC-Systems • taking care of acceptance • prompt success control • operability • comparability of results between different actors in different periods of time • alignment of incentives to network goals and network compliance 	<ul style="list-style-type: none"> • considering goals and needs of stakeholders in healthcare networks • controlling of medical treatment processes • planning and monitoring of performance on network and supplier level • comparability to other healthcare networks and suppliers • considering policy holder structure • identification of potentials to develop supplier structure • considering relationships to external healthcare suppliers • taking care of data privacy • avoiding additional effort for documentation • balance of process and outcome related measures

Table 2: Requirements for performance measurement systems in healthcare networks [ScKB06; Wenn03; Toph03]

4.2 Concept

The concept introduced in this section is based on the balanced scorecard which not only because of its flexibility is said to be the most promising approach in performance measurement [Glei01, 88f]. The balanced scorecard approach is particularly appropriate for controlling in networked organizations. It explicitly addresses the implementation of a strategy throughout an or-

ganization by developing and integrating several dependent scorecards [PiRW03, 573ff]. Figure 3 shows the performance measurement process as a basis for the concept to be introduced in this chapter.

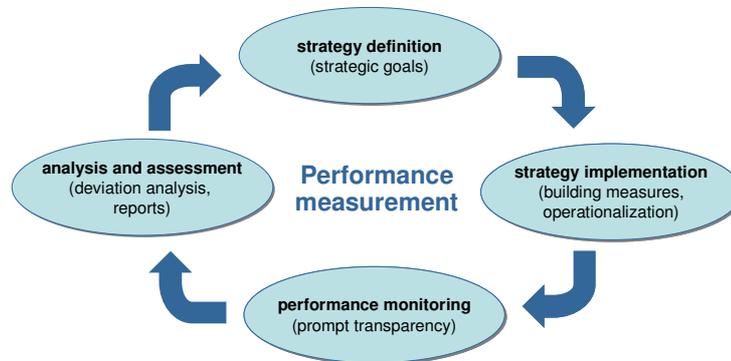


Figure 3: Performance Measurement Process

4.2.1 Strategy definition

Defining a network strategy is the first task in the performance measurement process. For this purpose the network strategy needs to be coordinated between the network management, the internal suppliers and the external partners (e. g. insurance company, association of CHI physicians). Strategy maps can be used as communication instrument [Horv04]. Figure 4 shows an exemplary strategy map for a healthcare network. Unlike in pure social organizations the mission in healthcare networks is related to economic and social aspects. Whereas in pure social organizations the financial perspective of the balanced scorecard can be placed below the process perspective to express that finance is the base for the work in the organization and for achieving the customer related goals, in healthcare networks financial goals play a more important role. Because of the causal relationship between financial and customer related goals, it is suggested to keep the financial perspective above the customer perspective. The importance of social issues can be expressed by connecting goals that are relevant for social aspects to the strategic imperatives that should be placed above the balanced scorecard perspectives [KaNo01, 120f].

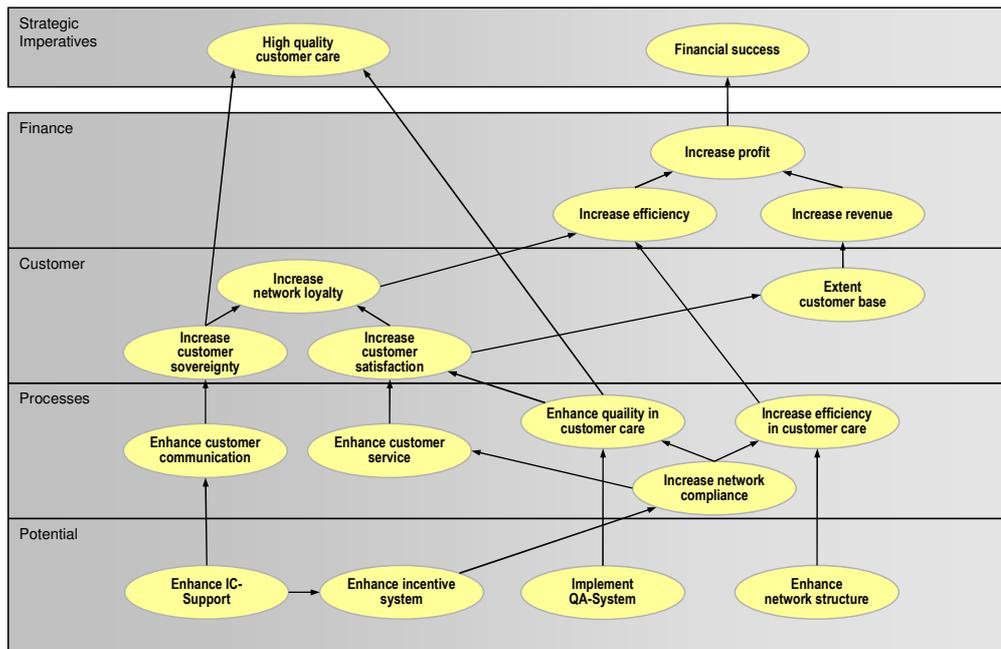


Figure 4: Strategy map for a healthcare network

4.2.2 Strategy implementation

Implementing the network strategy includes defining measures and targets, operationalizing the strategy and aligning incentives to the strategy. These tasks need to be fulfilled in coordination between the network management, the internal suppliers and the external partners. Measures need to indicate the degree of goal achievement. They must be clearly interpretable and influenceable by the actors. Collecting data should not cause too much effort [Horv04, 224]. As part of the research project for each measure in the scorecard of the healthcare network “QuE” potential data sources are retrieved. The measures were arranged in the dimensions relevance and ease of retrieving data to show which information demands can be fulfilled easily, what additional data is needed and what the retrieve costs are.

Operationalizing the strategy can be done by executing projects related to strategic goals. Another way of strategy operationalization is to create more detailed measures and scorecards across the network by building hierarchies. In Figure 5 the balanced scorecard for a healthcare network is translated to more detailed scorecards in order to specify the contribution of network actors to the network strategy. The scorecard for physicians is deduced from the network scorecard and further concretized in scorecards related to special types of physicians like gatekeepers. In this example the scorecard for gatekeepers has own measures and adopts all measures from the scorecard for network physicians. The targets can be specified.

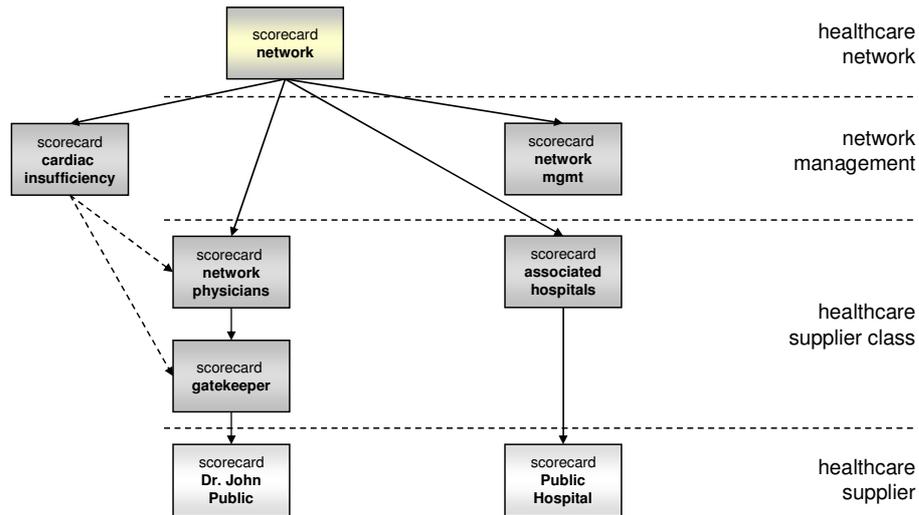


Figure 5: Building scorecard hierarchies

As suppliers in healthcare networks are organizationally independent the network management can not dictate scorecards to them. On the other hand there are directives substantiated in contracts that need to be controlled. For that reason compliance scorecards were implemented which do not necessarily reflect all goals of the suppliers but specify criteria for measuring the suppliers' network compliance. Figure 6 shows an example.

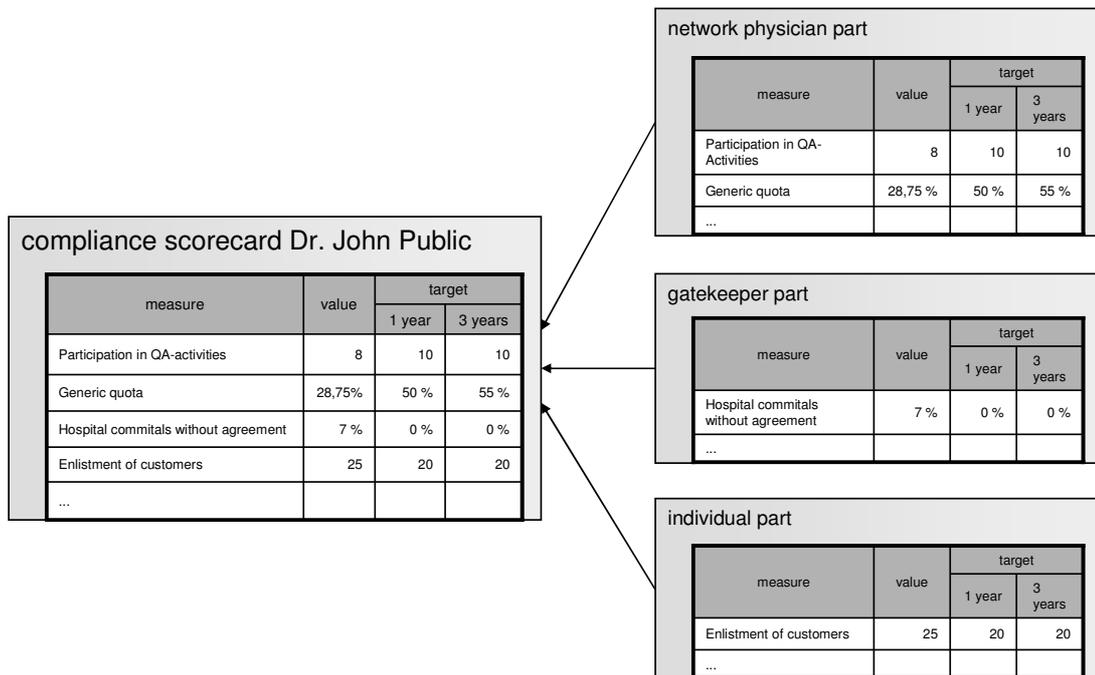


Figure 6: Exemplary compliance scorecard

In healthcare networks medical treatment processes and other network processes need to be considered when operationalizing the network scorecard. For that reason process related scorecards (e. g. for important indications like cardiac insufficiency or for medication) can be

built in an appropriate network board. These process scorecards need to be translated to and coordinated with the scorecards that are related to the network actors. Figure 5 shows an example for building a process-based scorecard hierarchy.

In the traditional German healthcare system there is a permanent incentive for suppliers to enlarge their services and hence cause immense costs for the system (hamster wheel effect). In contrast, the behavior of the network supplier needs to be aligned to network goals in full capitation healthcare networks. This can be done by implementing an incentive system. The measures in the compliance scorecard could form the assessment base for this incentive system. If the network is financed by a full capitation model, the amount of money to be distributed can be determined by the incentive system. The more money the network physicians save (e. g. by avoiding not necessary examinations), the more money can be shared. One effect of an incentive system could be that network physicians are more aligned with the assessment base than with the network goals. To avoid an abuse of the incentive system common values and mutual trust are indispensable [PiRW03, 543].

4.2.3 Performance monitoring

In order to implement a performance measurement system relevant data needs to be retrieved from different data providers. Internal health suppliers provide data related to medical treatment processes. In a web-service-based approach as described in section 3 data from external suppliers is integrated as well. External partners provide additional data regarding enlisted patients and internal physicians beyond the network processes. Network management provides additional data that was produced inside the network (e.g. surveys regarding customer and member satisfaction) or obtained externally (e.g. medication data).

Collecting, processing, and using personal data is subject to sever conditions regarding data privacy. In healthcare networks these conditions refer to patients and physicians. Processing health-related data needs to be approved by patients or governed by law. Approvals by patients are tied to the medical treatment process and to one institution. Data that was de-personalized by using anonyms or pseudonyms can be used without any patient approval. In contrast to anonyms, pseudonyms can be used to integrate data from different data sources when the same de-personalization key is used. This key must be kept secret, because otherwise mapping tables can be built to gather personalized information [Dier02, 232ff]. Because of the unclear legal position and the high demands related to the collection and processing of personalized data in the research project de-personalized data is used only.

To migrate data from different data providers in a central data pool an ETL-Process (Extraction, Transformation, Load) needs to be implemented. As data providers use different keys for de-personalization data can not be migrated on the level of pseudonyms. Therefore in a first step data is migrated in provider specific data marts. Afterwards parts of the data marts are migrated to a central data warehouse. Regarding patients the migration must be performed on a higher level of aggregation. In a long-term view it needs to be examined whether the single data providers could share the de-personalization key by using appropriate security mechanism.

As relational database systems are not an adequate solution for ad hoc analyses in vast databases, relational data can be converted to multidimensional data whenever necessary. The multidimensional data which is stored in an OLAP server can be used for many purposes as for calculating measures in scorecards and reports or performing ad hoc queries.

4.2.4 Analysis and assessment

Based on the retrieved data the network management creates reports related to the network processes and the network compliance. The healthcare suppliers can access reports automatically created regarding their individual network compliance. Figure 7 shows which features are implemented in the performance cockpit so far and what will be done in a stage of extension.

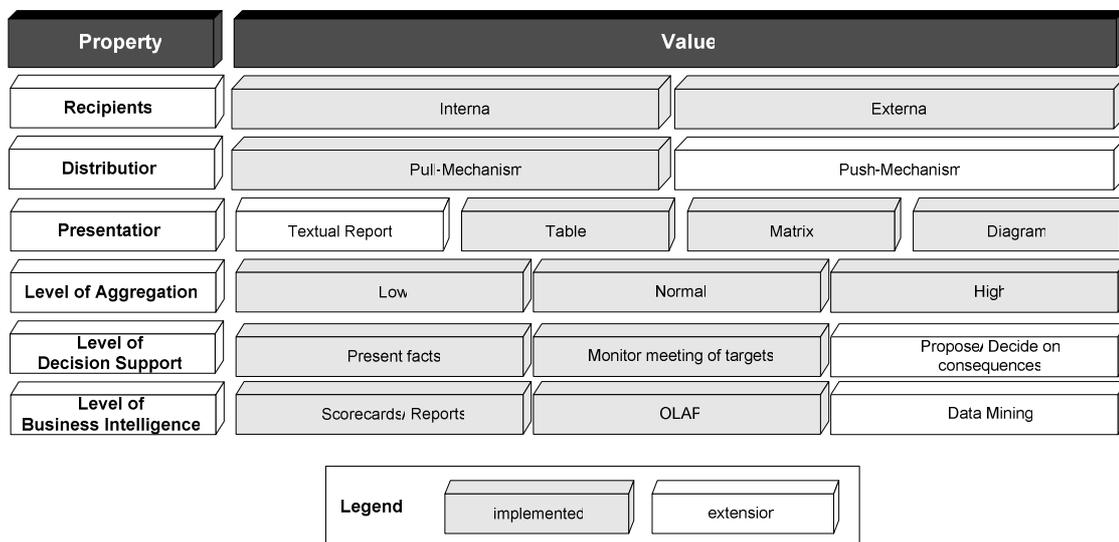


Figure 7: Analyses and report systems in healthcare networks

Reports are automatically created but can be reviewed and adjusted before being presented to external stakeholders. The reports are accessed on demand (pull-mechanism). In certain cases reports can also be automatically distributed depending on the situation and user needs (push-mechanism). Reports are presented as tables, matrices or diagrams. The performance cockpit

presents facts and monitors the achievement of goals, but does not yet propose or decide on consequences. This will be the next step when implementing an incentive system.

4.3 Technical Implementation

Figure 8 shows the architecture of the network performance cockpit. It follows the principle of SOA. Application frontends control user interaction, whereas the functionality is realized by e-services with a web service interface [KrBS04, 55ff].

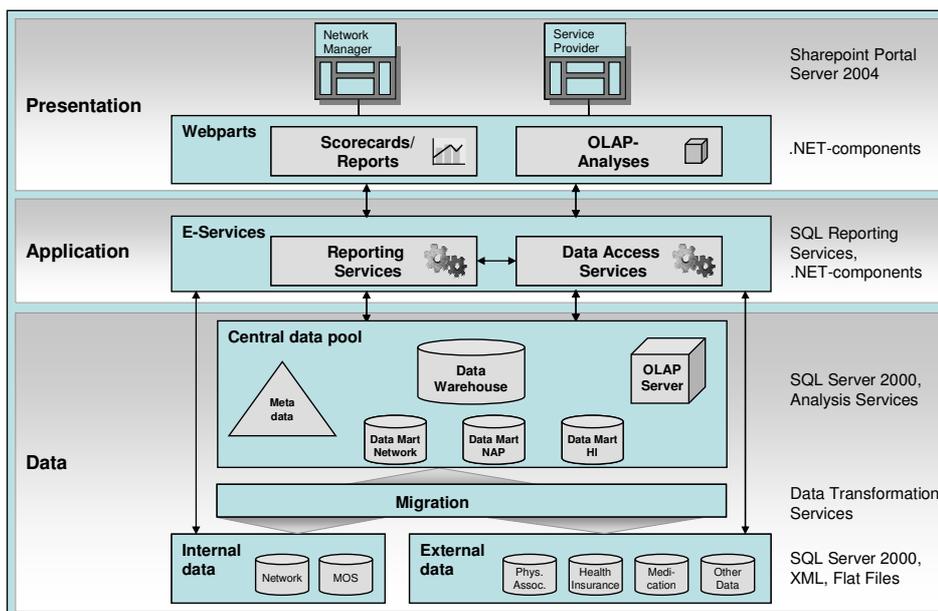


Figure 8: Performance Cockpit Architecture

4.3.1 Data Layer

Figure 9 represents important data sources for healthcare network controlling.

xDT: xDT is a collection of interfaces provided by the central institute for statutory medical care. The different xDT interfaces define a set of fields and a corresponding order. Today only ADT is a real standard as it is used for the exchange of billing data. This standard provides billing data and billing diagnoses but no medical data related to treatment processes (e. g. findings, therapies, medication). BDT includes treatment data, but most of it can not be interpreted because the contents of the corresponding fields are not well-defined [LiSe94]. However for many surgery information systems the medication data inside BDT can be interpreted. The STDT standard was implemented by two surgery information system providers only.

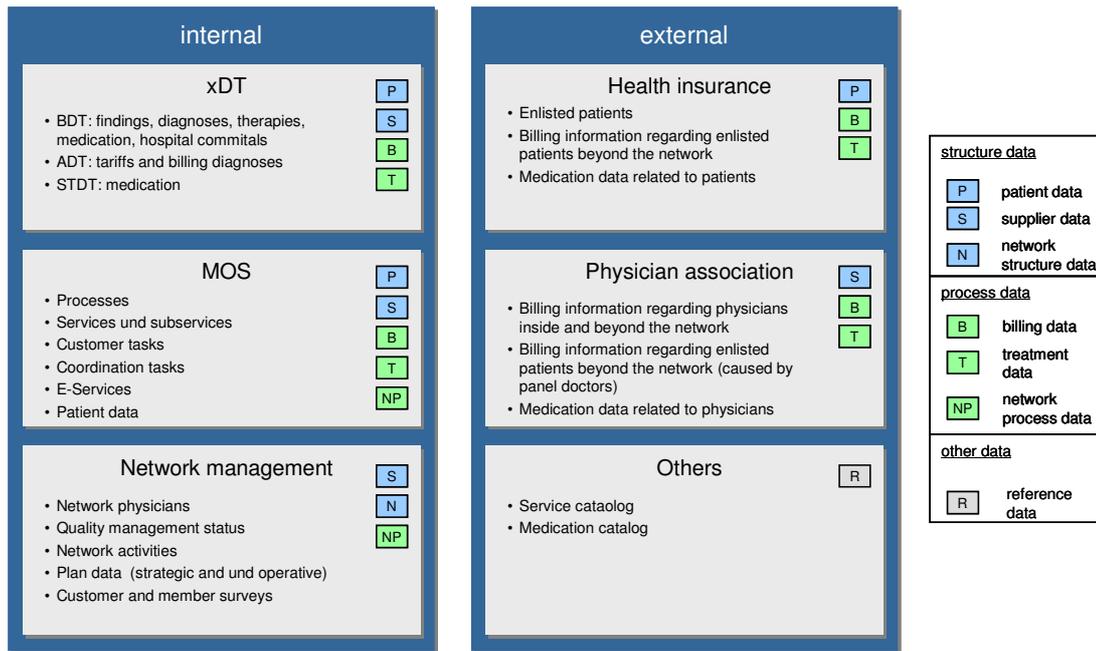


Figure 9: Data sources

MOS: The Meta-Orchestration-Server (MOS) is the central process management platform as described in section 3 which aims to support the coordination of healthcare processes. Therefore, the corresponding data scheme contains medical treatment processes, services, customer tasks, coordination tasks, e-services, and patient data. The treatment processes and corresponding elements are well-defined in a database scheme.

Network management: The network management provides data that is generated in the network, including physician master data, data regarding network activities, the status of quality management, strategic and operative planning data and data generated by customer and member surveys.

Health insurance: Insurance companies provide patient related data generated inside and beyond the network (e.g. treatment data from hospitals). This data is very important for the network management as in a full capitation model the network also has to pay for patient treatment beyond the network. The data is delivered after billing which causes a delay of about 9 months.

Association of CHI physicians: The association of CHI physicians provides data generated in the ambulant sector inside and beyond the network. The data is delivered after billing which causes a delay of about 9 months.

Other data: Depending on the need additional external data sources (e.g. medication catalog) can be integrated. In the future the telematics infrastructure especially the electronic patient record can become an important data source for patient and treatment data.

For controlling issues prompt transparency regarding the network performance is very important. As the data supply by external partners is carried out with a substantial delay, internal data needs to be retrieved. By use of the xDT interfaces data can be generated daily, but can only partly be interpreted. The implemented performance cockpit uses data generated by the MOS to show potentials related to controlling in healthcare networks. As the internal data is limited to treatment processes planned inside the network, external data supplied by health insurance companies and the association of CHI physicians needs to be integrated anyway.

4.3.2 Application Layer

Figure 10 shows the applied e-services grouped in two e-service modules. The SQL Reporting Services from Microsoft provide a web service interface to create and adjust reports. The Microsoft Analysis Services provide access to multidimensional infocubes by MDX (Multi-dimensional Expressions) a language with a syntax similar to SQL. All other e-services were developed within the research project.

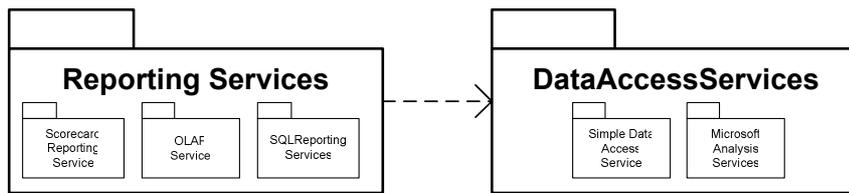


Figure 10: e-service modules

4.3.3 Presentation Layer

Figure 11 shows a management dashboard that visualizes the current network performance on a high aggregation level. For more details users can drill down to the process perspective with its goals and related measures (see Figure 12).

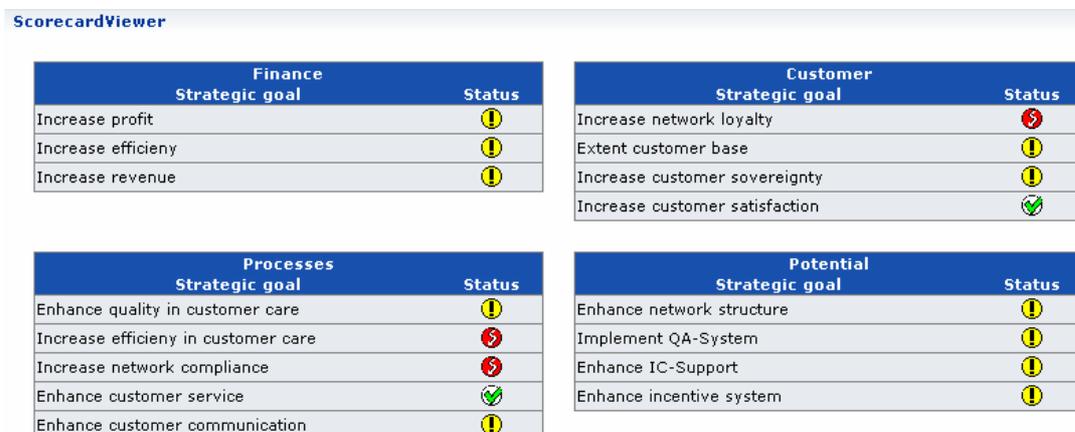


Figure 11: Dashboard for an exemplary balanced scorecard

ScorecardPerspectiveViewer						
Strategic goal / Measure	Processes					
	Status	Value	Trend	Target 06	Target 07	
<input type="checkbox"/> Enhance quality in customer care						
<input type="checkbox"/> Increase efficiency in customer care						
<input type="checkbox"/> Increase network compliance						
compliance to directives related to medical services		36 %		80 %	90 %	
compliance to directives not related to medical services		65 %		80 %	90 %	
<input type="checkbox"/> Enhance customer service						
<input type="checkbox"/> Enhance customer communication						

Figure 12: Process perspective of an exemplary balanced scorecard for healthcare networks

The status symbols are calculated by comparing the current value to the target value taking account of the target type (max, min, point). The compliance scorecard viewer webpart assesses the healthcare suppliers' network compliance. In Figure 13 the compliance scorecard consists of three parts. The first part applies to all network physicians, the second to all gatekeepers and the last one just to Dr. Public containing an individually negotiated target value for customer enlistment.

ComplianceScorecardViewer						
Report part / Measure	Dr. John Public					
	Status	Value	Trend	Target 06	Target 07	
<input type="checkbox"/> Network Physician						
Participation in quality activities		8		10	20	
Generic quota		28,75 %		50 %	55 %	
<input type="checkbox"/> Gatekeeper						
Hospital commitals without agreement		7 %		0 %	0 %	
<input type="checkbox"/> Dr. John Public						
Enlistment of customers		25		20	30	

Figure 13: Exemplary compliance scorecard related to a gatekeeper

In addition to that an OLAP Viewer enables users to navigate in multidimensional infocubes to process ad-hoc-queries.

5 Conclusion

A strategy-oriented concept for process-based performance measurement in healthcare networks was designed and prototypically implemented. The solution was developed in cooperation with an innovative healthcare network which is organized as a gatekeeper system with a full-capitation model. By monitoring medical treatment processes prompt transparency regarding network performance is reached. One important challenge for the future of performance measurement in healthcare networks is the extraction and integration of heterogeneous data by guaranteeing data privacy. Therefore standards regarding data interfaces need to be defined. In a full-capitation

model treatment processes must also be monitored beyond the borders of healthcare networks and sectors. To ease comparisons between healthcare suppliers and between healthcare networks standard measures need to be developed and implemented. Next steps will be the integration of incentive systems (“pay for performance”) and the use of more sophisticated methods for analyzing performance data like data mining and simulation.

Bibliography

- [AiDo05] Aier, Stephan; Dogan, Turgut: Indikatoren zur Bewertung der Nachhaltigkeit von Unternehmensarchitekturen. In: Wirtschaftsinformatik 2005, 7. Internationale Tagung, Bamberg 2005, S. 614-618.
- [AQUA02] AQUA: Institut für angewandte Qualitätsforschung im Gesundheitswesen: Einführung in das Indikatorenhandbuch. http://www.aqua-institut.de/pdf/indikatoren_einfuehrung.pdf, 2002-06-25, Abruf am 2006-04-12.
- [BGHS03] Burghardt Markus; Gehrke, Nick; Hagenhoff, Svenja; Schumann, Matthias.: Spezifikation und Abwicklung von Workflows auf Basis von Web-Services. In: Praxis der Wirtschaftsinformatik (Fröschle, H.-P. Hrsg.): Web Services, HMD 234, S. 61-69.
- [Cors00] Corsten, Hans.: Ansatzpunkte für die Koordination in heterarchischen und hierarchischen Unternehmensnetzwerken. Lehrstuhl für Produktionswirtschaft, Univ., Kaiserslautern 2000, S. 2-53.
- [Dier02] Dierks, Christian: Telemedizin aus juristischer Sicht. In: Klusen, Norbert; Meusch, Andreas (Hrsg.): Gesundheitstelematik: Medizinischer Fortschritt durch Informationstechnologien. Nomos, Baden-Baden 2002.
- [Glei01] Gleich, Ronald: Das System des Performance Measurement: Theoretisches Grundkonzept, Entwicklungs- und Anwendungsstand. Vahlen, München 2001.
- [Glei02] Gleich, Ronald: Performance Measurement: Grundlagen, Konzepte und empirische Erkenntnisse. In: Controlling 14 (2002) 8-9, S. 447-454.

- [GISS04] Glock, Gerhard; Sohn, Stefan; Schöffski, Oliver: IT-Unterstützung für den medizinischen Prozess in der integrierten Versorgung. Books on Demand, Burgdorf 2004.
- [GrMW03] Greiling, Michael; Mormann, Johanna; Westerfeld, Ruth: Klinische Pfade steuern. Baumann, Kulmbach 2003.
- [Horv04] Horváth & Partners (Hrsg.): Balanced Scorecard umsetzen. 3. Aufl., Schäffer-Poeschel, Stuttgart 2004.
- [John02] Johnson, Sue: Interdisziplinäre Versorgungspfade. 1. Aufl., Verlag Hans Huber, Bern 2002.
- [KaNo01] Kaplan, Robert S.; Norton, David P.: Die strategiefokussierte Organisation: Führen mit der Balanced Scorecard. Schäffer-Poeschel, Stuttgart 2001.
- [KrBS04] Krafzig, Dirk; Banke, Karl; Slama, Dirk: Enterprise SOA: Service Oriented Architecture Best Practices. Pearson Education, Maryland 2004.
- [LiSe94] Lichtner, Friedrich; Sembritzki, Jürgen: BDT-Satzbeschreibung: Schnittstellenbeschreibung zum systemunabhängigen Datentransfer von Behandlungsdaten. (unveröffentlicht), Köln 1994.
- [LiSS04] Lindenthal, Jörg; Sohn, Stefan; Schöffski, Oliver: Praxisnetze der nächsten Generation: Ziele, Mittelverteilung und Steuerungsmechanismen. Books on Demand, Burgdorf 2004.
- [MaCr90] Malone, Thomas W.; Crowston, Kevin.: What is Coordination Theory and How can it Help Design Cooperative Work Systems? In: CSCW'90, Los Angeles, Proceedings of the 3rd Conference on Computer-Supported Cooperative Work. New York 1990, S. 357-370.
- [PaBr01] Paul, Volker; Bresser, Bertram: Spezielle Probleme der Medizintelematik im Alltag. In: Hellmann, Wolfgang (Hrsg.): Management von Gesundheitsnetzen. Kohlhammer, Stuttgart 2001, S. 163-180.

- [PiRW03] Picot, Arnold; Reichwald, Ralf; Wigand, Rolf T.: Die grenzenlose Unternehmung. 5. Aufl., Gabler, Wiesbaden 2003.
- [Ramm04] Ramming, Joachim: Integrierte Gesundheitsversorgung. In: Jähn, Karl; Nagel, Eckhard (Hrsg.): e-Health. Springer, Heidelberg 2004, S. 147-151.
- [Toph03] Tophoven, Christina: Integrierte Angebotsstrukturen – Netze auf dem Weg zur markt- und vertragsfähigen Organisation. In: Tophoven, Christina, Lieschke, Lothar (Hrsg.): Integrierte Versorgung: Entwicklungsperspektiven für Praxisnetze. Deutscher Ärzte-Verlag, Köln 2003, S. 229-260.
- [ScBo05] Schicker, Günter; Bodendorf, Freimut: Portalunterstützte Behandlungspfade in Gesundheitsnetzen: Prozesse - Architektur – dynamische Navigation. In: Cremers, Armin B.; Manthey, Rainer; Martini, Peter; Steinhage, Volker (Hrsg.): INFORMATIK 2005: Informatik LIVE! Band 2. GI-Edition „Lecture Notes in Informatics“ (LNI): P68, Bonner Köllen Verlag, Bonn 2005, S. 7-11.
- [ScBo06] Schicker, Günter; Bodendorf, Freimut: Process-based E-Service-Logistics for Healthcare Networks. In: Gesellschaft für Informatik (Hrsg.): Lecture Notes in Informatics (LNI)- Proceedings. European Conference on eHealth (ECEH06). Fribourg 2006.
- [ScKB06] Schicker, Günter; Kohlbauer, Oliver; Bodendorf, Freimut: Praxisnetz-Studie 2006 - Status Quo, Trends & Herausforderungen. Arbeitspapier Wirtschaftsinformatik II Nr. 01/2006, Universität Erlangen-Nürnberg, Nürnberg, 2006.
- [WaLF05] Wambach, Veit; Lindenthal, Jörg; Frommelt, Mona: Integrierte Versorgung – Zukunftssicherung für niedergelassene Ärzte. Ecomed MEDIZIN, Landsberg 2005.
- [Wenn03] Wenninger-Zeman, Katrin: Controlling in Unternehmensnetzwerken: Eine organisationstheoretische Betrachtung. Deutscher Universitäts-Verlag, Wiesbaden 2003.

Zielgruppenspezifische Dienste für Virtuelle Patientengemeinschaften

Achim Dannecker und Ulrike Lechner

Institut für Angewandte Systemwissenschaften und Wirtschaftsinformatik
Universität der Bundeswehr München
85577 Neubiberg
Achim.Dannecker@unibw.de und Ulrike.Lechner@unibw.de

ABSTRACT

Virtuelle Patientengemeinschaften bieten Mitgliedern Informationen zum Thema Gesundheit und wechselseitige Unterstützung. In diesem Beitrag werden die wesentlichen Erfolgsfaktoren von virtuellen Patientengemeinschaften und der Bedarf nach elektronischen Diensten analysiert. Die Resultate beinhalten Erfolgsfaktoren, den Bedarf an elektronischen Diensten für virtuelle Patientengemeinschaften und für jüngere und ältere Mitglieder dieser Gemeinschaften. Kurz werden wesentliche Elemente eines generischen Bewertungsdienstes für medizinisch relevante Informationen dargestellt.

1 Einleitung und Motivation

Virtuelle Patientengemeinschaften (VPG) können –theoretisch– eine ganze Reihe von Aufgaben im Gesundheitswesen übernehmen. Beispiele für VPG, die in der Literatur genannt werden, sind das „Cancerforum“ von compuserve [Rhei94; HaAr97], Virtuelle Gemeinschaften für Brustkrebspatientinnen oder chronisch kranke Patienten [KADL02; Jose04; LeDK04]. Allein im Bereich Gesundheit („Health & Wellness“) bei Yahoo [Yaho05] gibt es mehr als 74.000 Gemeinschaftsforen. Unter VPG verstehen wir Online Gemeinschaften von Patienten, Angehörigen, Medizinern, Interessierten und Forschern, die sich mit einem Krankheitsbild auseinandersetzen und deren wesentliches Kommunikationsmedium eine Plattform im Internet ist. Es ist festzustellen, dass VPG sich heute im Wesentlichen auf wechselseitige Unterstützung von Patienten und Angehörigen untereinander und Information für Patienten und Angehörige zu be-

schränken scheinen. VPG haben häufig nur wenige aktive Mitglieder und nutzen nur wenige Arten von elektronischen Diensten [DaLe04]. VPG stellen eine wichtige Informationsquelle dar und die wechselseitige Unterstützung für Patienten ist wichtig. Daher stellt sich die Frage, ob VPG durch neue elektronische Dienste gestärkt werden können und vielleicht damit ihre Funktionen besser wahrnehmen oder weitere Funktionen im Gesundheitswesen übernehmen können. Es wurde eine empirische Studie bei Mitgliedern und Betreibern von VPG durchgeführt. Die Ergebnisse beinhalten eine Analyse der wesentlichen Faktoren und E-Service Komponenten und der speziellen Anforderungen von jüngeren und älteren Mitglieder von VPG, ein Modell basierend auf einer explorativen Faktoranalyse und eine kurze Beschreibung eines Dienstes, der auf Basis der Resultate der empirischen Studie entwickelt wurde.

Das Papier ist wie folgt strukturiert. Der aktuelle Stand der Forschung im Bereich Selbsthilfeorganisationen und Selbsthilfegruppen von Patienten wird in Kap. 2 dargestellt. Die Forschungsmethode wird in Kap. 3 erläutert. Die Resultate der empirischen Untersuchung werden in Kap. 4 und eine Skizze eines generischen Dienstes in Kap. 5 präsentiert. Eine Diskussion in Kap. 6 beschließt das Papier.

2 Stand der Wissenschaft und Praxis

Das Gesundheitswesen befindet sich in einem Reorganisationsprozess und somit eröffnen sich neue Möglichkeiten für elektronische Dienste und Marktplätze [KaTa05]. Das Internet wird zunehmend als Medium für gesundheitsrelevante Informationen genutzt. Es ist möglich vielfältige Dienste für das Management der eigenen Gesundheit genauso wie Informationen von Dienstleistern und medizinischen Institutionen zu finden [Gold05; HuTa05; PUCS06]. Eine offene Frage ist, wie die Information aus dem Internet mit den Informationen, die Patienten von Medizinern erhalten, bei wichtigen gesundheitsrelevanten Entscheidungen in Relation gesetzt wird [PUCS06]. Im Folgenden wird die aktuelle Situation in der Selbsthilfe von Betroffenen analysiert und wie VPG diese Selbsthilfe unterstützen können.

2.1 Selbsthilfeorganisationen und Selbsthilfegruppen von Patienten

Der Erfahrungsaustausch von Patienten oder als Angehörige von Betroffenen wird traditionell „Offline“ in Selbsthilfeorganisationen (SHO) und Selbsthilfegruppen (SHG) organisiert. SHO informieren typischerweise ihre Mitglieder über alle Aspekte der relevanten Krankheit und a-

gieren als „Repräsentanten“ (Patientenvereinigung). Janke et al. stellen fest, dass Patienten, die in einer SHO Mitglied sind, besser über ihre Krankheit informiert sind, als Patienten, die zu keiner SHO gehören [JKGM05]. Borgaonkar et al. auf der anderen Seite zeigt, dass das bloße zur Verfügung stellen von relevanten Informationen die gesundheitsspezifische Lebensqualität von Betroffenen (health-related quality of life – HRQOL) verschlechtert [BoTD02]. Dies deckt sich mit Resultaten aus Interviews mit Betreibern von SHO, die bestätigen, dass das reine zur Verfügung stellen von gesundheitsspezifischen Informationen und das kontinuierliche Erinnern an die Krankheit der Patienten (auf Basis von Mailings, Broschüren oder Newslettern) kontraproduktiv ist, da es zu einer Verschlechterung des HRQOL und zu nicht unerheblichen Beendigungen von Mitgliedschaften führt. In einer Langzeitstudie konnten Kennedy et al. zeigen, dass eine spezielle Form von Informationen, Informationen die von Patienten für Patienten erstellt wurden “(...) patients given a patient-developed guidebook of self-management skills experienced (...)”, HRQOL signifikant ansteigen lies [KRNR03]. Dies lässt den Schluss zu, dass elektronische Dienste, die nur Informationen anbieten, nicht hinreichend sind, um Patienten in ihrer Lebenssituation zu unterstützen.

Die Teilnehmer einer SHG treffen sich regelmäßig und werden hauptsächlich durch zwei Aspekte getrieben, die wechselseitige Unterstützung und der Austausch von Informationen [Borg04]. Hierbei können die Teilnehmer von den Erfahrungen der Anderen, gerade im Umgang mit der Krankheit im Alltag und in langfristigen Entwicklungen des Krankheitsbildes, profitieren. Teilnehmer empfinden es als positiv zu erfahren, dass sie nicht die einzigen Betroffenen sind. Themen, die besprochen werden, betreffen Ärzte, klinische Institutionen, alternative Behandlungsmethoden, Rehabilitationsinstitutionen, Medikamente und Medikamentierung, aktuelle Forschung und auch die Teilnahme an klinischen Studien.

2.2 VPG und Online Gemeinschaften

Aktuell betreiben viele SHO eine Online Gemeinschaft oder wenigstens ein Forum innerhalb ihres Internetauftritts. Ein Interviewpartner (Betreiber einer VPG einer SHO) beschreibt Ende 2004 die typische Situation: In den letzten Jahren hat sich kaum etwas geändert. Das Forum hat kaum über 100 reguläre Besucher, die i.d.R. nur für eine kurze Zeit im Forum aktiv sind. Die gleiche “Newbie” Frage wird immer und immer wieder gestellt. Gängige, nicht notwendigerweise evidenzbasierte Theorien über die Ursache der Krankheit und mögliche, ebenfalls nicht evidenzbasierte häufig alternative Behandlungsformen werden diskutiert oder „Newbies“ auf ihre Fragen hin angeboten. Neu diagnostizierte Personen stellen i.d.R. im Forum ein oder zwei

dringliche Fragen und verschwinden dann wieder aus dem Forum. Nur wenige bleiben dem Forum für einen längeren Zeitraum als aktive Mitglieder treu. Fundierte Diskussionen finden zum Großteil außerhalb des Forums statt und so kann das typische Mitglied davon nicht profitieren. Dennoch ist diese SHO zufrieden mit der von ihr bereitgestellten Auswahl an medizinischen Informationen und der großen Anzahl von Mitgliedern, die sie über einen Newsletter erreicht. Werden VPG betrachtet, die nicht mit einer Selbsthilfe-Organisation assoziiert sind, bietet sich ein anderes Bild. Viele dieser VPG widmen sich einer bestimmten Krankheit in Verbindung mit eher weltanschaulich geprägten Theorien über Ursache und Behandlungsmöglichkeiten, wenigen Diskussionen und wenig Anzeichen anderer Aktivitäten [DaLe04; Yaho05].

Die Forschung im Bereich der Online Gemeinschaften fokussiert auf aktive Mitglieder von Online-Gruppen und die Beiträge, die diese Mitglieder leisten. Vergleichsweise wenige Forschungsarbeiten beschäftigen sich mit passiven Mitgliedern, so genannten „Lurkern“, die sehr wenig bis gar nicht aktiv an den Online-Gruppen teilnehmen, sondern vielmehr die Inhalte „nur“ konsumieren. „Lurker“ stellen als solches in Online Gemeinschaften kein Problem dar, da sie sich als Teil der Gemeinschaft fühlen und auch als Teil der Gemeinschaft wahrgenommen werden. Ihr Vorhandensein motiviert andere Mitglieder, aktiv Online ihr Wissen bereitzustellen. [NoPr01; NoPA04; McFa05]. Virtuelle Gemeinschaften werden charakterisiert durch Regeln der Interaktion, Wertesystem, wechselseitige Unterstützung und gemeinsame Ziele und Interessen [WiOD97; Figa98; Pree00; Well01]. Leimeister et al. erfassen technologische, organisatorische und soziale Erfolgsfaktoren von Virtuellen Gemeinschaften im Allgemeinen [LeSK04]. Eine Besonderheit von VPG ist die Integration der Offline- (die Krankheit und welchen Einfluss diese auf das reale Leben hat) und Online-Welt (der Informationen und der wechselseitigen Unterstützung). Das „Integration Principle“ stellt sechs Beziehungen zwischen der „realen“ Welt und der „online“ Welt dar [Sule00]: *“translating troublesome issues from one realm to the other can be helpful, even therapeutic (...)”*.

3 Forschungsmethode

Das Ziel der Forschungsarbeit ist es, Dienste für VPG zu entwickeln. Im ersten Teil werden die Erfolgsfaktoren von VPG ermittelt und ein Modell entwickelt. Unterschiedliche Gruppen innerhalb VPG haben unterschiedliche Bedürfnisse für Dienste [DaLe06a]. Im vorliegenden Papier werden die Bedürfnisse jüngerer und älterer Mitglieder von VPG analysiert.

Der Fragebogen basiert auf Resultaten einer Vorstudie über Interaktion und Dienste in VPG [DaLe04], offenen Interviews mit Betreibern von VPG und einer Studie von [LeSK04] zu Erfolgsfaktoren von virtuellen Gemeinschaften. Der Fragebogen von [LeSK04] wurde um für VPG spezifische Fragestellungen erweitert und die bereits von Leimeister et al. als niedrig in der Rangliste der Erfolgsfaktoren erkannten und die für die befragte Zielgruppe irrelevanten Fragen wurden eliminiert.

Der Fragebogen besteht aus vier Teilen. (1) Demographische Aspekte (Alter, Geschlecht, Internetnutzung etc.), (2) Krankheit betreffende Aspekten, (3) Aspekte der Verknüpfung der Offline- und Online-Welt und (4) Fragen zu elektronischen Diensten und dem sozialen Netzwerk. Dieser vierte Teil besteht aus 34 Fragen zu medizinischen Informationen und Online Inhalten, qualitätssichernden Maßnahmen basierend auf Beiträgen von Mitgliedern und den Betreibern, zur Rolle der Betreiber von Online Gemeinschaften, zu technischen Aspekten, Interaktionsmöglichkeiten und der emotionalen Bindung der Mitglieder an die VPG. In Interviews hat sich herausgestellt, dass Angehörige und Mediziner auf unterschiedliche Weise in die VGP integriert sind und dass es wichtig ist, dass der Großteil der Mitglieder selbst von der in der VPG behandelten Krankheit betroffen ist. Daher wurden Fragen zur Integration von Angehörigen und Mediziner aufgenommen.

Eine Studie zu VPG [DaLe04], der Interaktion und des Grades der Verwendung von Diensten zeigt, dass es Unterschiede in der Handhabung von medizinischen Inhalten in VGP, der Kultur und der Interaktion zwischen Mitgliedern gibt (z.B. sind Diskussionen über alternative Behandlungsmethoden oder klinische Institutionen erlaubt?). Fragen nach Diensten für das Management medizinischer Informationen in VPG wurden in den Fragebogen aufgenommen.

Ein wesentlicher Punkt in den Interviews mit Betreibern von VPG war das Zusammenspiel zwischen Selbsthilfegruppen (Treffen in der realen Welt) und der VPG. Mitglieder von VPG nehmen an Treffen von Selbsthilfegruppen teil und es besteht aber die Wahrnehmung, dass beide Interaktionsformen in Konkurrenz stehen und dass sich Themen und Kultur in den beiden Organisationsformen unterscheiden. Fragestellungen zu Selbsthilfegruppen und der Verknüpfung zwischen der realen Welt und der VPG wurden daher aufgenommen. In den Interviews wurde von den Betreibern der Standpunkt vertreten, dass Neutralität (z.B. nicht durch die Pharmaindustrie unterstützt zu sein) für die Mitglieder sehr wichtig ist und eine entsprechende Frage nach der Wichtigkeit der Neutralität wurde ebenfalls aufgenommen.

In Anlehnung an die Studie von Leimeister et al. [LeSK04] wurde eine bipolare verbal beschriebene Ordinalskala verwendet (beginnend mit “Wichtig ist...”) wobei Zustimmung oder Ablehnung erfragt werden konnte (siehe Tab. 1).

Starke Zustimmung	Zustimmung	Neutral	Ablehnung	Starke Ablehnung	Keine Angabe
=4	=3	=2	=1	=0	=9

Tab 1: Bipolare Ordinalskala und Reinterpretation in Zahlen.

Für Mitglieder und Betreiber von VPG wurden zwei verschiedene Fragenkataloge entwickelt. Unter Betreibern von VPG verstehen wir die Personen, die Plattform und Dienste bereitstellen und typischerweise auch Informationen bereitstellen und als Moderatoren tätig sind. Im Allgemeinen sind auch die Betreiber Betroffene.

Für die Auswahl von VPG im deutschsprachigen Raum wurde eine Internet-Recherche (Yahoo, Google) angewandt. Im Rahmen dieser Recherche wurden ca. 250 VPG im deutschsprachigen Raum identifiziert. VPG mit weniger als 50 Mitgliedern sowie VPG, deren aktuellste Beiträge älter als ein Jahr waren, wurden verworfen. Übrig blieben 117 VPG, von denen 73 (63%) zufällig ausgewählt wurden. Zehn weiteren VPG, mit denen bereits eine Zusammenarbeit bestand, wurde eine erste Version der Fragebögen mit der Bitte um Verbesserungsvorschläge zugesandt und diese 10 VPG wurden ebenfalls befragt.

Zusammen mit einer Ankündigung der Studie wurden die überarbeiteten Fragebögen den Betreibern dieser 83 VPG, mit der Bitte diese Studie zu unterstützen, zugesandt. Die Online Fragebögen wurden 3 Wochen im Mai/Juni 2005 im Internet zur Verfügung gestellt. Nach Elimination von leeren Antworten oder Duplikaten wurden 295 Antworten von Mitgliedern von VPG und 21 Antworten von Betreibern von VPG hinterlegt. Im Folgenden werden vor allem die Antworten der Mitglieder von VPG betrachtet.

4 Resultate

Die Teilnehmer der Studie sind in insgesamt 145 unterschiedlichen VPG aktiv. Bei dieser Frage waren Mehrfachantworten möglich. Die „Top Ten“ der meist genannten VPG wurden hierbei von 50% der Teilnehmer genannt. Die zehn am häufigsten genannten VPG (inkl. Anzahl der Antworten und prozentualer Anteil) sind:

	VPG	Anzahl der Teilnehmer	Verteilung
1	rheuma-online.de (Rheumatismus)	50	11,74%

2	fibromyalgie-aktuell.de (Schmerzpatienten)	35	8,22%
3	dccv.de (Morbus Crohn / Colitis Ulcerosa)	31	7,28%
4	croehnchen-klub.de (Morbus Crohn / Colitis Ulcerosa)	24	5,59%
5	sylvia.at (Morbus Crohn / Colitis Ulcerosa)	15	3,50%
6	diabetes-world.net (Diabetes)	14	3,26%
7	prostatakrebse.de (Prostatakrebs)	14	3,26%
8	sd-krebs.de (Schilddrüsenkrebs)	10	2,33%
9	leukaemie-betroffene.de (Leukämie)	9	2,10%
10	kisp.de (Prostatakrebs)	9	2,10%

Tab. 2: Top.Ten Liste der meist genannten VPG

16 der VPG wurden zwei Mal und 100 VPG nur ein Mal genannt. Bei den Krankheiten werden Rheuma (inkl. Fibromyalgie) (20%), gefolgt von Morbus Crohn (17%), Krebs (11%), Diabetes (6%) und Tinnitus (5%) am häufigsten genannt. Es sind mehr als 95% der Befragten von einer chronischen Krankheit betroffen.

4.1 Demographische Informationen

	Mitglieder		
	m	w	Gesamt
Teilnehmer-Anzahl	87	208	295
Durchschnittsalter	47,78	39,13	41,68
Online Zeit (h pro Tag)	3,18	3,30	3,26
Zeit in VPG (h pro Tag)	0,80	0,89	0,87
Anzahl der Mitgliedschaften in VPG	1,21	1,62	1,50
Zeit der Mitgliedschaft oder des Betriebs der VPG (Jahre)	3,67	4,35	4,15
Selbst betroffen von der Krankheit?	ja: 83 nein: 4	ja: 198 nein: 10	
Zeitraum der Erkrankung (Jahre)	8,74	10,37	9,89
Haben Sie schon einmal einen Dienst wie „Ask the expert“ benutzt?	ja: 13 nein: 74	ja: 67 nein: 141	
Nehmen Sie an Selbsthilfegruppen (SHG)-Treffen teil (j / n)? (Sehr oft 4 – Ab und zu 2 – Niemals 0)	44 / 43 1,06	110 / 98 0,96	
Wo fühlen Sie sich besser aufgehoben? VPG / gleich / SHG	28 / 50 / 9	87 / 109 / 12	
Wie oft schreiben Sie Beiträge? (mehrmals tgl. 4 - wöchentl. 2 - nie 0)	1,61	2,00	1,88
Kennen Sie Mitglieder im realen Leben?	ja: 32 nein: 55	ja: 104 nein: 104	
Spielt die VPG eine zentrale Rolle in Ihrem Leben (Stimme stark zu 4 – Unentschieden 2 – Lehne stark ab 0)	2,27	2,76	2,62
Entwicklungszufriedenheit mit der VPG (Sehr zufrieden 4 - Unentschieden 2 – Vollkommen enttäuscht 0)	2,76	3,16	3,04
Stellen Sie Fragen in der VPG, die Sie einem Arzt nicht stellen (Sehr oft 4 – Ab und zu 2 – Niemals 0)	1,75	2,02	1,94

Tab. 3: Demographische Daten (nur Mitglieder)

Auffällig ist, dass 208 Frauen und 87 Männer geantwortet haben. Das Durchschnittsalter ist mit über 40 Jahren für eine Umfrage am Internet hoch (vgl. Abb. 1). Die Mitglieder verbringen ca. 45 Minuten pro Tag in ihren VPG, sind ihrer VPG über Jahre hinweg treu und nur in durch-

schnittlich 1,5 VPG engagiert. 50% der Frauen, aber nur 30% der Männer kennen andere Mitglieder persönlich (aus dem realen Leben). Die Zufriedenheit mit der Entwicklung der VPG liegt mit einem Wert von 1,96 (2,0 entspricht zufrieden; 1.0 sehr zufrieden) hoch. Die am häufigsten genannten Krankheiten passen zu der Alterstruktur der Teilnehmer.

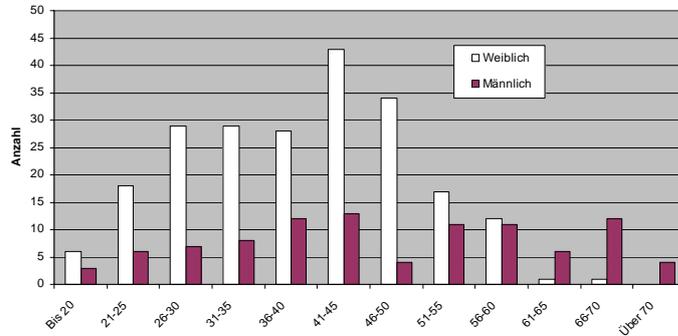


Abb. 1: Alterstruktur und Geschlecht der Teilnehmer

4.2 Wichtige Faktoren

In Abb. 2 sind die 10 Fragen, die von den Mitgliedern der VPG als am Wichtigsten bewertet wurden und die Auswirkungen auf E-Service Komponenten haben, nach der Reihenfolge der Wichtigkeit sortiert dargestellt und den Werten der Betreiber gegenübergestellt. Der sensible Umgang mit den Daten der Mitglieder wird von den Mitgliedern am wichtigsten erachtet. Auch die Betreiber erachten diesen Punkt als sehr wichtig.

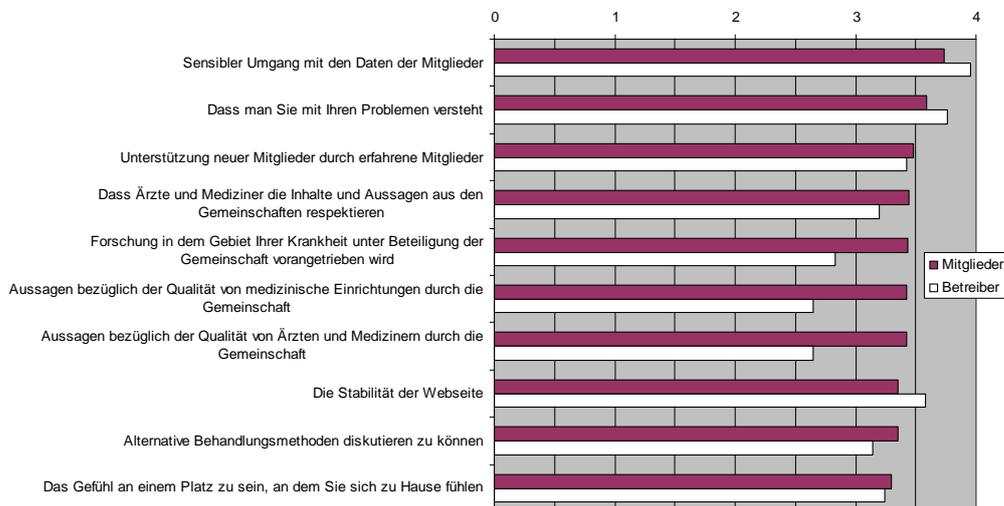


Abb. 2: Vergleich der Werte geordnet nach der Wichtigkeit aus Sicht der Mitglieder (4,0 – sehr wichtig, 0,0 – lehne stark ab)

Die nächsten Punkte behandeln Dienste, die es der VPG erlauben, Informationen über medizinische Einrichtungen, von Ärzten und Mediziner wie auch alternative Behandlungsmethoden be-

reitstellen zu können. Es ist für die Mitglieder wichtig, dass sie auf Aussagen hinsichtlich der Qualität von medizinischen Einrichtungen, Ärzten und Mediziner wie auch auf alternative Behandlungsmethoden zugreifen können, die von Mitgliedern für andere Mitglieder sind. Diese Aussagen werden im Folgenden als Qualitätssichernde Maßnahmen bezeichnet.

Es fällt auf, dass bei diesen Aspekten Mitglieder und Betreiber unterschiedliche Sichtweisen haben. Ein weiterer wichtiger Punkt ist die Etablierung von Verhaltensregeln. Ebenso wie die vorangegangenen Punkte der Qualitätssichernden Maßnahmen differiert die Ansicht über das Angebot von aktuellen relevanten medizinischen Studien zwischen den Betreiber und den Mitgliedern. Noch anzumerken ist, dass die Beteiligung von Medizinern innerhalb der VPG für die Mitglieder, nicht aber für die Betreiber wichtig ist. Zusammenfassend bestehen Unterschiede in dem, was die Betreiber und in dem was die Mitglieder in Bezug auf elektronische Dienste als wichtig in der VPG ansehen.

4.3 Modell

Für die weitere Analyse hinsichtlich der Analyse für Dienste, die für die betrachteten Zielgruppen von Bedeutung sind, wird auf das Modell (vgl. Abb. 3) der Einflussfaktoren von Komponenten VPG verwiesen, das von Dannecker und Lechner [DaLe06c] entwickelt wurde. Auf Basis des Modells wird erarbeitet, welche Komponenten für die einzelnen Zielgruppen von Bedeutung sind.

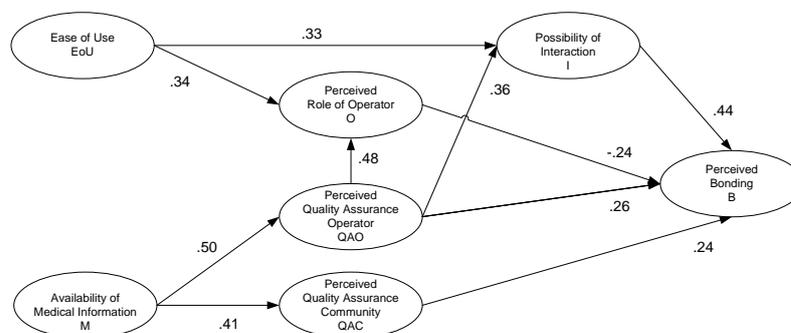


Abb. 3: Einflussfaktoren von Komponenten VPG

Alle latenten Variablen wirken auf die latente endogene Variable „Wahrgenommene emotionale Bindung“ (Perceived Bonding, B). Bemerkenswert ist, dass „Möglichkeit der Interaktion“ (Possibility of Interaction, I) hierbei den höchsten Einfluss hat. „Verfügbarkeit von medizinischen Informationen“ (Availability of Medical Information, M) hat keine direkte kausale Beziehung zu B, dennoch wirkt M nahezu gleich stark (.213) wie die „Wahrgenommene Qualitätssicherung durch die Gemeinschaft“ (Perceived Quality Assurance Community, QAC), „Wahrge-

nommene Qualitätssicherung durch den Betreiber“ (Perceived Quality Assurance Operator, QAO) oder „Wahrgenommene Rolle des Betreibers“ (Perceived Role of Operator, O).

Dies lässt die Interpretation zu, dass eine Verbesserung eines Dienstes hinsichtlich „der Verfügbarkeit medizinischer Inhalte“ (M) auch eine wahrgenommene Verbesserung bei emotionaler Bindung (B) zur Folge hat. Dies spiegelt sich auch in der Bewertung der individuellen Erfolgsfaktoren, die in M eingehen, wider. Der negative Einfluss von O auf B liegt vor allem darin begründet, dass die „Kontinuierliche Überprüfung“ in VPG keine Zustimmung findet. Die „Einfache Nutzung“ (Ease of Use, EoU) der VPG hat - wie M - nur eine indirekte kausale Beziehung zu B. Auswirkungen auf die Entwicklung von elektronischen Diensten in VPG, können auf der Basis des Modells besser abgeschätzt werden. Ein Dienst, der nur medizinisch relevante Informationen zur Verfügung stellt, ist aus der Sicht der Mitglieder nicht hinreichend, sondern viel mehr in einem Kontext qualitätssichernder Maßnahmen durch Betreiber und Mitglieder, zu sehen. Dies ist auch in Übereinstimmung mit der Forschung zu HRQOL (vgl. Kap. 2.1).

4.4 Dienste für jüngere und ältere Mitglieder einer VPG

Da chronische Krankheiten alle Altersschichten betreffen, ist es notwendig, VPG attraktiv für jüngere wie auch ältere Zielgruppen zu gestalten. Im Folgenden werden „Jüngere“ (bis 25 Jahre) und „Ältere“ (über 55 Jahre) Mitglieder betrachtet. Um zu bestimmen, ob eine paarweise Assoziation der einzelnen normal verteilten Variablen besteht, wurden die Variablen auf eine „Bivariate Correlation“ basierend auf „Pearson“ untersucht. Alle Korrelationen zwischen dem Alter der Mitglieder und den untersuchten Faktoren waren signifikant mit $p < .001$. In Abb. 4 sind die Faktoren abgebildet, die in der Sichtweise jüngerer (bis 25 Jahre) und älterer (über 55 Jahre) Mitglieder einen Unterschied in der Wichtigkeit aufweisen, der größer als 0,3 ist.

Bei den Erfolgsfaktoren, die für jüngere Mitglieder wichtiger sind, fallen zwei Punkte auf: Zwei Konstrukte von B, nämlich „Das Gefühl an einem Platz zu sein, an dem Sie sich zu Hause fühlen“ und „Dass man Sie mit Ihren Problemen versteht“ sind für jüngere Mitglieder wichtig, wie auch drei Konstrukte von QAC. Dies sind „Aussagen bezüglich der Qualität von Ärzten und Medizinern durch die Gemeinschaft“, „Alternative Behandlungsmethoden diskutieren zu können“ und „Aussagen bezüglich der Qualität von medizinischen Einrichtungen (z.B. Krankenhäuser) durch die Gemeinschaft“.

Für ältere Mitglieder sind Aspekte wichtiger, wie „Erreichen einer hinreichend großen Mitgliederzahl in kurzer Zeit“, „Die Einbindung von Angehörigen in die Gemeinschaft“ und „Aufrechterhaltung der eigenen Neutralität bei der Präsentation und Auswahl der Angebote“, welche

Teile der Komponente B des Modells sind. Weiter sind wichtig „Individueller Seitenaufbau der Gemeinschaft nach Präferenzen des Mitgliedes“ und „Förderung einer intuitiven Benutzerführung“. Beide sind Teile der Komponente EoU des Modells.

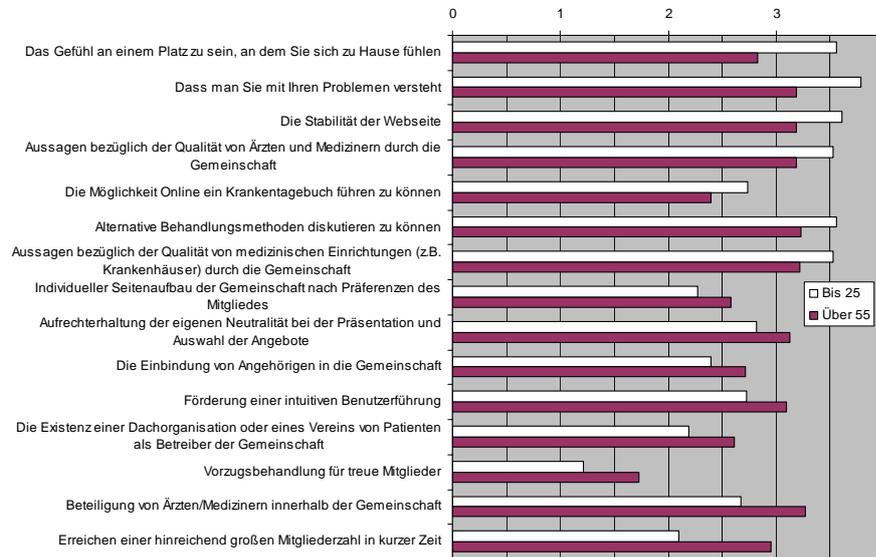


Abb. 4: Hauptunterschiede der Sichtweise jüngerer und älterer Mitglieder VPG (sortiert nach Unterschieden $\geq 0,3$; mit starke Zustimmung 4 – starke Ablehnung 0)

Auch die 10 wichtigsten Faktoren von jüngeren und älteren Mitgliedern unterscheiden sich.

Wichtiger Faktor			
Jüngere Mitglieder (bis 25 Jahre)		Ältere Mitglieder (über 55 Jahre)	
Faktor	Ø	Faktor	Ø
Das Gefühl an einem Platz zu sein, an dem Sie sich zu Hause fühlen	3,79	Dass die Forschung in dem Gebiet Ihrer Krankheit unter Beteiligung der Gemeinschaft vorangetrieben wird	3,56
Die Stabilität der Webseite	3,61	Dass die Mitglieder der Gemeinschaft aktiv an der Qualitätssicherung der Inhalte mitarbeiten	3,30
Alternative Behandlungsmethoden diskutieren zu können	3,56	Alternative Behandlungsmethoden diskutieren zu können	3,23
Aussagen bezüglich der Qualität von medizinische Einrichtungen (z.B. Krankenhäuser) durch die Gemeinschaft	3,53	Aussagen bezüglich der Qualität von medizinische Einrichtungen (z.B. Krankenhäuser) durch die Gemeinschaft	3,21
Aussagen bezüglich der Qualität von Ärzten und Medizinern durch die Gemeinschaft	3,53	Angebot von aktuellen relevanten medizinischen Studien	3,21
Dass die Forschung in dem Gebiet Ihrer Krankheit unter Beteiligung der Gemeinschaft vorangetrieben wird	3,47	Dass man Sie mit Ihren Problemen versteht	3,19
Dass die Mitglieder der Gemeinschaft aktiv an der Qualitätssicherung der Inhalte mitarbeiten	3,35	Aussagen bezüglich der Qualität von Ärzten und Medizinern durch die Gemeinschaft	3,19
Förderung der Interaktion zwischen Mitgliedern	3,22	Die Stabilität der Webseite	3,18
Eine schnelle Reaktionszeit der Webseite	3,21	Eine schnelle Reaktionszeit der Webseite	3,16
		Förderung einer intuitiven Benutzerführung	3,09

Tab 4: Auflistung der 10 wichtigsten Faktoren

Für jüngere Mitglieder sind „Dass man Sie mit Ihren Problemen versteht“ und „Die Stabilität der Webseite“ die wichtigsten zwei Faktoren. Für ältere Mitglieder „Dass die Forschung in dem Gebiet Ihrer Krankheit unter Beteiligung der Gemeinschaft vorangetrieben wird“ und „Dass die Mitglieder der Gemeinschaft aktiv an der Qualitätssicherung der Inhalte mitarbeiten“.

Damit ergeben sich für die unterschiedlichen Komponenten unterschiedliche Wichtigkeiten: Die Komponenten B und QAC sind deutlich wichtiger für die jüngeren Mitglieder als dies der Fall für ältere Mitglieder ist (vgl. Abb. 5). Medizinisch relevante Informationen hingegen sind für ältere Mitglieder wichtiger.

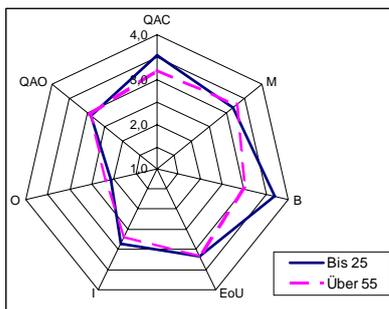


Abb. 5: Durchschnittswerte der Variablen der Komponenten (zur besseren Darstellung beginnt die Abbildung bei 1.0; starke Zustimmung 4 – starke Ablehnung 0)

Zielgruppe	Komponente						
	O	B	EoU	M	QAO	I	QAC
Jüngere Mitglieder	0	2	2	1	1	1	3
Ältere Mitglieder	0	1	3	2	1	0	3

Tab. 5: Anzahl der Faktoren aus den 10 wichtigsten Faktoren abgebildet auf das Modell

In Tab. 5 aufgelistet sind die Komponenten des Modells, mit der Anzahl der Faktoren, die unter den zehn wichtigsten Faktoren der betrachteten Zielgruppen sind (vgl. Tab. 4).

Einträge, die in Tab. 5 grau hinterlegt und fettgedruckt sind, geben an, dass alle Variablen, die Teil der Komponente sind, unter den zehn wichtigsten Faktoren aus Sicht der jeweiligen Zielgruppe sind. Für beide Gruppen von Mitgliedern ist Qualitätssicherung, die durch Mitglieder der VPG durchgeführt wird, von besonderer Bedeutung (vgl. Abb. 5 und Tab. 5). Für jüngere Mitglieder ist die Bindung an eine VPG ein wesentlich wichtiger. Für ältere Mitglieder hingegen ist eine stabile und einfach zu bedienende VPG im Vordergrund.

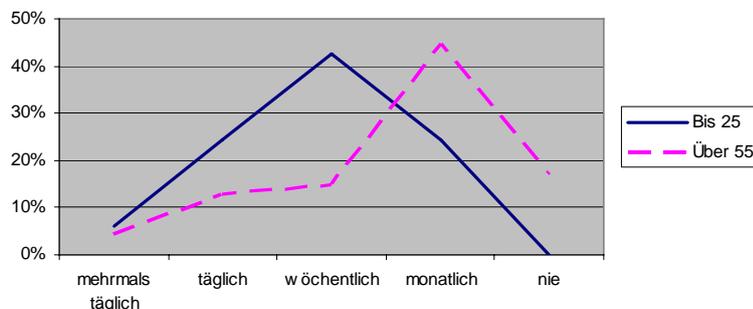


Abb. 6: Verteilung der Aktivität (Beiträge in der VPG)

Abb. 6 zeigt die Aktivität der jüngeren und älteren Mitglieder in Bezug auf die Anzahl der Beiträge, die diese in der VPG tätigen. Mit durchschnittlich einem Beitrag pro Woche, sind die jüngeren Mitglieder aktiver als die älteren Mitglieder, die durchschnittlich nur einmal im Monat einen Beitrag in der VPG leisten. Daraus lässt sich schließen, dass gerade für ältere Mitglieder andere Dienste notwendig sind, um ihnen Beiträge zu VPG zu ermöglichen und ihre Erfahrungen zur Verfügung zu stellen.

	Bis 25	Über 55
Zeit in VPG (h pro Woche)	3	2
Nehmen Sie an Selbsthilfegruppen (SHG) (Sehr oft 1 – Ab und zu 3 – Niemals 5)	4,3939	3,5745
Kennen sie Mitglieder im realen Leben? Ja / Nein	40% / 60%	25% / 75%
Stellen Sie Fragen in der VPG, die Sie einem Arzt nicht stellen (Sehr oft 1 – Ab und zu 3 – Niemals 5)	2,64	4,0
Spielt die VPG eine zentrale Rolle in Ihrem Leben (Stimme stark zu 1 – Unentschieden 3 – Lehne stark ab 5)	2,46	3,0

Tab. 6: Unterschiede in Demographischen Daten zwischen jüngeren und älteren Mitgliedern

Jüngere Mitglieder verbringen 50% mehr Zeit in den VPG als ältere Mitglieder, was sich implizit auch in der höheren Aktivität widerspiegelt und dem Umstand, dass die VPG eine zentralere Rolle im Leben der jüngeren Mitglieder spielt. Ältere Mitglieder nehmen wesentlich öfter an Selbsthilfegruppen teil als jüngere Mitglieder. Interessanterweise kennen anteilig mehr jüngere Mitglieder andere Mitglieder im realen Leben als dies bei älteren Mitgliedern der Fall ist. Der größte Unterschied allerdings findet sich in der Quote, wie oft sie Fragen in ihrer VPG stellen, die sie einem Arzt nicht stellen würde. Diese Art der Fragen wird wesentlich häufiger von jüngeren Mitgliedern gestellt. Für jüngere Mitglieder sind daher Möglichkeiten der Interaktion zentral. Beiden Zielgruppen sind gemeinsam, dass Qualitätssicherung, die durch Mitglieder durchgeführt wird, sehr wichtig ist. Aus diesem Grund wurde gemeinsam mit einem Partner ein Bewertungsmechanismus gestaltet und implementiert, der Mitglieder einer VPG in die Lage versetzt, Qualitätsaussagen auf Basis von Bewertungen durchzuführen.

5 Konzeption eines Dienstes: Ein generisches Bewertungssystem

Da Qualitätssicherung durch die Mitglieder ein wesentlicher Punkt für VPG ist (vgl. Abb. 2), wurde ein elektronischer Dienst entwickelt und in einem studentischen Projekt auch implementiert. Dieser Dienst wurde gemeinsam mit der Deutsche Morbus Crohn / Colitis ulcerosa Vereinigung DCCV e.V. (www.dccv.de, DCCV) konzipiert und ist seit Sommer 2006 in der Online Gemeinschaft der DCCV verfügbar. Mit diesem Dienst kann die DCCV einfach Online Bewer-

tungen entwickeln, realisieren und die Ergebnisse auswerten. Die Mitglieder erhalten die Möglichkeit, ihre Erfahrungen strukturiert beizutragen. Im Folgenden sollen kurz wesentliche Punkte aus der Gestaltung des Dienstes dargestellt werden.

Die Ziele, die mit diesem Dienst erreicht werden sollen, sind: Mit Online Bewertungen von medizinischen Institutionen oder Informationen soll das Informationsangebot des DCCV verbreitert werden und eine Interaktionsmöglichkeit geschaffen werden, die auch Mitgliedern, die nicht im Forum aktiv sind und sein wollen (v.a. auch älteren Mitgliedern, vgl. Kap. 4.4), eine strukturierte Möglichkeit gibt, zur VPG beizutragen. Die Entwicklung der Bewertungskriterien und die Auswahl, welche Informationen durch die VPG bewertet werden, kann Thema eines gemeinschaftsinternen Prozesses sein. Man erhofft sich, dass durch strukturierte Bewertungen die Diskussion von medizinisch relevanten Themen in einer Gemeinschaft ein Stück weit objektiver und qualitativ hochwertiger wird und damit die Gemeinschaften auch mehr Informationen im medizinischen Bereich anbieten können.

Entwickelt wurde ein generischer Dienst für die Gestaltung, die Durchführung und Evaluation von Online Bewertungen durch eine VPG. Nachdem die Dienstnutzung spezifisch für unterschiedliche Arten von Gemeinschaften ist [DaLe06c; DaLe06a] und sich in einer Vorstudie herausgestellt hat, dass sich auch VPG untereinander unterscheiden, musste ein generischer Dienst, der durch die VPG selbst anpassbar ist, entwickelt werden.

Der hierbei entwickelte Dienst wurde als eine Erweiterung zu Typo3 implementiert [AlFH06]. Die Erweiterung besteht aus einem Modul, das die generische Erstellung des Fragenkatalogs auf Administrator-Seite ermöglicht und einem Plugin, das die generische Darstellung des Fragenkatalogs und der Resultate auf den Typo3-Seiten realisiert.

Der Dienst bietet für eine Bewertung (eine Bewertung errechnet sich aus einer Menge an Einzelbewertungen) folgende Gestaltungsmöglichkeiten:

- Eine Einzelbewertung kann durch andere Mitglieder bewertet werden, ob die Einzelbewertung als hilfreich angesehen wird oder nicht.
- Eine Bewertung wird erst nach einer Mindestanzahl an Einzelbewertungen angezeigt. Damit soll die Aussagekraft der Bewertungen erhöht werden – einzelne, möglicherweise tendenzielle Aussagen erscheinen nicht und werden gegebenenfalls durch mehrere unabhängig abgegebene Bewertungen relativiert.

- Eine Bewertung wird erst angezeigt, wenn sie besser als ein individuell einstellbarer Schwellwert ist. So können schlechte, möglicherweise ungerechtfertigte Bewertungen ausgefiltert werden. Nur positive Bewertungen werden somit angezeigt.
- Es können nur registrierte Benutzer Bewertungen abgeben. Benutzernamen können bei der Darstellung der Bewertung und Einzelbewertungen ausgeblendet werden. Damit kann den Bewertenden Anonymität zugesichert werden und – wenn nötig – ist sichergestellt, dass die Personen hinter den Bewertungen in der Gemeinschaft bekannt sind. Dies soll die Seriosität der Bewertungen erhöhen.
- Bewertungen können auf Basis des Alters gefiltert werden. Damit wird die Tendenz der Entwicklung einer Bewertung sichtbar.

Für die Ausgestaltung der Fragen existieren drei Ausprägungen: Text-Frage, deren Antwort nicht in die Bewertung eingeht, Ja/Nein Fragen und Skalen Fragen. Bei Skalen Fragen wird eine Likert-Skala zu Grunde gelegt und die verbale Repräsentation der Zahlenwerte wird individuell für jede Frage definiert. Für die Fragen können eine Reihe von Parametern angegeben werden:

- Optionale Frage: Muss eine Frage beantwortet werden?
- Gewichtung: Bei der Erstellung des Fragebogens kann eine Gewichtung für jede Frage angegeben werden. Es kann auch festgelegt werden, ob bei der Antwort eine Gewichtung der Frage durch den Bewertenden festgelegt werden kann. Ferner kann die Art der Gewichtung der Frage durch den Bewertenden festgelegt werden. Somit können verschiedene Gewichtungen unterschiedlicher Aspekte in einem Fragebogen erfasst werden. Ein Beispiel: Bei der Bewertung von Reha-Kliniken sind Gesprächsrunden für Patienten ein Kriterium – dies wird bei der Erstellung des Fragebogens festgelegt. Es gibt Mitglieder in der Gemeinschaft, für die Gesprächsrunden eher unwichtig sind. Sie können ihre Antwort entsprechend für ihre Einzelbewertung gewichten. Mitglieder, für die Gesprächsrunden wichtig sind, bewerten die Qualität der Runden und geben eine entsprechend hohe persönliche Gewichtung an.

Der mit der DCCV gemeinsam entwickelte Dienst versetzt Betreiber von VPG in die Lage, elektronische Dienste zur Qualitätssicherung den Mitgliedern zur Verfügung zu stellen. Diese elektronischen Dienste können an beliebige medizinische Inhalte angepasst werden.

6 Zusammenfassung

VPG benötigen neue Dienste, um mehr und zielgerichtete Informationen für die Mitglieder und die Betroffenen anbieten zu können, um den Zusammenhalt der VPG besser zu stärken und mehr für die Mitglieder und die Betroffenen im Gesundheitswesen erreichen zu können. Die Studie gibt Aussagen darüber, was Mitglieder und Betreibern dieser Gemeinschaften wichtig ist. Das Modell zeigt auf, wie sich die emotionale Bindung, das zentrale Charakteristikum virtueller Gemeinschaften verstärken lässt. Bemerkenswert ist, dass das Anbieten medizinischer Informationen und die Qualitätssicherung durch Mitglieder als wichtig angesehen wird, es aber kaum Beispiele dafür in existierenden VPG gibt. Im vorliegenden Papier wurden außerdem die Unterschiede zwischen jüngeren und älteren Mitgliedern von VPG analysiert und aufgezeigt, welche elektronischen Dienste für diese speziellen Zielgruppen von Interesse sind, die die Akzeptanz und Nutzung von VPG fördern. Medizinische Informationen, eine gute, einfach zu nutzende Infrastruktur und ein personalisierter Zugang zur VPG ist für ältere Mitglieder wichtig. Für jüngere Mitglieder ist die Interaktion in der Gemeinschaft wichtig. Dies zeigt sich darin, dass jüngere Mitglieder im Verhältnis mehr Bekanntschaften im realen Leben pflegen. Ein Event-Management, das Treffen in der realen Welt unterstützt, ist für jüngere Mitglieder positiv. Außerdem hat die Interaktion auch einen positiven Einfluss auf die Bindung zu einer VPG.

Im Papier wurden wesentliche Gestaltungsmerkmale eines Dienstes kurz dargestellt. Dieser Dienst ist ein generischer Bewertungsmechanismus, der im Rahmen von VPG für beliebige Inhalte Einsatz finden kann. Qualitätssicherung im medizinischen Bereich durch die Mitglieder von VPG ist ein wichtiger Erfolgsfaktor. Dieser Dienst ermöglicht einer VPG, eigene Bewertungsdienste für medizinische Inhalte zu entwickeln und damit qualitätssichernde Maßnahmen eigenständig durchzuführen. Vier Aspekte sind in unserem Konzept wesentlich: (1) Die VPG kann den Inhalt bestimmen, der durch eine VPG, respektive der Mitglieder, auf Qualität (in Form einer Bewertung) überprüft wird. (2) Die VPG kann die Mitglieder aktiv in die Entwicklung eines Fragenkatalogs einbinden. Dies kann sich positiv auf eine Gemeinschaft und ihr Zusammengehörigkeitsgefühl, auswirken. (3) Die Mitglieder werden - wenn dies im Fragebogen festgelegt wird - auch in die Lage versetzt, Bewertungen auf ihre Qualität hin zu bewerten, indem sie eine Bewertung als hilfreich oder nicht hilfreich kennzeichnen. Damit wird die Aussagekraft eines Bewertungssystems erhöht. (4) Durch vordefinierte Strukturen wird es insbesondere älteren Mitgliedern, die nicht an den offenen Diskussionen im Forum teilnehmen wollen, einfach gemacht, beizutragen. Betreibern einer VPG wird durch diesen Dienst ein Baustein ge-

liefert, die Mitglieder zu aktivieren und mehr Beiträge von Mitgliedern zu erhalten und die Mitglieder auch damit stärker an die VPG zu binden. Ebenso wird das Vertrauen in die durch eine VPG bereitgestellten Inhalte gestärkt und die, von den Mitgliedern geforderte Unabhängigkeit der VPG untermauert. Geplant ist sowohl diesen Dienst weiterzuentwickeln als auch weitere Dienste auf Basis der Ergebnisse der Umfrage zu entwickeln und implementieren.

Danksagung: Wir bedanken uns bei allen Teilnehmern der Studie, bei H. Krcmar, J.M. Leimeister und P. Sidiras für die Bereitstellung ihres Fragebogens und bei unserem Kooperationspartner Deutsche Morbus Crohn / Colitis ulcerosa Vereinigung DCCV e.V.

Literaturverzeichnis

- [AlFH06] *Altmann, Werner; Fritz, Rene; Hinderink, Daniel*: Typo3: Enterprise Content Management. Open Source Press 2006.
- [BoTD02] *Borgaonkar, M.R.; Townson, G.; Donnelly, M.*: Providing Disease-Related information worsens Health-Related Quality of Life in Inflammatory Bowel Disease. In: *Inflamm Bowel Disease* 8 (2002), S. 264-269.
- [Borg04] *Borgetto, Bernhard*: Selbsthilfe und Gesundheit - Analysen, Forschungsergebnisse und Perspektiven. Verlag Hans Huber, Bern 2004.
- [DaLe04] *Dannecker, Achim; Lechner, Ulrike*: "Virtual Communities with a Mission" in the Health Care Sector. In: *Klein, Stefan* (Hrsg.): 11th Research Symposium on Emerging Electronic Markets (RSEEM 2004) (2004), S. 115-128.
- [DaLe06a] *Dannecker, Achim; Lechner, Ulrike*: An empirical analysis of the demand for e-services for virtual communities of patients. In: *Walden, Pirkko; Markus, M. Lynne; Gricar, Joze; Pucihar, Andreja; Lenart, Gregor* (Hrsg.): 19th Bled eConference (2006), S. 18.
- [DaLe06c] *Dannecker, Achim; Lechner, Ulrike*: Success Factors of Communities of Patients. In: *Ljungberg, Jan; Andersson, Magnus* (Hrsg.): 14th European Conference on Information Systems (2006), S. 12.
- [Figa98] *Figallo, Cliff*: Hosting Web Communities: Building Relationships, Increasing Customer Loyalty, and Maintaining a Competitive Edge. John Wiley & Sons, Inc. 1998.
- [Gold05] *Goldschmidt, Peter C.*: HIT and MIS: Implications of Health Information Technology and Medical Information Systems. In: *Communications of the ACM* 48 (2005) 10, S. 69-74.
- [HaAr97] *Hagel III, John; Armstrong, Arthur G.*: Net gain: expanding markets through virtual communities. Harvard Business School Press 1997.
- [HuTa05] *Hulstijn, Joris; Tan, Yao-Hua*: Design Aspects of a Personalized Information System about Healthcare Regulations. In: *Tan, Yao-Hua* (Hrsg.): 12th Research Symposium on Emerging Electronic Markets (RSEEM 2005) (2005), S. 135-149.
- [JKGM05] *Janke, K.H.; Klump, B.; Gregor, M.; Meisner, C.; Haeuser, W.*: Determinants of life satisfaction in inflammatory bowel disease. In: *Inflamm Bowel Disease* 11 (2005) 3, S. 272-286.

- [Jose04] *Josefsson, Ulrika*: Patients Creating Self-Help on the Internet - Lessons for Future Design of Internet Based Healthcare Resources. In: *Sprague, Ralph H.* (Hrsg.): 37th Annual Hawaii International Conference on System Sciences (HICSS'04) (2004), S. 9.
- [KaTa05] *Kartseva, Vera; Tan, Yao-Hua*: Designing Controls for a Marketplace of Health Care services: a Case Study. In: *Tan, Yao-Hua* (Hrsg.): 12th Research Symposium on Emerging Electronic Markets (RSEEM 2005) (2005), S. 116-134.
- [KRNR03] *Kennedy, Anne; Robinson, Andrew; Nelson, E; Rogers, A; Reeves, D; Sculpoher, M; Richardson, G; Thompson, David; Roberts, C*: A randomised controlled trial to assess the impact of a package comprising a patient-orientated, evidence-based self-help guidebook and patient-centred consultations on disease management and satisfaction in inflammatory bowel disease. In: *Health Technology Assessment* 7 (2003) 28, S. 140.
- [KADL02] *Krcmar, Helmut; Arnold, Y.; Daum, Miriam; Leimeister, Jan Marco*: Virtual communities in health care: the case of "krebsgemeinschaft.de". In: *SIGGROUP Bull.* 23 (2002) 3, S. 18-23.
- [LeDK04] *Leimeister, Jan Marco; Daum, Miriam; Krcmar, Helmut*: Towards mobile communities for cancer patients: the case of krebsgemeinschaft.de. In: *International Journal of Web Based Communities* 2004 1 (2004) 1, S. 58-70.
- [LeSK04] *Leimeister, Jan Marco; Sidiras, Pascal; Krcmar, Helmut*: Success Factors of Virtual Communities from the Perspective of Members and Operators: An Empirical Study. In: *Sprague, Ralph H.* (Hrsg.): 37th Annual Hawaii International Conference on System Sciences (HICSS'04) (2004), S. 10.
- [McFa05] *McLure Wasko, Molly; Faraj, Samer*: Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. In: *MIS Quarterly* 29 (2005) 1, S. 35-57.
- [NoPA04] *Nonnecke, Blair; Preece, Jenny; Andrews, Dorine*: What Lurkers and Posters Think of Each Other. In: *Sprague, Ralph H.* (Hrsg.): 37th Annual Hawaii International Conference on System Sciences (HICSS'04) (2004), S. 9.
- [NoPr01] *Nonnecke, Blair; Preece, Jennifer*: Why Lurkers Lurk. In: *Strong, D.; Straub, D.* (Hrsg.): *Americas Conference on Information Systems* (2001), S. 1521-1531.
- [PUCS06] *Pratt, Wanda; Unruh, Kenton; Civan, Andrea; Skeels, Meredith*: Personal Health Information Management. In: *Communications of the ACM* 49 (2006) 1, S. 51-55.
- [Pree00] *Preece, Jennifer*: *Online Communities: Designing Usability and Supporting Socialbilty.* John Wiley & Sons, Inc. 2000.
- [Rhei94] *Rheingold, Howard*: *Virtual Community: Homesteading on the Electronic Frontier.* HarperTrade 1994.
- [Sule00] *Suler, John*: Bringing Online and Offline Living Together: The Integration Principle. In: *The Psychology of Cyberspace.* www.rider.edu/suler/psyber/integrate.html (2005).
- [Well01] *Wellmann, Barry*: Computer networks as social networks. In: *Science* 293 (2001) 14, S. 2031-2034.
- [WiOD97] *Whittaker, Steve; Isaacs, Ellen; O'Day, Vicki*: Widening the net: workshop report on the theory and practice of physical and network communities. In: *SIGCHI Bull.* 29 (1997) 3, S. 27-30.
- [Yaho05] *Yahoo*: Health & Wellness.
http://health.dir.groups.yahoo.com/dir/Health___Wellness/, 2005, Abruf am: 29.04.2005.

Formalisierung und Automatisierung von SOPs in der Intensivmedizin

Martin Sedlmayr, Thomas Rose

Fraunhofer-Institut für angewandte Informationstechnik
53754 St. Augustin
martin.sedlmayr@fit.fraunhofer.de

Rainer Röhrig, Markus Meister

Klinik für Anaesthesiologie, Intensivmedizin, Schmerztherapie
Universitätsklinikum Gießen und Marburg – Standort Gießen
35392 Gießen
rainer.roehrig@chiru.med.uni-giessen.de

Achim Michel-Backofen

Ressort Klinische und Administrative Datenverarbeitung
Universitätsklinikum Gießen und Marburg – Standort Gießen
35392 Gießen
achim.michel-backofen@akad.uniklinikum-giessen.de

Abstract

Dieser Beitrag präsentiert einen Ansatz zur Prozessunterstützung in der Medizin. Ziel des Ansatzes ist die Erfassung medizinischen Wissens über Vorgehensweisen (Standard Operating Procedures) und die effektive Unterstützung medizinischen Personals in der Befolgung dieser Vorgehensweisen, da es empirisch belegt ist, dass sich die Qualität der Versorgung von Patienten durch die Umsetzung von Standard Operating Procedures erhöht [Mart06]. Zudem ermöglicht die Nutzung standardisierter Vorgehensweisen eine quantitative Analyse und schrittweise Fortschreibung zur Qualitätsverbesserung.

Der hier vorgestellte Ansatz basiert auf einem mehrstufigen Modellierungskonzept, das zunächst auf einer textuellen Ebene Vorgehensweisen konsolidiert und sie dann schrittweise formalisiert bis hin zu Ausführungsvorschriften, die durch Vorgangunterstützungstechniken automatisiert werden.

1 Einführung

Kliniken sehen sich durch die Einführung eines fallpauschalenbasierten Entgeltsystems durch Diagnosis Related Groups (DRG) mit einem zunehmenden Druck zur Qualitätssicherung der Behandlungen und zur Einschränkung der Kosten konfrontiert [MSKK03]. Ein Kernelement in der Lösung dieser Herausforderung ist Prozesswissen über sowohl klinische (technisch-administrative) als auch medizinische (fachliche) Vorgehensweisen. Es konnte bereits gezeigt werden, dass standardisierte Behandlungspfade geeignet sind, sowohl die Effizienz und Wirtschaftlichkeit, als auch die Qualität der medizinischen Leistungen zu erhöhen [BHBW00; MDHB96; SEGO95]. Dies spiegelt sich etwa in einer verkürzten Liegezeit des Patienten oder einer signifikanten Reduktion der Mortalität (Sterblichkeit) wider.

Während sich die Unterstützung klinischer Abläufe im Rahmen von Clinical Pathways (siehe Kapitel 2) bereits in der Etablierung befindet, ist die Unterstützung medizinischer Prozesse durch klinische Informationssysteme (KIS) noch weitestgehend offen. Dies liegt zum einen an der hohen Komplexität der medizinischen Prozesse, als auch an einer mangelnden Flexibilität von Vorgangunterstützungssystemen. Zudem ist zumeist nur eine unzureichende Datenbasis als Entscheidungsgrundlage verfügbar. Eine Unterstützung und Assistenz in der Ausführung der Prozesse findet bisher nur für solche Behandlungen statt, die strukturell eindeutig beschreibbar sind und zumeist aus wiederkehrenden Aktivitäten bestehen, z.B. der Therapie chronischer Erkrankungen, wie die Behandlung mit Chemotherapiezyklen, deren Unterstützung eine Domäne von ONCOCIN darstellt [Muse05; SSBC84]. Verlässt man aber das Gebiet dieser strukturierten Vorgehensweisen, so schwindet jede Form der Unterstützung ausgenommen die Dokumentation der Behandlung im Kontext der (elektronischen) Patientenakte.

Ziel unseres Projektes Online Guideline Assist (OLGA) ist die Unterstützung medizinischer Prozesse durch ein wissensbasiertes und prozessorientiertes Assistenzsystem in der Intensivmedizin. Unser Ansatz besteht darin, (1) vorhandene Leitlinien zu sichten, (2) zu bewerten und zu formalisieren, (3) sie in einer Ausführungsumgebung zu automatisieren und (4) den Effekt der Unterstützung zu messen. Bei der Realisierung sollen existierende Standards zur Formalisierung und Automatisierung bewertet und so weit wie möglich genutzt werden, d.h. der Fokus liegt auf dem lauffähigen System und nicht so sehr auf der Entwicklung einer neuen Prozessbeschreibungssprachen für medizinische Vorgehensweisen.

Im Folgenden werden dabei die Ausgangssituation in der Intensivmedizin dargestellt (Kapitel 2), die bestehenden Standards zur Prozessmodellierung, Formalisierung und Unterstützung diskutiert (Kapitel 3), unser Ansatz der mehrstufigen Modellierung vorgestellt (Kapitel 4) sowie die Erfahrungen bei der Umsetzung unseres Ansatzes beschrieben (Kapitel 5).

2 Ausgangssituation: Therapierichtlinien in der Intensivmedizin

2.1 Intensivmedizin

In der Intensivmedizin werden kritisch kranke Patienten überwacht und behandelt. Intensivpatienten besitzen meist mehrere Erkrankungen oder Verletzungen, die häufig einen Ausfall lebenswichtiger Organsysteme zur Folge haben, der durch Unterstützungs- und Ersatzverfahren - wie z.B. kreislaufunterstützende Therapien, künstliche Beatmung, Nieren- oder Leberdialyse - zeitweise kompensiert werden muss.



Abbildung 1: Integrierter Intensivarbeitsplatz mit Vitaldatenüberwachung, Beatmungsgerät und Patienten-Daten-Management-System, Rechts Dialysegeräte als Nierenersatzverfahren

Aufgrund des eingeschränkten Allgemeinzustandes und der invasiven Therapien kann es jederzeit zu Befundänderungen und Interaktionen zwischen den einzelnen Erkrankungen und Therapien kommen. Intensivmedizin stellt somit ein komplexes Interaktionssystem zwischen Patient und Behandlung mit hoher Dynamik in einem zeitkritischen Hochrisikobereich dar. Dies erfordert eine gut strukturierte interdisziplinäre Zusammenarbeit. Demzufolge zählt die Intensivmedizin zu den kostenintensivsten Bereichen der stationären Patientenversorgung.

2.2 Klinische Behandlungspfade, Standard Operating Procedures

Bei den diagnostischen oder therapeutischen Standards unterscheidet man zwischen offiziellen und innerklinischen Handlungsanweisungen. Die offiziellen Handlungsanweisungen werden von der Bundesärztekammer nach ihrer justiziablen Verbindlichkeit in „Empfehlungen“, „Leitlinien“ und „Richtlinien“ unterschieden und von den medizinischen Fachgesellschaften in der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) er-

stellt und veröffentlicht [Bund98]. Bei den innerklinischen Standards werden „klinische Behandlungspfade“ (Clinical Pathways) und „Standard Operating Procedures“ (SOPs) unterschieden. Während klinische Behandlungspfade den vollständigen Diagnostik- und Therapieplan für einen Krankenhausaufenthalt in Abhängigkeit von einer Diagnose oder einem Symptom abbilden, stellen SOPs diagnose-, symptom- oder maßnahmenbezogene Indikations- und Handlungsanweisungen dar [CHBP98, <http://www.bda.de>]. Der englische Terminus „Guideline“ wird als undifferenzierter Oberbegriff für die verschiedenen offiziellen und innerklinischen Standards verwendet.

Die Intensivmedizin bietet aufgrund der hohen Kosten und der zahlreichen, oft komplex ineinander verzahnten Prozessabläufen ein natürliches Einsatzgebiet von SOPs, da mit deren Hilfe Arbeitsschritte verzögerungsfreier und störungsstabiler angegangen werden können. Ziel ist es dabei, mit hoher Qualität eine effiziente und wirtschaftliche Leistung zu erbringen, um eine möglichst optimale Behandlung des Patienten (niedrige Sterblichkeit / Letalität) zu angemessenen Kosten (kurze Liegedauer) zu erbringen.

2.3 Entwicklung und Einschränkungen von SOPs

Basis für die Entwicklung von SOPs ist evidenzbasierte Medizin und der von den Experten der Fachgesellschaften erarbeiteten Maßstäbe [MSKK03]. Weltweit sind mehrere tausend Guidelines definiert, die in nationalen oder fachgesellschaftlichen „Clearinghouses“ im Internet publiziert werden (z.B. bei <http://www.guidelines.gov> und <http://www.leitlinien.de>). Dabei müssen SOPs an örtliche Gegebenheiten und bereits existierende Abläufe angepasst werden. Daraus resultiert, dass eine SOP für andere Abteilungen oder Krankenhäuser nur als Vorlage dienen kann die einer individuellen Ausarbeitung bedarf. Trotzdem stellt die Veröffentlichung von SOPs einen enormen Wissensaustausch und eine Arbeitserleichterung für das Erstellen eigener Standards dar. In den Tauschbörsen wird die überwiegende Zahl der SOPs als strukturierter Fließtext präsentiert. Dies führt bei der Formalisierung zu dem Problem, dass zur Überführung in medizinische Algorithmen eine weitere Präzisierung notwendig wird, die ein hohes Maß an Aufwand und Fachkompetenz erfordert. Zudem erfordert eine Formalisierung als Vorbereitung einer Automatisierung weitere Präzisierungsschritte, da hier auch die Anpassung an die lokale Software, Datenstruktur und Semantik (Bezeichnung, Einheiten) erfolgen muss.

Es gibt bereits seit den 70er Jahren die Möglichkeit zur Unterstützung medizinischer SOPs durch DV-Systeme [SSBC84]. Die Systeme waren meist Standalone-Expertensysteme, die sich nicht in der Routine durchsetzen konnten. Eine Ursache hierfür ist die mangelnde Faktengrundlage, da es den meisten klinischen Informationssystemen bisher an der vollständigen, zeitnahen

Erfassung aller Patientendaten (Befunde, Therapiemaßnahmen, Medikation, etc.) mangelt. Eine zweite Ursache ist die mangelnde Einbindung der Assistenzsysteme in die ärztlichen Arbeitsprozesse: Die meisten klinischen Informationssysteme unterstützen vor allem die Erfassung administrativer und organisatorisch relevanter Daten. Eine Integration in den Behandlungszyklus (Diagnostik, Befund, Therapieentscheidung, Therapieüberwachung) findet bei den herkömmlichen Systemen nicht statt. Eine dritte Ursache ist die mangelnde Formalisierung der SOPs, die zur automatischen Ausführung notwendig ist.

2.4 DV-Umgebung auf der Intensivstation

Mit dem Einsatz moderner Patienten-Daten-Management-Systemen (PDMS), die den Anspruch auf eine vollständige Patientendokumentation haben, entsteht die Möglichkeit, ein aktives Unterstützungssystem auf Basis einer umfangreichen Faktenlage in den ärztlichen Arbeitsprozess zu integrieren.

Ein Repräsentant eines PDMS ist das, seit 1999 auf der Operativen Intensivstation des Universitätsklinikum Gießen und Marburg Standort Gießen implementierte, Intensiv-Informationen-Daten-Management-System (IMS) ICUData (IMESO GmbH, Hüttenberg). Dieses PDMS ist mit dem Vitaldatenmonitoring des einzelnen Patienten und dem Beatmungsgerät vernetzt und kommuniziert bidirektional über HL7 mit den klinischen Informationssystemen.

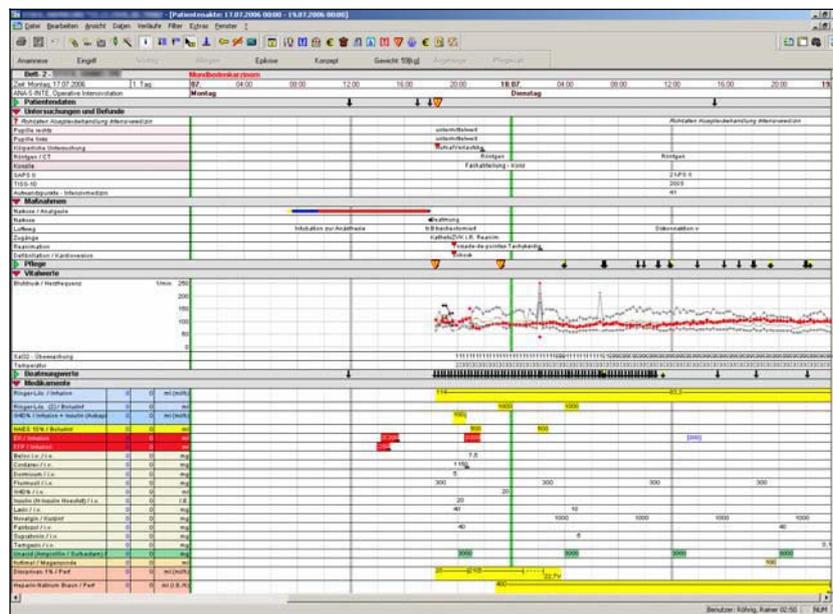


Abbildung 2: Darstellung der vollständigen Patientendaten in der Benutzeroberfläche des PDMS.

Health Level 7 (HL7) ist ein internationaler Standard für den Austausch von Daten zwischen Computersystemen im Gesundheitswesen (<http://www.hl7.org>). Es erfolgt ein automatischer Import der Diagnosen und Prozeduren, sowie der Labor-, Radiologie- oder Mikrobiologiebefun-

de. Auf der Station werden die vollständige Basis- und Verlaufsdokumentation als auch die gesamte Therapie- und Pflegeplanung in dem PDMS durchgeführt (Abbildung 2). Damit stehen in dem PDMS eine Vielzahl der zur Auswahl und Einhaltung von Standardprozeduren benötigten Informationen zur Verfügung, die in Verbindung mit einer Wissensbank eine teilautomatisierte Unterstützung zur Auswahl und Einhaltung von SOPs ermöglichen.

3 Prozessunterstützung

Durch die zunehmende Verbreitung von EDV-Arbeitsplätzen in Kliniken liegt der Einsatz von wissensbasierten und prozessorientierten Systemen zur Unterstützung des ärztlichen und pflegerischen Personals nahe. Trotz der genannten Vorteile haben solche Unterstützungssysteme bisher keinen Eingang in die klinische Routine der Intensivmedizin gefunden.

3.1 Modellierung automatisierbarer SOPs

Seit dem Ende der neunziger Jahre wurden verschiedene Ansätze zur computergestützten Modellierung klinischer SOPs entwickelt (Computerized Guidelines). Sie reichen von Hypertext-basierten Dokumentensammlungen bis hin zu klassischen, entscheidungsunterstützenden Systemen (<http://www.openclinical.org>). Allen Ansätzen gemeinsam ist, dass sie versuchen, das medizinische Wissen einer Guideline möglichst vollständig zu formalisieren, d.h. die Entscheidungs-bäume mit den dazugehörigen Regeln, die Abfolge von Aktivitäten, sowie die assoziierten Datenmodelle.

Es werden verschiedene Formalismen und Techniken verwendet: regelbasiert (Arden Syntax), logik-basiert (PROforma), netzwerk-basiert (PRODIGY) oder vorgangs-basiert (GUIDE) [PTBC03]. Dabei erhebt keiner dieser Formalismen Anspruch auf eine universelle Lösung. Vielmehr konzentrieren sich die einzelnen Forschungsprojekte auf spezifische Problemstellungen: Asbru betrachtet Aspekte der Planung und der zeitlichen Entwicklung, PROforma basiert auf einer formal verifizierbaren Logiksprache, während GLIF die besten Aspekte anderer Projekte zu kombinieren versucht und insbesondere in frühen Versionen den Austausch von Guidelines zum Ziel hatte.

Ultimativ sollen die formal erstellten Modelle von einer Ausführungsumgebung (Execution Engine) automatisch unterstützt werden können. Während klassische entscheidungsunterstützende Systeme hierbei hauptsächlich die eigentliche Entscheidung unterstützen, zielen viele Guidelinesysteme auch auf die Unterstützung bei der Ausführung der notwendigen und empfohlenen Aktivitäten im Sinne der klassischen Vorgangsunterstützung.

3.2 Vorgangunterstützung

Im Bereich der Office-Automation und der Produktion unterstützen seit vielen Jahren Workflowsysteme die automatische Ausführung von Arbeitsprozessen. Zunächst wird ein Modell aus Prozessen, Aktivitäten und Ressourcen spezifiziert. Dieses Modell kann durch eine Laufzeitumgebung interpretiert und damit anstehende Aufgaben und Daten den zuständigen Bearbeitern (Benutzer oder Applikationen) zugeleitet werden. Es existieren zahllose Standards (siehe z.B. www.ebpm1.org oder www.bpm-guide.de), die sich in der Unterstützung durch Produkte, Firmen, Institutionen und fachlichen Schwerpunkten unterscheiden [AHKB003].

In der Anwendung können Workflowsysteme vor allem dort erfolgreich eingesetzt werden, wo standardisierte, wohldefinierte Arbeitsabläufe existieren. In vielen Anwendungsfällen und Domänen ist dies jedoch nicht der Fall, z.B. weil die Prozesse nicht vollständig a-priori spezifiziert werden können (Komplexität) oder weil zu viele situationsspezifische Abhängigkeiten den Prozess steuern (Dynamik) [HHJN99]. Daher ist besonders die Frage der Flexibilisierung Gegenstand intensiver Forschung. Hier werden unterschiedliche Ansätze verfolgt, von denen wir nur einige kurz betrachten.

- Bestehende Workflows werden während der Ausführung geändert: Dabei reagiert die Ausführungsumgebung auf vorab definierte Ereignisse und Zustände und verändert anhand einer Liste von Regeln die laufende Workflowinstanz. Bekannter Vertreter dieser Regeln sind Event Condition Action Rules (ECA), wie sie z.B. in AgentWork [MüGR04] verwendet werden.
- Im Prozessmodell stellen generische Prozesse Platzhalter für Prozessfragmente, die während der Ausführung aus einer Liste alternativer Implementierungen dynamisch ersetzt werden (z.B. Worklets in YAWL [AHEA05]).
- Statt eines fest definierten Prozessmodells werden Prozessfragmente unter Ausnutzung von Regeln und Constraints während der Laufzeit dynamisch zusammengesetzt [DYWH04; MaSa02] oder ihre Konfiguration wird interaktiv unterstützt [RuPR99].
- Bei der Fallbehandlung wird der explizierte Ablauf von Prozessen durch ein implizites Routing ersetzt, d.h. Aktivitäten werden ausgeführt, wenn ihre Vorbedingungen erfüllt sind [ScWe02; AaWG05].

3.3 Einschränkungen

Klinische Guidelines sind keine vollständigen Prozessbeschreibungen im eigentlichen Sinne. Sie stellen vielmehr ein Regelwerk medizinischen Wissens dar, das Prozessfragmente enthält.

Bisherige Projekte konzentrieren sich vornehmlich auf die Modellierung des Expertenwissens und untersuchen damit Probleme der Planung, temporalen Logik und Datenmodellierung. Sie befassen sich jedoch (noch) nicht mit der Spezifikation zur Ausführung benötigter Laufzeiteigenschaften.

Obwohl es nahe liegt, Techniken zur Implementierung einzusetzen, die sich bereits im wirtschaftlichen und industriellen Umfeld bewährt haben, fußt keines der Guidelinesysteme auf bekannten, klassischen Workflowstandards oder Werkzeugen. Lauffähige Ausführungsumgebungen sind – falls existent – vornehmlich wissenschaftliche Prototypen, die sich nicht um Anforderungen der Stabilität, Sicherheit und Skalierbarkeit kümmern (müssen). Neue Entwicklungen beziehen jedoch Workflowsysteme gezielt mit ein, geben jedoch dabei den Anspruch der Automatisierung von Guidelines im Top-Down Verfahren auf: vielmehr werden Prozesse durch Regeln reaktiv auf Event im Workflowsystem getriggert (Bottom up) [TMS04].

Neben der Flexibilisierung klassischer Workflows durch die oben beschriebene Techniken, z.B. durch ECA Regeln, wurden neuartige Ansätze wie Fallbehandlung und Ad-hoc Vorgangunterstützung zur Spezifikation einer Ablauflogik entwickelt. Obwohl die medizinische Domäne dabei häufig als Referenz genannt wird, wurden die Abbildung medizinischer Guidelines auf die Workflowtechnologie und das Mapping zwischen Computerized Guideline und Workflow bisher noch nicht systematisch betrachtet.

4 Unser Ansatz

Das Projekt OLGA verfolgt eine mehrstufige Modellierung der medizinischen Prozessabläufe in der Intensivmedizin. Zunächst wird in einem Konsensfindungsprozess der medizinische Prozessablauf bestimmt. Hierfür ist zum einen das Fachwissen der Beteiligten zu konsultieren als auch eine Recherche relevanter Ressourcen vorzunehmen. Ergebnisse dieser Konsensfindung sind zunächst einfache Ablaufdiagramme und textuelle Beschreibungen. Dieses Vorgehen ist konform zur medizinischen Fachwelt, in der der Austausch von SOPs fast ausschließlich auf Textdateien basiert. Nachfolgend werden die durch Flussdiagramme und Texte beschriebenen Prozessabläufe formalisiert, d.h. in einer Prozessmodellierungssprache dargestellt. Für diese Formalisierung verwenden wir das Guideline Interchange Format GLIF [OGMJ98], da GLIF zunehmend zur Modellierung von Guidelines in der Medizin verwendet wird. Anschließend wird dieses GLIF Modell mit Abbildungsassistenten in ausführbare Vorgangsbeschreibungen transformiert. Das Abbildungsziel sind Prozessmodelle, wie sie beispielsweise mit YAWL [AaHo05] und JBPM (siehe Kapitel 5.1) ausgeführt werden können. Obwohl sich die ver-

schiedenen Sprachen hinsichtlich ihrer Modellierungskonstrukte unterscheiden, so sind diese doch fast zu vernachlässigen gegenüber den offenen Fragen, die der Abbildungsassistent adressieren muss: nicht alle GLIF Konstrukte lassen sich eindeutig auf Vorgangskontrollstrukturen abbilden (siehe 5). Schließlich können die Erfahrungen und Ergebnisse durch die Evaluation und Auswertung von Vorgangsdaten genutzt werden, um die Leistungsfähigkeit unterschiedlicher Prozessabläufe zu bewerten und ggf. Anpassungen bei der Konsensfindung und Formalisierung vorzunehmen.

Unser Ansatz unterscheidet sich dabei wesentlich in zwei Punkten von ähnlichen Projekten: Die umfassende elektronische Patientenakte der Intensivstation in Giessen (siehe 2.4) bietet eine Datengrundlage zum Testen wie sie nur wenigen anderen Projekten zur Verfügung steht. Diese erlaubt umfangreiche Simulationen durch lückenlosen Zugriff auf alle Patientendaten seit 1999 (durch erneutes „Abspielen“ aller HL7 Nachrichten) sowie den Einsatz des OLGA Prototypen im realen Umfeld mit Einwilligung der Ethikkommission im Rahmen der Evaluierung 2007/2008. Diese Voraussetzungen ermöglichen aber auch die Entwicklung eines operativ einsetzbaren Assistenzsystems. Daher kombiniert OLGA in einem eher pragmatischen Ansatz existierende Technologien in Hinblick auf ein im Alltag evaluierbaren Systems (vgl. auch 3.3).

4.1 Formalisierung und Repräsentation

In Deutschland wurde gemeinsam vom Bund Deutscher Anästhesisten (BDA) und der Deutschen Gesellschaft für Anästhesiologie und Intensivmedizin (DGAI) eine SOP-Tauschbörse für ihre Mitglieder im Internet zur Verfügung gestellt. In diesem Forum fanden wir insgesamt 23 SOPs, die wir hinsichtlich ihrer Formalisierbarkeit als Workflow in GLIF untersuchten und deren Automatisierbarkeit beurteilten [MMJR06].

Dabei zeigte sich, dass auch die sehr gut als Algorithmus abbildbaren SOPs hinsichtlich einer automatisierten Assistenz eine detailgenauere Abbildung des Prozesses erfordern. So sind zeitliche Abläufe bzw. Prioritäten von Behandlungsschritten im Textformat häufig unscharf beschrieben oder müssen in Einzelfällen parallel erfolgen. Als Beispiel sei hier der gleichzeitige Volumenersatz mit Infusionen und Blutprodukten (Transfusionen) verbunden mit einer kreislaufunterstützenden Medikation bei starken Blutungen erwähnt. Hier kann selbstverständlich nicht unbegrenzt die alleinige Flüssigkeitssubstitution mit z.B. Kochsalzlösungen bis zu einem definierten Volumenstatus ohne Berücksichtigung der Konzentration der Sauerstoffträger im Blut erfolgen. Auch die Diagnose „Herzinfarkt“ stützt sich nicht auf ein einziges Symptom, sondern vielmehr auf eine komplexe Befundkonstellation, die in einem zeitlichen Zusammen-

hang stehen muss. Hier ist zur Formalisierung von Algorithmen die Verwendung von Parallelprozessen und Synchronisationsschritten nötig.

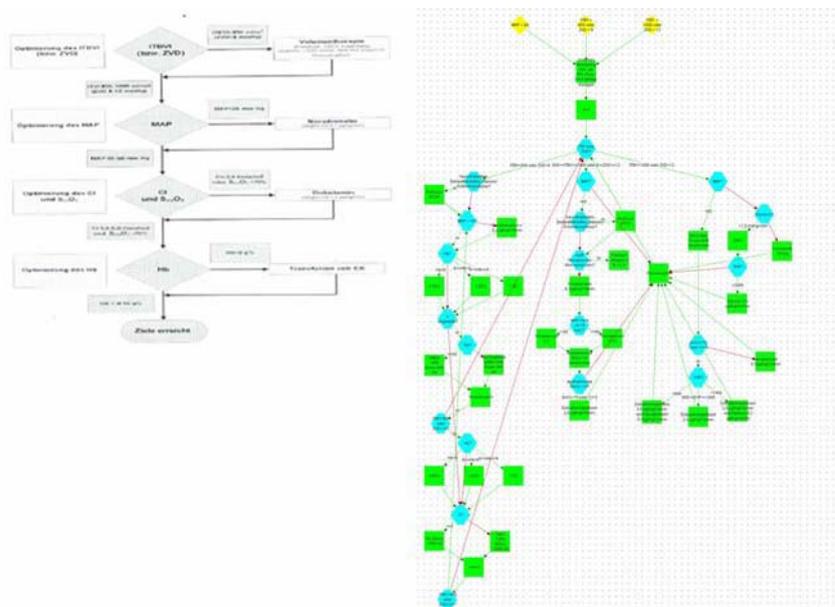


Abbildung 3: Links Darstellung einer SOP aus der SOP-Tauschbörse (BDA/DGAI), rechts die Darstellung der SOP nach der Modellierung in GLIF (Protege).

Die Präzisierung der Modelle führt zu einer erhöhten Komplexität. Die daraus resultierenden Unterschiede sind durch die beiden Flussdiagramme in Abbildung 3 exemplarisch dargestellt.

4.2 Automatisierung und Unterstützung

Wir haben verschiedenen Guidelinemodelle [PTBC03] betrachtet und das Guideline Interchange Format GLIF 3.5 gewählt (<http://www.glif.org>). GLIF entstand in Kooperation mehrerer Vorgängerprojekte und ist eine bekannte und breit angelegte Methode zur Spezifikation computerisierbarer Guidelines [PBTZ04]. Diese werden zunächst auf einer konzeptuellen Ebene als Flowchart graphisch gezeichnet („conceptual layer“), welche anschließend um Entscheidungsregeln und Datenstrukturen angereichert wird. Dazu werden auf dem „computational layer“ die Details der Entscheidungen, Datendefinitionen, Kontrollfluss und Entscheidungsregeln spezifiziert. Schließlich werden die Aktionen und Datenreferenzen auf ein konkretes Krankenhausinformationssystem abgebildet („implementable layer“). Mit Protege ist ein Werkzeug mit Beispielen für GLIF zum Editieren erhältlich (<http://protege.stanford.edu>). Die GLIF Execution Engine GLEE [WPTB04] hat gezeigt, dass GLIF Modelle hinreichend beschrieben werden können, um sie automatisch auszuführen. Wir sehen als weitere Vorteile, dass das Datenmodell von GLIF auf dem Datenmodell (RIM) von HL7 basiert und die Guideline Expression Language GELLO [SBOG04] von HL7 und ANSI als Standard akzeptiert wurde.

Um formalisierte Guidelines auszuführen bieten sich prinzipiell zwei Möglichkeiten an: die direkte Ausführung oder die indirekte Ausführung durch Mapping in ein anderes Prozessformat.

Bei der direkten Ausführung, wie sie z.B. die Guideline Execution Engine (GLEE) für GLIF realisiert [WPTB04], weist das Ausführungsmodell den Elementen (*Guideline_Steps*) Zustände wie *active*, *prepared*, *started* oder *finished* zu und ermöglicht ein benutzerdefiniertes Scheduling zwischen den Stati. Eine solche direkte Implementierung stellt sicher, dass alle Möglichkeiten und Elemente eines Standards unterstützt werden, ist jedoch stets proprietär. Einen interessanten Ansatz verfolgte das Projekt GESDOR [WPBC03], das eine generische, virtuelle Maschine zur Ausführung verschiedener Standards zum Ziel hatte.

Die mittelbare Ausführung durch Übersetzung der GLIF Guideline in ein Workflow Format und Ausführung durch eine Workflowengine kann auf etablierten Business Standards aufbauen und teilweise schon vorhandene Infrastruktur nutzen sowie gegebenenfalls mit anderen Anwendungen transparent zusammenarbeiten. Zwar lassen sich viele Konzepte von GLIF nicht durch Standardworkflowtechniken realisieren (z.B. eventbasierte Konzepte, siehe Diskussion), die jedoch bei geeigneter Auswahl und Modellierung der Guidelines vermieden werden können.

4.3 System Design

Das in Giessen eingesetzte PDMS ICUData unterstützt eine modulare Datenverarbeitungsumgebung. Die Nachrichtenkommunikation erfolgt über so genannte „Kommunikations-Master“, die die Informationen im HL7-Standard von einem Modul zum anderen leiten. Immer wenn eine Oberflächenapplikation (z. B. das grafische Krankenblatt) auf einem Client gestartet wird, etabliert der jeweilige Master eine Daten-Pipeline zu den Applikations-Servern der Datenbank und den Mastern anderer Clients. Jede Änderung in dem aktuellen Datenpool eines Patienten wird auf diese Weise gleichzeitig an die Datenbank wie auch an andere Clients mit dem geöffneten Krankenblatt des gleichen Patienten weitergeleitet (eventbasierte Kommunikation). Ein einheitlicher Datenbestand eines Krankenblattes desselben Patienten, mit dem mehrere Personen an unterschiedlichen Clients gleichzeitig arbeiten, kann somit gewährleistet werden. Sowohl interne PDMS-Module als auch externe Programme können über diesen standardisierten Mechanismus über die Applikations-Server mit der Datenbank kommunizieren.

Die OLGA Ausführungsumgebung ist ein solches externes Programm, welches über HL7 mit anderen Modulen und dem Datenserver des PDMS kommuniziert und alle Daten- und Eventströme empfängt und somit diese auf relevante Nachrichten filtern kann.

Automatisch generierte Anweisungen, Empfehlungen und Nachfragen des Systems werden über HL7 verschickt und sind Bestandteil der Patientenakte. Bis auf spezifische Alarmierung

findet keine direkte Kommunikation der OLGA Ausführungsumgebung mit dem Endanwender statt, d.h. Therapieempfehlungen gemäß einer SOP erfolgen für den Benutzer transparent über die Benutzeroberfläche des PDMS. Damit integriert sich die Benutzung von OLGA-Guidelines in die normale Stationstätigkeit und ist für den Mediziner und das Pflegepersonal keine zusätzliche Anwendung.

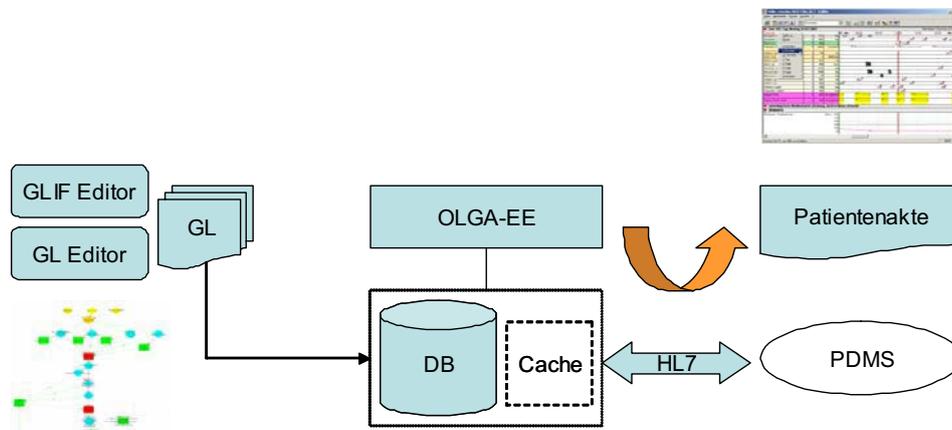


Abbildung 4: Konzept der Systemarchitektur

Zwei Möglichkeiten eröffnen sich für die Architektur von OLGA: clientseitige Installation (an jedem Patientenbett) oder als zentraler Server. In „bettseitiger“ Installation, d.h. auf den Terminals neben jedem Krankenbett, kann OLGA lokal auf alle aktuellen Patientendaten zugreifen. Da sämtliche Datenkommunikation via IMESO Master über die zentrale Patientenakte abläuft, benötigt OLGA keine weitere Synchronisation mit anderen Klienten. Durch den dezentralen Aufbau des PDMS mit lokalen Servern (Master) können Guidelines ausgeführt werden, auch wenn die Verbindung zum Server abbricht. Dies ist die bevorzugte Variante bezüglich Anforderungen der Zuverlässigkeit, Ausfallsicherheit, Skalierbarkeit und Stabilität.

5 Automatisches Mapping und Diskussion

Wir haben einen Übersetzer in Java implementiert, das GLIF Modelle liest und als Prozessspezifikation in JPDL ausgibt. Diese Modelle lassen sich mit JBPM ausführen.

5.1 Übersetzung von GLIF nach JPDL

Die Java Business Process Management Framework der JBoss Guppe (<http://www.jbpm.org>) ist eine flexible und erweiterbare Ausführungsumgebung für Workflows (Open Source Workflow Engine). Die Prozessmodelle werden in der Java Process Definition Language JPDL spezifiziert. Das zugrunde liegende graphische Programmiermodell lässt sich leicht um eigene Kno-

tentypen erweitern und erscheint dadurch als Basis für eine GLIF Ausführungsumgebung interessant. JBPM lässt sich zudem mit dem Applikationsserver von JBoss sowie mit anderen relevanten Werkzeugen wie der Rule Engine Drools (zur Auswertung von Entscheidungsbäumen) und dem Persistenzframework Hibernate (für lokale Datencaches) kombinieren.

Mit Hilfe der Protege API wird ein GLIF Modell gelesen und sukzessive nach JPDL übersetzt. Für jeden Schritt einer Guideline (*Guideline_Step*) wird ein korrespondierender Knoten in der Prozessdefinition erzeugt (siehe Tabelle). Nachfolgebeziehungen werden als Transitionen geschrieben. Somit erhält man ein Modell des Kontrollflusses, das JBPM ausführen kann. Die Interpretation von Entscheidungsregeln (*Decision_Step*) wird an eigens implementierte Klassen delegiert, die direkt auf dem GLIF Modell operieren. Ebenso werden Ausführungsanweisungen (*Action_Specification*) durch spezifische Java Klassen mittels Protege API interpretiert.

In Referenz zu den drei Modellierungsebenen in GLIF [PBTZ04] wird also das Flowchart der obersten Ebene (Conceptual Layer) übersetzt, wohingegen die Details (Entscheidungen, Datenkommunikation) des Computational und Implementable Layers an händisch implementierte Java Klassen delegiert werden, die das Workflowsystem aufruft.

Es hat sich für geeignete, einfache Beispiele gezeigt, dass sich die darin verwendeten Kontrollstrukturen von GLIF Guidelines relativ leicht nach JPDL übersetzen lassen. Eigene Experimente mit XPDL sowie Literaturstudien lassen schließen, dass dies ebenso für andere Workflowsprachen gilt.

Der Konverter wurde mit prozessorientierten Beispielmotellen in GLIF erfolgreich getestet, und auch einfache Guidelines konnten erfolgreich getestet werden. Gegenwärtig werden die Klassen für *Guideline_Steps*, *Action_Specification* und die Bedingungen für Entscheidungen erweitert. Im nächsten Schritt werden der Eventmechanismus (*Triggering_Events*), Iterationen und die Anbindung an das HL7 Netzwerk realisiert.

5.2 Diskussion

Der beschriebene Ansatz ist zunächst ähnlich geradlinig wie [PTMA04], bei dem Peleg et. al. unter anderem GLIF Modelle auf Petri-Netze abbilden, um Aussagen zu strukturellen und dynamischen Eigenschaften biologischer Prozesse zu machen. Das dort beschriebene Mapping beschränkt sich bei GLIF jedoch auf ein vereinfachtes Modell, in dem z.B. komplexe Bedingungen in Synchronisationsschritten (continuation) ebenso wenig berücksichtigt werden wie Parallelitätsbedingungen in Verzweigungen. Außerdem machen die Autoren keine Aussagen zur Ausführung von Entscheidungsregeln. Vielmehr steht im Vordergrund, welche Fragen sich mit Petrinetzen bei biologischen Prozessen beantworten lassen.

Vielleicht liegt die Ursache darin begründet, dass GLIF zwar die statischen und strukturellen Eigenschaften von Guidelines in der Medizin sehr gut unterstützt, jedoch keine Aussagen zu Laufzeiteigenschaften trifft, bzw. diese nicht modellieren kann. Dubey [DuCh00] nennt bei den Problemen, die bei der Automatisierung von GLIF auftreten - neben dem Umgang mit unvollständigen Daten sowie der problematischen Anbindung von Datenquellen - vor allem das Fehlen eines Prozesszustandskonzeptes, um den Status einer Guideline über die Zeit zu verfolgen. Der Grund ist sicherlich, dass GLIF nicht vorrangig zur Ausführung von Guidelines entwickelt wurde sondern zur formalen Erfassung und zum Austausch. Wohl daher wird GLIF von verschiedenen Forschergruppen an die jeweiligen Anforderungen und Bedingungen der konkreten Projekte adaptiert. So wird in [MZPT03] berichtet, wie zur Implementierung von Leitlinien zur Behandlung chronisch Kranker GLIF durch einen *Eligibility_Step* zur Prüfung der Indikation und ein *Notification* Objekt für Nachrichten adaptiert und erweitert wurde. Zudem passte man GLIF in Hinblick auf die Anbindung an das lokale Informationssystem und die elektronische Patientenakte an. Bei der systematischen Umsetzung von GLIF offenbart sich auch, dass die Spezifikation des Standards (Version 3.5, <http://www.glif.org>) selbst nicht eindeutig scheint. So kann beispielsweise offenbar nicht automatisch zwischen deterministischen und nicht-deterministischen Entscheidungen unterschieden werden.

Wir haben zur Implementierung unserer Ausführungsumgebung verschiedene Workflowengines verglichen (für eine umfangreiche Liste von Open Source Workflowengines siehe z.B. http://www.manageability.org/blog/stuff/workflow_in_java). Einer der Aspekte dabei war die Unterstützung bezüglich Workflowpatterns [SRRM06], aber auch die Erweiterbarkeit und der Support. Wir haben uns für JBPM entschieden, experimentieren jedoch auch mit Übersetzungen nach XPDL [WfMCoJ] und YAWL [AaHo05]. Obwohl sich die einzelnen Werkzeuge bei Spezifikationssprache, Werkzeugen und Support unterscheiden, so sind sie doch bezüglich des GLIF-Mappings und den dabei auftretenden Schwierigkeiten - wie oben beschrieben – ähnlich. Darüber hinaus ist beispielsweise die Abbildung des Event Mechanismus in JPDL (ähnlich wie XPDL und andere WF Sprachen) schwierig, wenn nicht gar unmöglich. Sofern sich der Einsatz von *Triggering_Events* auf Zeitconstraints oder Eintrittsbedingungen in Folge eines Vorgängers beschränkt, ist die Umsetzung möglich. Jedoch erlauben beliebige Events in GLIF ein unstrukturiertes Ausführen von Abläufen jenseits der Möglichkeiten der strukturierten Workflowmodellierung und sind damit nicht direkt abbildbar.

Ein interessanter Ansatz zur Implementierung von datengetriebenen Workflows ist die Fallbehandlung (Case Handling), für den sich jedoch in der Literatur bisher keine Verweise auf die Anwendung gerade im medizinischen Bereich finden.

6 Zusammenfassung

Wir haben gezeigt, wie klinische Leitlinien formalisiert und mit Standardworkflowtechniken automatisiert werden können. Die Formalisierung klinischer SOPs erfordert eine medizinische und eine technische Präzisierung, da die medizinischen Interpretationsspielräume generischer SOPs auf automatisierbare Systemalgorithmen und Datenstrukturen abgebildet werden müssen. Das in OLGA eingesetzte automatische Mapping von GLIF auf Workflowtechniken stellt eine Basis für den Test und die Evaluation von SOPs in einem PDMS dar, auch wenn sich nicht alle theoretischen Möglichkeiten von GLIF direkt abbilden lassen. Damit implementiert OLGA eine durchgängige, operativ einsetzbare Ausführungsumgebung für SOPs. Zur Zeit werden die gewählten SOPs schrittweise umgesetzt und getestet, damit sie in 2007 im PDMS der anästhesiologischen Intensivstation der Uniklinik Gießen operativ evaluiert, und ihre Effekte im klinischen Einsatz bewertet werden können.

Danksagung: Das Projekt Online Guideline Assist (OLGA) wird gefördert durch die Deutsche Forschungsgemeinschaft DFG (RO 3053/1-1; RO 2030/2-1).

Literaturverzeichnis

- [AaHo05] *Aalst, W. M. P. v. d.; Hofstede, A. H. M. t.: YAWL: yet another workflow language. In: Inf. Syst. 30 (2005) 4, S. 245-275.*
- [AaWG05] *van der Aalst, W. M. P.; Weske, M.; Grunbauer, D.: Case Handling: A New Paradigm for Business Process Support. In: Data and Knowledge Engineering 53 (2005) 2, S. 129–162.*
- [AHKB003] *Aalst, W. M. P. v. d.; Hofstede, A. H. M. t.; Kiepuszewski, B.; Barros, A. P.: Workflow Patterns. In: Distributed and Parallel Databases 14 (2003), S. 5-51.*
- [AHEA05] *Adams, M.; ter Hofstede, A. H. M.; Edmond, D.; van der Aalst, W. M. P.: Facilitating Flexibility and Dynamic Exception Handling in Workflows through Worklets. In: CAiSE Short Paper Proceedings (2005).*
- [BHBW00] *Basse, L.; Hjort Jakobsen, D.; Billesbolle, P.; Werner, M.; Kehlet, H.: A clinical pathway to accelerate recovery after colonic resection. In: Ann Surg 232 (2000) 1, S. 51-7.*

- [Bund98] *Bundesärztekammer: Zur Frage der Verbindlichkeit von Richtlinien, Leitlinien, Empfehlungen und Stellungnahmen.* <http://www.bundesaerztekammer.de/30/Richtlinien/90Verbindlich.html>, 1998, Abruf am 2006-07-24.
- [CHBP98] *Campbell, H., Hotchkiss, R., Bradshaw, N., Porteous, M.: "Integrated care pathways".* *Bmj*, 316 (1998), S. 133-137
- [DuCh00] *Dubey, A. K.; Chueh, H. C.: An XML-based format for guideline interchange and execution.* In: *Proc AMIA Symp* (2000), S. 205-9.
- [DYWH04] *Deng, S. G.; Yu, Z.; Wu, Z. H.; Huang, L. C.: Enhancement of workflow flexibility by composing activities at run-time.* In: *Proceedings of the 2004 ACM symposium on Applied computing* (2004), S. 667-673.
- [HHJN99] *Heinl, P.; Horn, S.; Jablonski, S.; Neeb, J.; Stein, K.; Teschke, M.: A comprehensive approach to flexibility in workflow management systems.* In: *SIGSOFT Softw. Eng. Notes* 24 (1999) 2, S. 79-88.
- [MaSa02] *Mangan, P. J.; Sadiq, S.: A Constraint Specification Approach to Building Flexible Workflows.* In: *Journal of Research and Practice in Information Technology* 35 (2002) 1, S. 21-39.
- [Mart2006] *Martin, J.: SOP-Tauschbörse.* *Deutscher Anästhesiecongress DAC* (2006); Vortrag H 409. 3.
- [MDHB96] *Mohr, V. D.; Dogan, N.; Haupt, W.; Bengler, K.; Zirngibl, H.; Hohenberger, W.: Optimizing a conventional documentation system on the surgical intensive care unit - a contribution to cost reduction and preventive quality management?.* In: *Anesthesiol Intensivmed Notfallmed Schmerzther* 31 (1996) 4, S. 228-38.
- [MMJR06] *Meister M, Michel-Backofen A, Jost A, Röhrig R, Junger A, Hempelmann G: Evaluation der Abbildbarkeit und Automatisierbarkeit von Standard Operating Procedures mit dem Guideline Interchange Format (GLIF);* 53. Jahrestagung der DGAI, Leipzig; Abstraktband DAC 2006; PO-1.4.2.
- [MSKK03] *Martin, J.; Schleppers, A.; Kastrup, M.; Kobylinski, C.; König, U.; Kox, W. J.; Milewski, P.; Spies, C.: Entwicklung von Standard Operating Procedures in der Anästhesie und Intensivmedizin.* In: *Anästhesiologie & Intensivmedizin* 44 (2003), S. 871-876.

- [MüGR04] *Müller, R.; Greiner, U.; Rahm, E.*: AGENT WORK: a workflow system supporting rule-based workflow adaptation. In: *Data & Knowledge Engineering* 51 (2004) 2, S. 223-256.
- [Muse05] *Musen, M. A.*: Getting Decision Support "Right": The Role of Ontologies. http://clinicalinformatics.stanford.edu/scci_seminars/slides/SCCImusen.pdf, 2005, Abruf am 2006-07-24.
- [MZPT03] *Maviglia, S. M.; Zielstorff, R. D.; Paterno, M.; Teich, J. M.; Bates, D. W.; Kuperman, G. J.*: Automating complex guidelines for chronic disease: lessons learned. In: *J Am Med Inform Assoc* 10 (2003) 2, S. 154-65.
- [OGMJ98] *Ohno-Machado, L.; Gennari, J. H.; Murphy, S. N.; Jain, N. L.; Tu, S. W.; Oliver, D. E.; Pattison-Gordon, E.; Greenes, R. A.; Shortliffe, E. H.; Barnett, G. O.*: The guideline interchange format: a model for representing guidelines. In: *J Am Med Inform Assoc* 5 (1998) 4, S. 357-72.
- [PBTZ04] *Peleg, M.; Boxwala, A. A.; Tu, S.; Zeng, Q.; Ogunyemi, O.; Wang, D.; Patel, V. L.; Greenes, R. A.; Shortliffe, E. H.*: The InterMed approach to sharable computer-interpretable guidelines: a review. In: *J Am Med Inform Assoc* 11 (2004) 1, S. 1-10.
- [PTBC03] *Peleg, M.; Tu, S.; Bury, J.; Ciccarese, P.; Fox, J.; Greenes, R. A.; Hall, R.; Johnson, P. D.; Jones, N.; Kumar, A.; Miksch, S.; Quaglini, S.; Seyfang, A.; Shortliffe, E. H.; Stefanelli, M.*: Comparing computer-interpretable guideline models: a case-study approach. In: *J Am Med Inform Assoc* 10 (2003) 1, S. 52-68.
- [PTMA04] *Peleg, M.; Tu, S.; Manindroo, A.; Altman, R. B.*: Modeling and analyzing biomedical processes using workflow/Petri Net models and tools. In: *Medinfo* 11 (2004) Pt 1, S. 74-8.
- [RuPR99] *Rupprecht, C.; Peter, G.; Rose, T.*: Ein modellgestuetzter Ansatz zur kontextspezifischen Individualisierung von Prozessmodellen. In: *Wirtschaftsinformatik* 41 (1999) 3, S. 226-236.

- [SBOG04] *Sordo, M.; Boxwala, A. A.; Ogunyemi, O.; Greenes, R. A.*: Description and status update on GELLO: a proposed standardized object-oriented expression language for clinical decision support. In: *Medinfo 11 (2004) Pt 1*, S. 164-8.
- [ScWe02] *Schuschel, H.; Weske, M.*: Fallbehandlung: Ein Neuer Ansatz zur Unterstützung Prozessorientierter Informationssysteme. In: *Proceedings of the Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen - Promise*. GI: Potsdam, 2002, S. 52-63.
- [SEGO95] *Schreiber, T. L.; Elkhatib, A.; Grines, C. L.; O'Neill, W. W.*: Cardiologist versus internist management of patients with unstable angina: treatment patterns and outcomes. In: *J Am Coll Cardiol* 26 (1995) 3, S. 577-82.
- [SRRM06] *Sedlmayr, M.; Rose, T.; Röhrig, R.; Meister, M.*: A Workflow Approach towards GLIF Execution. In: *Proceedings of the European Conference on Artificial Intelligence (ECAI)*. Riva del Garda, Italy, 2006.
- [SSBC84] *Shortiffe, E. H.; Scott, A. C.; Bischoff, M. B.; Campell, A. B.; Van Melle, W.; Jacobs, C. D.*: An Expert System for Oncology Protocol Management. In: *Buchanan, B. G.; Shortiffe, E. H. (Hrsg.) Rule-Based Expert Systems*. Addison-Wesley Publishing Company, 1984, S. 653-65.
- [TMSC04] *Tu, S. W.; Musen, M. A.; Shankar, R.; Campbell, J.; et. al.*: Modeling guidelines for integration into clinical workflow. In: *Medinfo 11 (2004) Pt 1*, S. 174-8.
- [WfMCoJ] *Workflow Management Coalition: XPDL Resources*. <http://www.wfmc.org/standards/XPDL.htm>, Abruf am 2006-07-24.
- [WPBC03] *Wang, D.; Peleg, M.; Bu, D.; Cantor, M.; Landesberg, G.; Lunenfeld, E.; Tu, S. W.; Kaiser, G. E.; Hripcsak, G.; Patel, V. L.; Shortliffe, E. H.*: GESDOR - a generic execution model for sharing of computer-interpretable clinical practice guidelines. In: *AMIA Annu Symp Proc (2003)*, S. 694-8.
- [WPTB04] *Wang, D.; Peleg, M.; Tu, S. W.; Boxwala, A. A.; Ogunyemi, O.; Zeng, Q.; Greenes, R. A.; Patel, V. L.; Shortliffe, E. H.*: Design and implementation of the GLIF3 guideline execution engine. In: *J Biomed Inform* 37 (2004) 5, S. 305-18.



“eOrganisation: Service-, Prozess-, Market-Engineering“ ist das Leitthema der 8. internationalen Tagung Wirtschaftsinformatik 2007 (WI 2007). Die Beiträge der WI 2007 greifen folgende Schwerpunkte auf:

Service-Engineering:

- | Das Internet der Dinge
- | eGovernment
- | Information Services in der Logistik
- | Mobile Business und Communications
- | Outsourcing und IT-Governance
- | Product Life Cycle Management in Unternehmenssoftwaresystemen
- | Semantic Web
- | Service Oriented Computing
- | Verkehr und Mobilität

Prozess-Engineering:

- | Business Intelligence
- | Business Process Engineering
- | Collaborative Business
- | Customer und Supplier Relationship Management
- | eHealth
- | eLearning
- | IS-Architekturen
- | IT-Risikomanagement
- | Modellierung als Innovationsmotor
- | Softwareprozessverbesserung
- | Supply Chain Management
- | Operations Research
- | Wissensmanagement

Market-Engineering:

- | Agenten- und Multiagenten-Technologien für betriebliche Anwendungen
- | eFinance
- | eLearning Geschäftsmodelle
- | Elektronische Märkte und elektronische Verhandlungen
- | eMedia
- | eServices
- | Informationsmanagement in der Energie- und Umweltinformatik
- | Internetökonomie
- | Rechtsfragen der Informationsgesellschaft

Die Tagungsbände richten sich an Wissenschaftler und Praktiker, und bieten einen Einblick in aktuelle Forschungsfelder der Wirtschaftsinformatik

ISBN: 978-3-86644-094-4 (Band 1)

ISBN: 978-3-86644-093-7 (Set)

www.uvka.de