



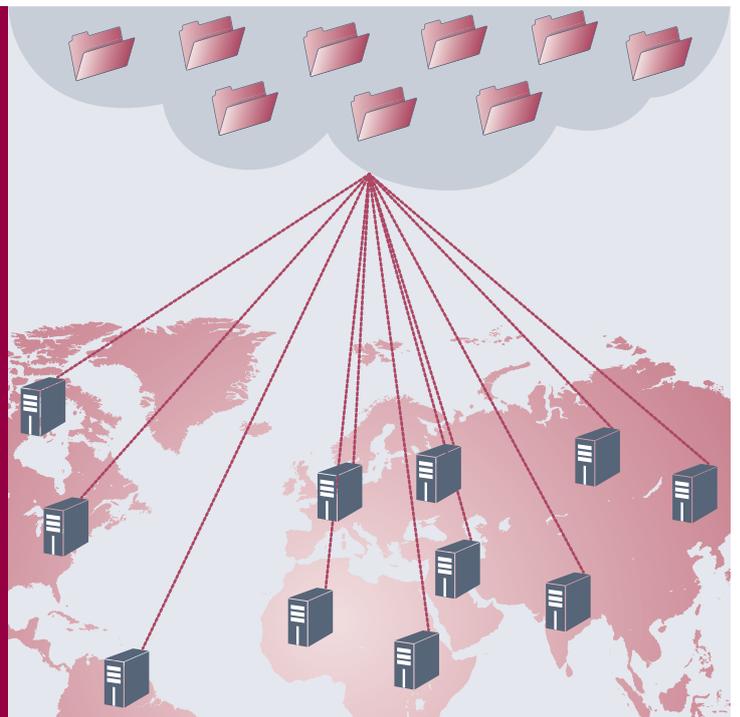
BÜRO FÜR TECHNIKFOLGEN-ABSCHÄTZUNG
BEIM DEUTSCHEN BUNDESTAG

Timo Leimbach
Daniel Bachlechner

Big Data in der Cloud

TA-Vorstudie

Juli 2014
Hintergrundpapier Nr. 19





Big Data in der Cloud



Das Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB) berät das Parlament und seine Ausschüsse seit 1990 in Fragen des technischen und gesellschaftlichen Wandels. Das TAB ist eine organisatorische Einheit des Instituts für Technikfolgenabschätzung und Systemanalyse (ITAS) im Karlsruher Institut für Technologie (KIT). Das TAB kooperierte zur Erfüllung seiner Aufgaben von 2003 bis 2013 mit dem Fraunhofer-Institut für System- und Innovationsforschung (ISI), Karlsruhe.



Timo Leimbach
Daniel Bachlechner

Big Data in der Cloud

TA-Vorstudie



Büro für Technikfolgen-Abschätzung
beim Deutschen Bundestag (TAB)
Neue Schönhauser Straße 10
10178 Berlin

Fon: +49 30 28491-0
Fax: +49 30 28491-119
buero@tab-beim-bundestag.de
www.tab-beim-bundestag.de

2014

ISSN-Print 2199-7128
ISSN-Internet 2199-7136



INHALT

ZUSAMMENFASSUNG	9
<hr/>	
I. EINLEITUNG	31
1. Digitale Gesellschaft und Konvergenz von Technologien	31
2. Ziel und Aufbau des Berichts	33
3. Überblick über die Forschungslage	36
<hr/>	
II. GRUNDLAGEN	39
1. Big Data	39
1.1 Begriffsdefinition	39
1.2 Entwicklung und Abgrenzung	41
1.3 Technische Grundlagen	44
1.4 Einsatzmöglichkeiten	50
2. Cloud Computing	52
2.1 Begriffsdefinition	52
2.2 Entwicklung und Abgrenzung	56
2.3 Technische Grundlagen	58
2.4 Einsatzmöglichkeiten	60
3. Fazit – Cloud Computing als Basis von Big Data	61
<hr/>	
III. MÄRKTE UND ANBIETER	63
1. Big Data	63
1.1 Markt	63
1.2 Anbieter	66
2. Cloud Computing	70
2.1 Markt	71
2.2 Anbieter	75
3. Fazit – Konvergenz der Märkte	78
<hr/>	
IV. DIFFUSION UND NUTZUNG	79
1. Big Data	79
1.1 Unternehmen	79
1.2 Öffentliche Verwaltung	86
1.3 Wissenschaft	88
1.4 Privatpersonen	90



INHALT

2.	Cloud Computing	91
2.1	Unternehmen	91
2.2	Öffentliche Verwaltung	94
2.3	Wissenschaft	97
2.4	Privatpersonen	99
3.	Fazit – Big Data als Treiber für Cloud Computing	100
<hr/>		
V.	POTENZIALE	101
1.	Sozioökonomische Potenziale	101
1.1	Potenziale für Organisationen	101
1.2	Potenziale für die Wissenschaft	108
1.3	Potenziale für Privatpersonen	111
1.4	Gesamtgesellschaftliche Potenziale	113
2.	Anwendungspotenziale	124
2.1	Klima/Energie	124
2.2	Gesundheit/Ernährung	128
2.3	Mobilität	131
2.4	Sicherheit	133
2.5	Kommunikation	136
3.	Fazit	139
<hr/>		
VI.	HERAUSFORDERUNGEN	143
1.	Technische Herausforderungen	143
1.1	Analysearchitektur	143
1.2	Datenmaterial	145
1.3	Interoperabilität	147
2.	Betriebliche Herausforderungen	147
2.1	Wirtschaftlichkeit	148
2.2	Fachwissen	149
2.3	Planung und Durchführung	150
2.4	Kontrolle	150
3.	Rechtliche Herausforderungen	151
3.1	Datenschutz	152
3.2	Eigentum an Daten	157
3.3	Vertragsgestaltung	159
3.4	Haftung	160
3.5	Wettbewerb	161
3.6	Weitere rechtliche Herausforderungen	162

4. Gesellschaftliche Herausforderungen	163
4.1 Vertrauen und Verlässlichkeit	164
4.2 Verengung des Sichtbaren	166
4.3 Autonomie und Beeinflussung	168
4.4 Diskriminierung und Teilhabe	170
4.5 Vulnerabilität und Abhängigkeit	171
5. Strukturelle Herausforderungen	174
5.1 Humankapital	174
5.2 Infrastruktur/Breitbandversorgung	175
5.3 Digitale Wettbewerbsfähigkeit	177
5.4 Internetgovernance	181
<hr/>	
VII. RESÜMEE: HANDLUNGS- UND FORSCHUNGSFELDER	185
<hr/>	
LITERATUR	197
<hr/>	
ANHANG	219
1. Tabellenverzeichnis	219
2. Abbildungsverzeichnis	219



ZUSAMMENFASSUNG

Big Data und Cloud Computing wurden in den letzten Jahren – neben Themen wie Mobile Computing und Social Media – besonders häufig als Trendthemen oder als zentrale Einflüsse bezeichnet, die nicht nur den Markt für Informationstechnologie (IT), sondern die Gesellschaft insgesamt nachhaltig verändern. Der Fokus dieser Studie liegt auf dem Zusammenspiel der Phänomene Big Data und Cloud Computing. Während der Begriff Big Data im Zusammenhang mit Ansätzen zur effizienten Analyse besonders großer, heterogener Datenmengen verwendet wird, beschreibt Cloud Computing einen Ansatz zur bedarfsgerechten Bereitstellung von IT-Ressourcen über ein Netzwerk. Für wie groß das in Big Data bzw. Cloud Computing liegende Potenzial gehalten wird, verdeutlichen häufig gezogene Vergleiche. Beispielsweise werden Daten zunehmend als neuer Produktionsfaktor oder als neue Form von Kapital bezeichnet. Beim Cloud Computing wird angenommen, dass der Ansatz den IT-Markt ähnlich verändern wird, wie einst der Übergang von lokaler Stromerzeugung zu flächendeckenden Stromnetzen die Industrie revolutioniert hat.

Auch wenn Big-Data-Analysen ohne Cloud Computing durchgeführt werden können und Cloud Computing neben Big-Data-Analysen zahlreiche andere Anwendungsfelder hat, wird der Kombination von Big Data und Cloud Computing großes Potenzial zugeschrieben. Ziel dieser Vorstudie ist eine überblicksartige Analyse und Bewertung der Potenziale und Herausforderungen von »Big Data in der Cloud«. Die Studie setzt sich dabei intensiv mit unterschiedlichen, teils widersprüchlichen Positionen auseinander. Eine ausgewogene Betrachtung von Potenzialen und Herausforderungen von Big Data in der Cloud bzw. Big Data und Cloud Computing in Wirtschaft, Politik, Forschung und Gesellschaft stellt eine notwendige Voraussetzung für die Identifikation und Beschreibung von zukünftigen Handlungs- und Forschungsfeldern dar. Methodisch basiert die Studie auf einer umfassenden Analyse der vorhandenen Literatur und einer Reihe von Experteninterviews.

BIG DATA IN DER CLOUD

Der Begriff »Big Data in der Cloud« beschreibt verschiedene Varianten der Kombination von Big Data und Cloud Computing, wobei die Nutzung von Clouddiensten grundsätzlich als Basis für die Durchführung von Big Data verstanden wird. Big Data in der Cloud stellt also eine Alternative zum Aufbau und zur Nutzung einer klassischen Datenanalysearchitektur dar, bei der sowohl die notwendige Hard- und Software vor Ort installiert ist. Organisationen wie Unternehmen, staatliche Stellen und Behörden sowie wissenschaftliche Einrichtungen



gen, aber auch Privatpersonen, die große, heterogene Datenmengen cloudbasiert analysieren möchten, können sich entscheiden zwischen Angeboten, bei denen der Cloudanbieter nur die IT-Infrastruktur zur Verfügung stellt und sie selbst die für die Datenanalyse notwendige Software betreiben, und Angeboten, bei denen die gesamte Big-Data-Lösung vom Cloudanbieter bereitgestellt wird. Um die Chancen und Risiken von Big Data in der Cloud besser einschätzen zu können, ist es notwendig, sich mit den Grundlagen von Big Data und Cloud Computing vertraut zu machen.

BIG DATA

Charakteristisch für Big Data sind nicht nur die Menge und die Heterogenität der Daten, sondern die Geschwindigkeit ihrer Erzeugung sowie die hohen Anforderungen an die Geschwindigkeit ihrer Analyse. Über die Kriterien, aufgrund derer Daten zu Big Data werden, bestand lange Uneinigkeit. Mittlerweile hat sich die Ansicht durchgesetzt, dass es sich dann um Big Data handelt, wenn Daten mit existierenden Technologien wie relationalen Datenbanken nicht mehr effizient verarbeitet werden können. Bei der Durchführung von Analysen großer Datenmengen, die häufig allgemein als Big-Data-Analytik bezeichnet wird, geht es vor allem um die Erkennung von Zusammenhängen, Mustern und Bedeutungen. Zum Einsatz kommen neben statistischen Verfahren Vorhersagemodelle, Optimierungsalgorithmen, Data Mining sowie Text- und Bildanalytik. Big-Data-Analytik ist nicht etwas gänzlich Neues, sondern vielmehr eine Weiterentwicklung von bereits seit Längerem existierenden Ansätzen zur Analyse von Daten. Big-Data-Analytik hat ihren Ursprung im Bereich der betrieblichen Entscheidungsunterstützungssysteme und wird als Weiterentwicklung von Business Intelligence (BI), welche sich mit der systematischen Analyse von Unternehmensdaten beschäftigt, gesehen. Auffällig ist, dass zahlreiche Definitionen von Big Data bzw. Big-Data-Analytik auf eine Unterstützung von wettbewerbskritischen Entscheidungen in einem betriebswirtschaftlichen Kontext fokussieren. Auf die Tatsache, dass Big-Data-Analytik z. B. auch in der Wissenschaft oder in der öffentlichen Verwaltung eine bedeutende Rolle spielt, wird nur selten explizit eingegangen.

Einen wesentlichen Unterschied zwischen Big-Data-Analytik und klassischer BI stellt die Tatsache dar, dass BI primär auf strukturierten Daten beruht, während im Kontext von Big Data in der Regel polystrukturierte Daten verarbeitet werden. Darüber hinaus werden im Rahmen von Big-Data-Analysen neben internen verstärkt auch externe Daten verwendet. Dementsprechend unterscheidet sich die analytische Infrastruktur für Big Data von jener für BI. Während für BI in der Regel existierende Werkzeuge für die Integration von Daten ausreichen, kommen im Kontext von Big Data speziell für den Import von großen, heterogenen Datenmengen entwickelte Verfahren zum Einsatz. Diese werden in der Vorstudie im Überblick dargestellt.

CLOUD COMPUTING

Für Cloud Computing konnte sich lange Zeit keine einheitliche Definition durchsetzen, unter anderem weil das Konzept nicht völlig neu ist, sondern eine Weiterentwicklung verschiedener bereits existierender Ansätze darstellt. Für Publikationen und Vorträge wurde in den letzten Jahren häufig die Definition des in den USA ansässigen National Institute of Standards and Technology (NIST) herangezogen, woran sich auch diese Vorstudie orientiert. Cloud Computing wird dabei als ein Modell beschrieben, das es erlaubt, bei Bedarf jederzeit und überall bequem über ein Netzwerk auf einen geteilten Pool von konfigurierbaren IT-Ressourcen zuzugreifen (Multitenancy), die schnell, mit minimalem Managementaufwand und geringer Interaktion mit dem Anbieter zur Verfügung gestellt werden können. Jede Ressource kann mehreren Nutzern zugewiesen werden. Die Nutzer teilen sich eine physische Cloudinfrastruktur und können Teile der genutzten Anwendungen, aber nicht deren Quellcode an ihre Bedürfnisse anpassen. Virtualisierung stellt einen Ansatz zur Realisierung des Multi-Tenant-Modells dar und ist eine wesentliche Grundlage von Cloud Computing. Bei der Nutzung von gebührenfinanzierten Clouddiensten ist in der Regel eine Abrechnung üblich, die sehr eng mit der tatsächlichen Nutzung eines Dienstes zusammenhängt. Im Kontext von Cloud Computing werden drei Hauptkategorien von Diensten unterschieden: Software as a Service (SaaS), Platform as a Service (PaaS) und Infrastructure as a Service (IaaS). Während es bei SaaS um die Bereitstellung von Anwendungen geht, wird unter PaaS die Bereitstellung von Programmier- und Laufzeitumgebungen mit flexiblen Rechen- und Datenkapazitäten verstanden, unter IaaS diejenige von Hardwareressourcen (z. B. Rechner, Netzwerke, Speicher). Darüber hinaus wird zwischen vier wesentlichen Formen der Bereitstellung unterschieden: Private Cloud, Community Cloud, Public Cloud und Hybrid Cloud. Während bei einer Private Cloud die Cloudinfrastruktur einer Organisation exklusiv zur Verfügung steht, steht sie bei einer Community Cloud einem bestimmten Nutzerkreis mit gemeinsamen Anliegen zur Verfügung. Bei einer Public Cloud kann die Cloudinfrastruktur von der breiten Öffentlichkeit genutzt werden. Eine Hybrid-Cloud-Infrastruktur ist dadurch gekennzeichnet, dass sie sich aus mehreren unabhängigen Infrastrukturen zusammensetzt, die verbunden wurden. Bei den Infrastrukturen muss es sich nicht ausschließlich um Cloudinfrastrukturen handeln.

MÄRKTE UND ANBIETER

Den großen Marktforschungsunternehmen zufolge gehören sowohl Big Data als auch Cloud Computing zu den am schnellsten wachsenden Teilmärkten für Software und IT-Services, die in den kommenden Jahren beachtliche Auswirkungen auf die Nutzung von IT im Allgemeinen haben werden. Studien zur Größe des Marktes für cloudbasierte Big-Data-Lösungen sind allerdings rar. Es



fällt auf, dass cloudbasierte Big-Data-Lösungen bei Start-ups, die mit dem Ziel, den Big-Data-Markt zu bedienen, entstanden sind, eine deutlich größere Rolle spielen als bei etablierten Anbietern, die versuchen ihr Geschäftsfeld auf den Markt für Big Data auszudehnen. Eine Ausnahme stellt Amazon dar, das mit Elastic MapReduce bereits vor mehreren Jahren eine entsprechende cloudbasierte Big-Data-Anwendung auf den Markt brachte.

MARKT UND ANBIETER FÜR BIG DATA

Die globalen Umsätze mit Big Data lagen dem Bundesverband Informationswirtschaft, Telekommunikation und neue Medien (BITKOM) zufolge 2011 bei rund 3,3 Mrd. Euro. Für das Jahr 2012 wurde mit einem Umsatz von 4,5 Mrd. Euro gerechnet. Bis 2016 sollen die globalen Umsätze auf 15,7 Mrd. Euro wachsen. Das entspricht einer mittleren Wachstumsrate von 36 % pro Jahr. Der Markt für Big Data zählt somit zu den wachstumsstärksten Teilmärkten des gesamten IT-Marktes. Die Wachstumsspitze wurde für die Jahre 2013 und 2014 vorhergesagt. Zum Abflachen des Wachstums soll einerseits der Preisverfall auf der Technologieseite und andererseits das zunehmende Fachwissen bei den Anwendern beitragen. Die Verteilung der globalen Umsätze auf die Kategorien Dienstleistungen, Software und Hardware ist ausgewogen – auf jede der drei Kategorien entfällt bislang ungefähr ein Drittel des Gesamtumsatzes. Mit zunehmender Marktreife soll es in Zukunft jedoch zu einer Verschiebung der Umsätze in Richtung Dienstleistungen kommen. Ein wesentlicher Teil der Umsätze wurde im Jahr 2011 mit 42 % im amerikanischen Raum erzielt, mit den USA als derzeit zentralem Big-Data-Entwicklungsstandort. Der BITKOM führt die starke Position des amerikanischen Raums insbesondere auf den volkswirtschaftlichen Stellenwert jener Branchen in den USA zurück, die federführend beim Einsatz von Big-Data-Analysen sind. Allerdings muss auch berücksichtigt werden, dass sich in den USA Regierungsbehörden und Nachrichtendienste für einen wesentlichen Teil der Umsätze verantwortlich zeigen. Mit 27 % der globalen Umsätze nimmt auch Europa einen wichtigen Platz ein, weitere 26 % entfallen auf den asiatisch-pazifischen Raum. Dies basiert laut BITKOM einerseits auf dem verbreiteten Einsatz von Big Data in Japan und andererseits auf den weitgehenden Befugnissen von Behörden und Unternehmen, vor allem in China, im Umgang mit personen- und firmenbezogenen Daten. Der mittlere Osten und Afrika spielen mit 5 % der weltweiten Umsätze nur eine untergeordnete Rolle. Auf deutsche Unternehmen entfällt derzeit nur ein Fünftel der Big-Data-Umsätze in Europa. Der BITKOM geht allerdings davon aus, dass im Jahr 2016 Deutschland für rund die Hälfte der europäischen Big-Data-Umsätze verantwortlich sein wird.

Bereits im Jahr 2012 gab es weltweit mehr als 80 Anbieter mit konkreten Big-Data-Lösungen. Die Anbieterlandschaft kann im Hinblick auf Big Data in zwei große Gruppen unterteilt werden: auf der einen Seite Start-ups, die mit dem Ziel, den Big-Data-Markt zu bedienen, entstanden sind, auf der anderen Seite etablier-

te Anbieter von Datenbanken und Data-Warehouse-Lösungen, die versuchen, ihr Geschäftsfeld zu erweitern. Vertreter der zweiten Gruppe haben nicht nur den Vorteil einer bestehenden Kundenbasis, sondern auch den einer bereits erprobten Produktpalette. Viele der jungen Unternehmen, deren Unternehmensstrategie und Portfolio ausschließlich auf Big Data ausgerichtet ist, agieren als Technologie- und Know-how-Partner etablierter und weltweit operierender Technologie- und IT-Serviceanbieter. Die auf Big Data spezialisierten Unternehmen erwirtschaften immerhin rund 8 % des globalen Big-Data-Umsatzes. Wie für einen jungen Markt nicht unüblich, werden für die nächste Zeit zahlreiche Übernahmen und Zusammenschlüsse erwartet. Es wird angenommen, dass Open-Source-Angebote sowie cloudbasierte Angebote im Big-Data-Markt eine größere Rolle spielen werden, als dies im Markt für BI der Fall ist. Zu den umsatzstärksten Unternehmen in der Gruppe der etablierten Anbieter zählen IBM, HP, Teradata, Dell, Oracle, SAP und EMC. Im Jahr 2012 war IBM mit einem Umsatz von rund 1,3 Mrd. US-Dollar im Bereich Big Data der umsatzstärkste Anbieter auf dem Markt. Big Data trug allerdings nur gut 1 % zum Gesamtumsatz von IBM bei. Zu den bekanntesten Unternehmen in der Gruppe der Start-ups gehören Cloudera, Hortonworks, Splunk, 10Gen und MapR. Splunk hat von allen Unternehmen, die ausschließlich den Big-Data-Markt bedienen, den größten Marktanteil.

MARKT UND ANBIETER FÜR CLOUD COMPUTING

Aufgrund von Unterschieden im Hinblick auf den gewählten methodischen Ansatz, die Abgrenzung von Marktsegmenten und Teilmärkten sowie Annahmen zur wirtschaftlichen Entwicklung unterscheiden sich die Ergebnisse der Studien der großen Marktforschungsunternehmen zum Teil erheblich. Trotzdem lassen sich bestimmte Entwicklungen des Marktes für Cloud Computing ablesen. Die globalen Umsätze am Markt für Cloud Computing lagen 2011 laut Studien von IDC und Forrester bei rund 40 Mrd. US-Dollar, Gartner kam gar auf einen Umsatz von 91,4 Mrd. US-Dollar. Berücksichtigt wurden in den drei Studien ausschließlich Umsätze mit Public-Cloud-Diensten. Für das Jahr 2016 errechnete Gartner einen Umsatz von 206,6 Mrd. US-Dollar. Forrester sagt für 2020 einen globalen Umsatz von 241 Mrd. US-Dollar voraus. Alle großen Marktforschungsunternehmen gehen über die nächsten Jahre von einer durchschnittlichen jährlichen Wachstumsrate von mehr als 20 % aus. Noch offen ist allerdings, wie sich die Enthüllungen zu den Überwachungspraktiken der in den USA ansässigen National Security Agency (NSA) und anderer Geheimdienste auswirken werden. Die Auswirkungen wurden bisher von Marktforschern nur wenig thematisiert. Bemerkenswert ist, dass sich der Fokus auf Private-Cloud-Lösungen und hybride Nutzungsformen weiter zu verstärken scheint. Zudem ist ein Trend in Richtung deutscher und europäischer Anbieter erkennbar. SaaS-Dienste bilden nicht nur derzeit die umsatzstärkste Dienstkategorie, sondern werden das auch in Zukunft



tun. Derzeit ist der nordamerikanische Markt nicht nur der umsatzstärkste Markt für Cloud Computing, sondern auch der am schnellsten wachsende. Den aufstrebenden asiatischen Märkten, insbesondere denen von Indien, Indonesien und China, wird großes Potenzial zugeschrieben. Das schnelle Wachstum in den asiatischen Märkten wird darauf zurückgeführt, dass viele Unternehmen in diesen Regionen keine komplexen, über Jahre hinweg gewachsenen IT-Infrastrukturen haben. Daher fällt ihnen der Wechsel auf neue Modelle wie Cloud Computing leichter. Der europäische Markt ist derzeit nach dem nordamerikanischen der zweitgrößte. Die Nutzung von Cloud Computing wird sich laut Gartner in Europa, insbesondere im Vergleich zu den USA, jedoch weiterhin langsamer entwickeln. Als Gründe werden neben einer allgemein größeren Zurückhaltung gegenüber Cloud Computing unter anderem die Eurokrise und die anhaltende Rezession in Europa sowie die zahlreichen Unterschiede zwischen den einzelnen Staaten genannt. In Deutschland lag dem BITKOM zufolge der Umsatz im Bereich Cloud Computing im Jahr 2012 bei rund 5 Mrd. Euro. Für 2013 wurde eine Umsatzsteigerung von knapp 50 % erwartet, und auch in den Folgejahren sollen die Zuwächse im zweistelligen Bereich liegen.

Die Anzahl der Anbieter im Bereich Cloud Computing nimmt rasant zu. Neben zahlreichen kleineren Unternehmen, die auf dem Cloud-Computing-Markt tätig sind, gibt es nur wenige große Anbieter. Die Anbieter lassen sich in der Regel einer der vier Kategorien Early Movers, spezialisierte Technologie- und Softwareanbieter, große IT-Dienstleister und Cloud-born-Unternehmen zuordnen. Zu den Early Movers, Unternehmen, die früh in den Markt eingetreten sind, gehören Anbieter wie Amazon, Google oder Salesforce.com. Unternehmen wie VMware, Citrix, Terremark und Rackspace haben sich auf die Bereitstellung von Technologien und Software spezialisiert, die für den Aufbau und Betrieb von Cloudinfrastrukturen benötigt werden. Einige Unternehmen dieser Gruppe haben sich mittlerweile zu Public-Cloud-Anbietern entwickelt. In dieser Kategorie kam es zuletzt verstärkt zu Übernahmen und Zusammenschlüssen. Bei IBM, HP, Dell oder auch Cisco handelt es sich nicht nur um große IT-Dienstleister, sondern auch um Unternehmen, die im Cloud-Computing-Markt aktiv sind. Die bestehende Kundenbasis, eine gute Vernetzung in der Branche und der problemlose Zugriff auf einschlägiges Expertenwissen erleichterte das Geschäft mit Clouddiensten für diese Unternehmen deutlich. Auch große Softwareunternehmen wie Microsoft, SAP und Oracle sind mittlerweile mit ihren Angeboten in der Cloud vertreten. Cloud-born-Unternehmen bieten ausschließlich Clouddienste an, häufig auf der Basis von Diensten anderer Unternehmen. Ein bekanntes Beispiel ist Dropbox, das 2008 zur Synchronisation von Dateien zwischen verschiedenen Computern und Personen entwickelt wurde.

DIFFUSION UND NUTZUNG

Im Rahmen einer Studie von Gartner gab knapp ein Drittel der Teilnehmer an, dass sie für ein Unternehmen tätig sind, das cloudbasierte Angebote zur Unterstützung seiner Fähigkeiten zur Datenanalyse in ausgewählten Geschäftsbereichen bereits nutzt oder plant, dies innerhalb eines Jahres zu tun. Nur 17 % der Befragten antworteten, dass die vorhandene Infrastruktur im Unternehmen teilweise oder komplett durch cloudbasierte Angebote ersetzt wurde oder entsprechende Schritte geplant sind.

DIFFUSION UND NUTZUNG VON BIG DATA

Während gut zwei Drittel der US-amerikanischen Unternehmen im Bereich Big Data aktiv sind, haben in Europa oder der Region Asien/Pazifik weniger als die Hälfte der Unternehmen Erfahrungen mit Big Data. Dabei gibt es starke regionale Unterschiede. In Europa sind beispielsweise Unternehmen mit Sitz im Vereinigten Königreich deutlich aktiver als Unternehmen mit Sitz in Deutschland oder den Niederlanden. Nur gut ein Drittel der deutschen Unternehmen hat bereits erste Erfahrungen im Bereich Big Data gesammelt. Big-Data-Analysen kommen vor allem in den Bereichen Controlling, Produktion, Vertrieb, Marketing sowie Forschung und Entwicklung zum Einsatz. In Unternehmen werden in erster Linie Transaktionsdaten, Protokolldaten und Ereignisdaten analysiert. Während gut zwei Drittel der Daten intern anfallen, stammt knapp ein Drittel von Quellen außerhalb des Unternehmens. In Unternehmen werden überwiegend Standardwerkzeuge zur Analyse großer Datenmengen verwendet. Spezielle Big-Data-Werkzeuge kommen nur in knapp jedem zweiten Unternehmen mit Erfahrung im Bereich Big Data zum Einsatz.

Die *öffentliche Verwaltung* führt einerseits selbst Big-Data-Analysen durch, andererseits fördert sie die Weiterentwicklung und Nutzung von Big-Data-Analysen durch Unternehmen und wissenschaftliche Einrichtungen. Bisher liegen noch keine umfassenden Studien über die Nutzung von Big Data in der öffentlichen Verwaltung vor. Einzelne Beispiele belegen jedoch, dass die Durchführung von Big-Data-Analysen in der öffentlichen Verwaltung Vorteile bringen kann. So setzte beispielsweise die Königlich Technische Hochschule in Stockholm bei der Verbesserung des Verkehrsmanagements auf Big-Data-Analysen; in Indien wird an einer zentralen Datenbank gearbeitet, in der für jeden indischen Staatsbürger neben einer Identifikations-ID unter anderem auch biometrische Daten gespeichert werden sollen; und das deutsche Unternehmen Empolis Information Management entwickelte in enger Zusammenarbeit mit dem Europäischen Patentamt mit EPOQUE-Net eine Lösung zur weltweiten Patentrecherche. Auch wenn Big Data heutzutage vor allem als Herausforderung für Unternehmen sowie staatliche Stellen und Behörden gesehen wird, gehen die Ursprünge von Big Data eigentlich auf den wissenschaftlichen Anwendungsbereich zurück. Ein bekanntes



Beispiel für die Durchführung von Big-Data-Analysen als Forschungsmethode ist die Untersuchung eines Konsortiums rund um die Harvard School of Public Health über den Zusammenhang zwischen der Mobilität von kenianischen Staatsbürgern und der Verbreitung von Malaria. Neben der Epidemiologie spielt Big Data vor allem in Disziplinen wie Kernphysik, Astronomie, Geologie, Genetik und Klimaforschung eine wichtige Rolle. Auch wenn Privatpersonen in der Regel selbst keine Big-Data-Analysen durchführen (lassen), profitieren sie in vielerlei Hinsicht von Big-Data-Analysen, die von Unternehmen, der öffentlichen Verwaltung oder der Wissenschaft durchgeführt werden.

DIFFUSION UND NUTZUNG VON CLOUD COMPUTING

Im Rahmen einer in Europa durchgeführten Studie gab knapp die Hälfte der gut 1.000 befragten *Unternehmen* an, dass sie zumindest in einem Bereich Dienste aus der Cloud nutzen. Knapp zwei Drittel nutzen Cloudangebote in mehr als einem Bereich. Eine in den USA durchgeführte Studie kam zu dem Ergebnis, dass dort mehr als drei Viertel der Unternehmen zumindest in einem Bereich Dienste aus der Cloud nutzen. Auch wenn die Ergebnisse der beiden Studien aufgrund unterschiedlicher methodischer Ansätze nicht ohne Weiteres vergleichbar sind, deuten sie doch auf eine stärkere Nutzung von Cloudangeboten in den USA hin. Aus beiden Studien geht hervor, dass größere Unternehmen eher Cloudangebote nutzen als kleinere und dass vor allem einfachere Dienste, wie E-Mail- und Securitydienste, aus der Cloud bezogen werden. Für Deutschland kam eine Studie zum Ergebnis, dass gut ein Drittel der Unternehmen Clouddienste nutzt. Bei Großunternehmen liegt der Anteil deutlich höher als bei KMU. Für deutsche Unternehmen kommen insbesondere Private-Cloud-Angebote infrage, nur jedes zehnte Unternehmen setzt auf Angebote aus einer Public Cloud. Im Branchenvergleich liegt in Deutschland sowohl bei der Nutzung von Private-Cloud-Angeboten als auch bei der Nutzung von Public-Cloud-Angeboten der Bereich IT und Telekommunikation an der Spitze. Private-Cloud-Angebote werden vor allem auch in der Chemie- und Pharmaindustrie und im Bereich Verkehr und Logistik häufig genutzt.

Die *öffentliche Verwaltung* nutzt nicht nur selbst Clouddienste und tritt unter Umständen auch als Cloudanbieter auf, sondern fördert auch die Weiterentwicklung und Nutzung von Cloud Computing durch Unternehmen und wissenschaftliche Einrichtungen. Allerdings zeigten die zuvor genannten Studien, dass staatliche Stellen und Behörden sowohl in Europa als auch in den USA beim Einsatz von Cloud Computing deutlich hinter Unternehmen zurückliegen, wobei das Interesse wächst. Die Herangehensweise unterscheidet sich von Staat zu Staat deutlich. Während beispielsweise in Dänemark keine Pläne zum Aufbau einer nationalen Cloudinfrastruktur bekannt sind und vielmehr auf existierende Dienste aufgebaut wird, sind Frankreich und das Vereinigte Königreich dabei, nationale Cloudumgebungen zu realisieren, die als Plattform für den gesamten

öffentlichen Sektor dienen sollen. In der Wissenschaft spielt Cloud Computing vor allem dann eine Rolle, wenn hohe Rechenleistung gefordert ist oder bestimmte Dienste oder Daten einem verteilten Nutzerkreis zur Verfügung stehen sollen. Studien zur Verwendung von Cloudangeboten durch Privatpersonen zeigen, dass die Abgrenzung des Begriffs »Cloud Computing« in diesem Bereich besonders schwierig ist. Es fällt auf, dass Cloud Computing in Europa breiter ausgelegt wird als in den USA. Besonders häufig werden Such-, Blogging- und Streamingdienste genutzt. Deutlich geringer ist das Interesse an Diensten, bei denen Daten in der Cloud gespeichert werden. Eine Ausnahme stellen allerdings Social-Network-Dienste dar. Weder in Europa noch in den USA sind viele Privatpersonen bislang bereit, für Clouddienste zu bezahlen.

POTENZIALE

Während es kaum Studien gibt, die sich detailliert mit den Potenzialen von Big Data in der Cloud auseinandersetzen, wurden die Potenziale von Big Data einerseits und Cloud Computing andererseits bereits häufig untersucht. Sowohl im Hinblick auf sozioökonomische Potenziale als auch im Hinblick auf Anwendungsmöglichkeiten erlaubt eine Auseinandersetzung mit den spezifischen Potenzialen in den beiden einzelnen Bereichen Schlüsse auf die Potenziale von cloudbasierten Big-Data-Lösungen. Der Einsatz von Big Data in der Cloud ist vor allem dann vielversprechend, wenn die auszuwertenden Daten im Netz entstehen oder dort bereits aggregiert wurden, wenn Ergebnisse von Analysen komfortabel zugänglich gemacht werden sollen und wenn auf Daten aus verschiedenen Quellen zugegriffen werden soll.

SOZIOÖKONOMISCHE POTENZIALE

Vor allem im Hinblick auf Unternehmen können auf Grundlage aktueller Studien zahlreiche Potenziale von Big Data und Cloud Computing identifiziert werden. Während im Zusammenhang mit Big Data Umsatzsteigerungen eine größere Bedeutung beigemessen wird, überwiegt im Zusammenhang mit Cloud Computing das Potenzial für Kostensenkungen. Auch hinsichtlich einer Stärkung der Kundenorientierung, die eine systematische Erfassung und Analyse der Wünsche, Bedürfnisse und Erwartungen von Kunden erfordert, wird in der Durchführung von Big-Data-Analysen großes Potenzial gesehen. Zur Steigerung der Produktivität können sowohl Big-Data-Analysen als auch die Nutzung von Cloud Computing beitragen. Während im Zusammenhang mit Big Data die Herbeiführung besserer Entscheidungen hervorgehoben wird, steht beim Cloud Computing die Steigerung der organisatorischen Flexibilität im Mittelpunkt. Auch die Innovationsfähigkeit soll durch Big-Data-Analysen und die Nutzung von Clouddiensten verbessert werden.



Mit Blick auf *staatliche Stellen* und *Behörden*, die *Wissenschaft* oder auch *Privatpersonen* gibt es nur sehr wenige Studien, die sich mit sozioökonomischen Potenzialen von Big Data und Cloud Computing beschäftigen. Sieht man von Umsatzsteigerungen und dem Aufbau strategischer Wettbewerbsvorteile ab, die bei staatlichen Stellen und Behörden sowie bei wissenschaftlichen Einrichtungen eine untergeordnete Rolle spielen, ergeben sich in diesen Bereichen im Prinzip dieselben sozioökonomischen Potenziale durch Big Data und Cloud Computing wie bei Unternehmen. Neben Kostensenkungen sind Verbesserungen bei Kundenorientierung, Produktivität und Innovationsfähigkeit auf Grundlage von besseren Entscheidungen und höherer Flexibilität in allen Bereichen gefragt. Besonderes Potenzial für wissenschaftliche Einrichtungen wird einerseits im einfacheren Zugriff auf vorhandene Forschungsdaten und andererseits in der Möglichkeit gesehen, die primär theorie- und modellgetriebene Forschung vermehrt durch datengetriebene Ansätze unterstützen zu können.

Privatpersonen eröffnet die Nutzung von Clouddiensten komfortable Möglichkeiten beim Zugriff auf sowie beim Teilen von Daten. Zunehmend werden Lösungen zur Einbindung von Privatpersonen in die Entwicklung von neuen Produkten oder Dienstleistungen auf Basis von Cloud Computing bereitgestellt (Stichwort: Open Innovation), wovon sowohl die jeweiligen Organisationen als auch Privatpersonen profitieren können. Ähnlich stellt sich die Situation im Bereich Big Data dar, wo Privatpersonen zur Sammlung großer Datenmengen beitragen, selbst aber in der Regel nicht in deren Analyse eingebunden sind. Von den Ergebnissen der Datenauswertung profitieren wiederum potenziell beide, Organisationen und Privatpersonen.

Die *gesamtgesellschaftlichen* Potenziale von Big Data sind vielfältig und betreffen Bereiche wie Produktivität und Unternehmensgründungen, Wachstum und Beschäftigung, Innovationsfähigkeit, Nachhaltigkeit sowie Transparenz und Partizipation. Sowohl Cloud Computing als auch Big Data können zu Produktivitätssteigerungen in Unternehmen und anderen Organisationen sowie zu Unternehmensgründungen beitragen. Produktivitätssteigerungen werden in der Regel auf Kosteneinsparungen und eine positive Geschäftsentwicklung zurückgeführt. Im Hinblick auf Big Data gibt beispielsweise die OECD an, dass US-amerikanische Unternehmen, die ihre Entscheidungen wesentlich auf Datenanalysen stützen, eine 5 bis 6 % höhere Produktivität aufweisen. Unternehmen sollen durch die Nutzung von Cloud Computing 10 bis 30 % der Kosten einsparen können. Der Ausbau und die Weiterentwicklung existierender Unternehmen wirken sich zusammen mit Gründungen neuer Unternehmen tendenziell positiv auf Wachstum und Beschäftigung aus. Eine genauere Einschätzung der Größe dieser Effekte ist allerdings kaum möglich, da die Ergebnisse sehr stark von den verwendeten Daten, Modellen und Vorgehensweisen abhängen. Auch wenn es durchaus konkrete Hinweise auf positive Auswirkungen auf die Produktivität sowie auf Unternehmensgründungen gibt, bleibt die Frage nach der Größe dieser Effekte weitgehend unbeantwortet.

Produktivitätssteigerungen und Unternehmensgründungen und damit auch Wachstum und Beschäftigung sind eng mit *Innovationen* verbunden. Sowohl Big Data als auch Cloud Computing können nicht nur Teil technischer Innovationen sein, sondern auch im Hinblick auf Prozess- oder Marketinginnovationen eine wichtige Rolle spielen. So ermöglicht beispielsweise Big Data eine deutlich stärkere Kundenorientierung, insbesondere die Entwicklung individuell zugeschnittener Angebote. Grundsätzlich sollen Big-Data-Analysen in Verbindung mit neuen Produktionstechnologien dazu beitragen, die Massenproduktion zu flexibilisieren und zu individualisieren. Cloud Computing ermöglicht den Zugang zu IT-Ressourcen, die entweder früher nicht vorhanden waren oder die sich KMU oder Start-ups bis dahin nicht leisten konnten. Außerdem wird Cloud Computing großes Potenzial im Hinblick auf Energie- und Ressourceneffizienz als wichtige Nachhaltigkeitsanforderungen zugesprochen. Einerseits trägt Cloud Computing zu einer besseren Auslastung von Rechenzentren bei und andererseits erlaubt es Nutzern von vollwertig ausgestatteten Desktopcomputern, vermehrt sparsamere Thin-Clients oder mobile Geräte zu nutzen (Computer, die auf eine Verbindung zu entfernten Rechenressourcen angewiesen sind). Es muss allerdings davon ausgegangen werden, dass der generelle Energie- und Ressourcenverbrauch im IT-Bereich weiter zunehmen wird, sodass die Verbreitung von Cloud Computing bestenfalls zu einer Abflachung dieser Wachstumskurve führen wird.

Vor allem bei Big Data werden zwei Entwicklungen mit Bedeutung für Fragen der Transparenz und Partizipation gesehen. Einerseits geht es um die Idee, dass Nutzer selbst entscheiden (können) sollen, ob, wem und wie viele Informationen sie von sich preisgeben – und zu welchem Preis. Dabei ist weitgehend unklar, inwieweit eine Ökonomisierung von Privatheit gesellschaftlich gewünscht ist. Andererseits geht es um mit Open Data verbundene Ideen, dass Privatpersonen Einblick in Daten und Prozesse der öffentlichen Verwaltung nehmen können.

ANWENDUNGSPOTENZIALE

In der Vorstudie werden ausgewählte Anwendungspotenziale von Big Data unter Bezugnahme auf die fünf Bedarfswelder der Hightech-Strategie der Bundesregierung behandelt: Klima/Energie, Gesundheit/Ernährung, Mobilität, Sicherheit und Kommunikation.

Im Hinblick auf das Bedarfsweld *Klima/Energie* hat sich die Bundesregierung eine Reihe von Zielen gesetzt. Einerseits soll die Wissensbasis im Hinblick auf den Klimawandel erweitert und die Anwendung von Klimaschutztechnologien und Anpassungslösungen vorangetrieben werden. Andererseits soll die Energieversorgung auf eine nachhaltigere Basis gestellt werden. Für Big Data gibt es in diesem Zusammenhang viele verschiedene Anwendungspotenziale. Vor allem in den Bereichen Klimaforschung, Smart Cities, Smarte Energienetze und Umweltschutz spielt Big Data bereits heute eine wichtige Rolle. Am Beispiel von Hurri-



kan Sandy lässt sich der Nutzen von Wettervorhersagen auf Grundlage von Big-Data-Analysen verdeutlichen. Dank der Analyse großer Datenmengen ließ sich der Weg des Hurrikans bereits drei Tage, bevor er in den Vereinigten Staaten auf Land traf, relativ genau vorhersagen und damit ein noch größeres Unglück vermeiden. Die Genauigkeit war vor allem deshalb beeindruckend, da der Hurrikan im Hinblick auf Größe und Stärke ein nahezu einzigartiges Wetterereignis darstellte. Für die Berechnungen wurden neben einer leistungsfähigen IT-Infrastruktur Modelle verwendet, die Informationen über Ereignisse in der Vergangenheit, Informationen über die Physik der Atmosphäre und Informationen über aktuelle und zukünftige Wetterbedingungen kombinieren. Die Daten, die mit geografischen Informationssystemen zusammengeführt werden mussten, stammten beispielsweise von Satelliten, Wetterbojen oder Flugzeugmessungen.

Das Bedarfsfeld *Gesundheit/Ernährung* zielt vor allem auf die Entwicklung besserer Diagnoseverfahren und Therapien ab, aber auch Ansätze zur Prävention und zur Kostensenkung sind von Interesse. Vor allem bei der Erforschung von Krankheitsursachen und Behandlungsmethoden, für die Epidemiologie sowie im Bereich der individualisierten Medizin spielt die Analyse von Big Data bereits heute eine zentrale Rolle und bietet vielfältige Anwendungspotenziale. In der Vergangenheit sorgte vor allem der Einsatz des Expertensystems Watson von IBM für Aufsehen. Watson ist ein System aus dem Bereich der künstlichen Intelligenz, das auf die Beantwortung von Fragen abzielt, die in natürlicher Sprache eingegeben werden können. Big-Data-Analysen unterstützen dabei nicht nur die Diagnose von Erkrankungen, sondern auch die Auswahl der passenden Therapie. Der Schwerpunkt liegt derzeit im Bereich der Onkologie. Watson kennt bereits mehr als 600.000 Fakten und mehr als 2 Mio. Seiten einschlägiger wissenschaftlicher Zeitschriften. Außerdem kann Watson auf 1,5 Mio. Krankengeschichten zurückgreifen. Einer Untersuchung zufolge diagnostiziert Watson Lungenkrebs in 90 % der Fälle korrekt, im Vergleich dazu schaffen dies Ärzte nur in 50 % der Fälle. Watson kann auch als Clouddienst genutzt werden.

Im Hinblick auf das Bedarfsfeld *Mobilität* geht es der Bundesregierung einerseits um eine schnelle, sichere und komfortable und andererseits um eine ressourcenschonende Beförderung von Menschen und Gütern. Neben der Entwicklung von neuen Antriebssystemen zielen die Bemühungen vor allem darauf ab, Verkehrsinfrastrukturen intelligenter zu machen. In den Bereichen autonome Fahrzeuge und intelligente Transportsysteme spielt Big Data bereits heute eine Rolle. Während bei autonomen Fahrzeugen der Fokus zunächst darauf lag, diese in die Lage zu versetzen, mithilfe verschiedener Sensoren ihre Umgebung wahrzunehmen, ein Ziel anzusteuern und Kollisionen auf dem Weg zu vermeiden, wurde in den letzten Jahren die Kommunikation mit anderen Fahrzeugen und der Verkehrsinfrastruktur (Car-to-X-Kommunikation) immer wichtiger. Im Rahmen des sim^{TD}-Projekts wurde unlängst in einem großen Feldversuch im Rhein-Main-Gebiet gezeigt, dass Car-to-X-Kommunikation prinzipiell reif für den Alltagseinsatz ist.



Im Rahmen eines hybriden Systemansatzes kamen sowohl WLAN- als auch mobilfunkbasierte Kommunikation zum Einsatz. Eine Rolle spielten bei diesem Projekt nicht nur Informationen über die Verkehrslage und Informationen von Lichtsignalanlagen oder Wechselverkehrszeichen, sondern auch Umfeldinformationen, beispielsweise zum Wetter.

Im Bedarfsfeld *Sicherheit* geht es primär um den Schutz der Gesellschaft und ihrer Infrastrukturen vor Terrorismus, Sabotage, organisierter Kriminalität und den Folgen von Naturkatastrophen oder Unfällen. Da die Infrastrukturen hoch vernetzt und technisiert sind, spielt das Thema IT-Sicherheit eine wichtige Rolle. Vor allem in den Bereichen Kriminalitäts- und Terrorbekämpfung, Schutz vor Naturkatastrophen und IT-Sicherheit kommt Big Data bereits heute zum Einsatz. Beispielsweise lassen sich im Bereich der Kriminalitäts- und Terrorbekämpfung durch die Analyse großer Datenmengen aus unterschiedlichen Quellen Straftaten unter bestimmten Voraussetzungen vorhersagen. Systeme zur polizeilichen Vorhersage sind in der Vergangenheit vor allem in den USA sowie im Vereinigten Königreich auf Interesse gestoßen, aber auch das deutsche Bundeskriminalamt beschäftigt sich mit entsprechenden Ansätzen. Während IBM und Microsoft bereits seit einiger Zeit Lösungen für das Predictive Policing anbieten, sorgte die Software PredPol des gleichnamigen US-amerikanischen Start-ups zuletzt für besondere Aufmerksamkeit in der Öffentlichkeit. PredPol greift auf Statistiken früherer Ereignisse zurück. Relevant sind nicht nur verzeichnete Straftaten, sondern auch eingegangene Notrufe. Die Wissenschaftler von PredPol gehen davon aus, dass die Wahrscheinlichkeit, dass Straftäter es in derselben Gegend, in der sie erfolgreich waren, noch einmal versuchen, viel höher ist als die Wahrscheinlichkeit, dass sie sich nach einem Erfolg ein neues Zielgebiet suchen.

Das Bedarfsfeld *Kommunikation* zielt auf Information und Kommunikation als Grundlage effizienter Wertschöpfungsprozesse ab. Die Ziele dieses Bedarfsfelds werden von der Bundesregierung sehr weit gefasst und überschneiden sich teils mit Zielen, die auch im Rahmen anderer Bedarfsfelder verfolgt werden. Im Kern geht es um die Intensivierung und Ausweitung des Einsatzes von IKT, damit Deutschland seine Position als Hightech-Standort festigen und ausbauen kann. Das Bedarfsfeld Kommunikation bietet naturgemäß besonders große Anwendungspotenziale für Big Data. Besondere Aufmerksamkeit wurde zuletzt Big-Data-Anwendungen mit Fokus auf den Bereich Kundenorientierung, auf die intelligente Fabrik (Smart Factory) sowie den Hochfrequenzhandel an den Finanzmärkten zuteil. Aber auch bei der Optimierung von Wahlkampagnen konnte Big Data vor allem in den USA seine Stärken bereits unter Beweis stellen. Als Vorreiter im Bereich Smart Factory gilt beispielsweise der taiwanesischer Elektronikauftragshersteller Foxconn, der für Apple große Mengen an iPhones herstellt. Da Apple die elektronischen Komponenten bei iPhones häufig ändert, muss Foxconns Fertigung entsprechend anpassungsfähig sein, was nur durch den Einsatz von neuen Verfahren zur Datenanalyse in Verbindung mit einem weitgehend computergesteuerten Fertigungsprozess möglich ist.



HERAUSFORDERUNGEN

Die zentralen technischen Herausforderungen von Big Data betreffen den Aufbau der Analysearchitektur und den Umgang mit dem Datenmaterial. Während davon auszugehen ist, dass die Nutzung cloudbasierter Lösungen den Aufbau der Analysearchitektur aus Sicht des Nutzers in vielerlei Hinsicht vereinfacht, kommen neue Herausforderungen im Hinblick auf die Interoperabilität von Diensten dazu. Die betrieblichen Herausforderungen betreffen vor allem die Wirtschaftlichkeit von Einsatzszenarien und die eingeschränkte Verfügbarkeit von Experten mit einschlägigem Fachwissen, aber auch Widerstände bei Planung und Durchführung. Die Nutzung von Clouddiensten führt dazu, dass Nutzer deutlich weniger Kontrolle über die analysierten Daten und die Analyseergebnisse haben. Gleichzeitig kann angenommen werden, dass sich die Nutzung von Clouddiensten überwiegend positiv auf die Wirtschaftlichkeit von Einsatzszenarien auswirkt. Hier unterscheiden sich allerdings cloudbasierte Big-Data-Anwendungen nicht wesentlich von anderen Anwendungen, die in eine Cloud ausgelagert werden. Im Hinblick auf rechtliche Herausforderungen zeigt sich, dass sowohl bei Big Data als auch bei Cloud Computing hinsichtlich zahlreicher Fragen Rechtsunsicherheit herrscht bzw. sich das anwendbare Recht als wenig geeignet erweist und einer Anpassung an den Stand der Technik bedarf. Betroffen sind vor allem die Bereiche Datenschutz, Eigentum an Daten, Vertragsgestaltung, Haftung und Wettbewerb. Die gesellschaftlichen Herausforderungen, die Big Data und Cloud Computing mit sich bringen, sind vielfältig und betreffen Bereiche wie Vertrauen und Verlässlichkeit, Autonomie und Beeinflussung, Diskriminierung und Teilhabe, aber auch Vulnerabilität und Abhängigkeit. Die Ausbildung von Experten sowie die Förderung des Fachwissens auf Anbieter- und Nutzerseite (das sog. Humankapital) stellt die zentrale strukturelle Herausforderung im Kontext von Big Data dar. Durch die Nutzung von Clouddiensten kommen Herausforderungen in den Bereichen Infrastruktur/Breitbandversorgung und Internetgovernance sowie der Wettbewerbsfähigkeit der IT-Industrie hinzu.

TECHNISCHE HERAUSFORDERUNGEN

Aktuelle Studien identifizieren zahlreiche technische Herausforderungen von Big Data und Cloud Computing. Neben fehlendem Fachwissen, fehlenden überzeugenden Einsatzszenarien und Kostenaspekten spielen vor allem technische Probleme im Kontext von Big Data eine zentrale Rolle. Etwa ein Drittel der Unternehmen hat mit technischen Problemen zu kämpfen. Technisch anspruchsvoll werden Big-Data-Analysen genau durch jene Charakteristika, über die sich Big Data vorrangig definiert. In der Praxis gibt es vor allem Probleme beim Aufbau einer geeigneten Analysearchitektur, bei der Realisierung des Zugriffs von verschiedenen Endgerätetypen sowie bei der Erstellung einer für den Endnutzer angemessenen Benutzerschnittstelle. Auch im Rahmen der Experteninterviews wurden Herausforderungen im Zusammenhang mit den zentralen Big-Data-

Charakteristika und dem Aufbau geeigneter Analysearchitekturen angesprochen. Beispielsweise wurde betont, dass es im Big-Data-Bereich keine Lösungen gibt, die über alle Anwendungsfälle hinweg gut abschneiden. In der Regel ist es daher notwendig, eine für die jeweiligen Anforderungen geeignete Analysearchitektur aufzubauen. Die Experten stellten außerdem fest, dass der Aufbau von Analysearchitekturen vor allem deshalb so herausfordernd ist, weil sich die Anwendungsfälle und damit auch die Anforderungen von Anwendern im Laufe der Zeit weiterentwickeln, sich die hoch spezialisierten Big-Data-Lösungen aber nur sehr beschränkt an neue Anforderungen anpassen lassen. Mögliche Ursachen für die beschriebenen Schwierigkeiten sind neben fehlendem Fachwissen vermutlich einerseits in der Verwendung veralteter Technologien und andererseits in der mangelnden Reife neuer Big-Data-Technologien zu finden. Im Hinblick auf die mangelnde Reife neuer Big-Data-Technologien wurde im Rahmen der Interviews vor allem kritisiert, dass weiterführende, fortschrittliche Analysen nur sehr eingeschränkt möglich sind. Die vorhandenen Lösungen sind den befragten Experten zufolge nicht dafür geeignet, komplexes Data Mining durchzuführen, wie es beispielsweise für Cluster- oder Netzwerkanalysen erforderlich ist.

Neben dem Aufbau der Analysearchitektur erweist sich im Kontext von Big Data und Cloud Computing häufig auch der Umgang mit dem Datenmaterial als schwierig. Herausforderungen im Zusammenhang mit Datensicherheit, Datenschutz und Datenqualität haben zwar eine technische Komponente, sind aber nicht als ausschließlich technische Probleme zu verstehen. Sowohl Datensicherheit als auch Datenschutz werden als Kernherausforderungen im Umfeld von Big Data und Cloud Computing gesehen. Mangelnde Datenqualität ist vor allem im Hinblick auf die Durchführung von Big-Data-Analysen ein bekanntes Problem. Nicht selten wird für Big-Data-Analysen beispielsweise auf Archive zurückgegriffen, die ursprünglich für eine Auswertung nicht vorgesehen waren. Eine angemessene Aufbereitung umfangreicher Datenmengen ist häufig mit großem Aufwand verbunden, im Hinblick auf Vertrauenswürdigkeit der Ergebnisse aber unverzichtbar. Im Hinblick auf die Nutzung von Cloud Computing stellt neben Problemen in den Bereichen Datensicherheit und Datenschutz mangelnde Interoperabilität eine zentrale technische Herausforderung dar. Interoperabilität ist nicht nur im Hinblick auf die Integration von Cloudlösungen mit Inhouselösungen ein Thema, sondern auch im Hinblick auf die Integration von Cloudlösungen untereinander.

BETRIEBLICHE HERAUSFORDERUNGEN

Abgesehen von technischen Herausforderungen sehen sich Organisationen sowohl bei der Durchführung von Big-Data-Analysen als auch bei der Nutzung von Cloud Computing mit betrieblichen Herausforderungen konfrontiert. Im Kontext von Big Data gehören dazu primär fehlende ökonomisch und technisch überzeugende Einsatzszenarien, mangelndes Fachwissen, Datenschutzbedenken,



aber auch relativ hohe Kosten und Rivalitäten zwischen Abteilungen spielen eine nicht unbedeutende Rolle. Im gegenwärtigen Wirtschaftsumfeld haben Organisationen nur eine geringe Motivation, in neue Technologien zu investieren, solange sich keine messbaren Vorteile ergeben. Daher überrascht es nicht, dass in vielen Organisationen der Investitionswille fehlt. Im Hinblick auf betriebliche Herausforderungen spielen auch Defizite beim erforderlichen Fachwissen eine wichtige Rolle. Im Rahmen der Experteninterviews wurde unterstrichen, dass die derzeit entstehenden Lehr- und Studienangebote allein aufgrund der Vielschichtigkeit der Anforderungen an entsprechende Mitarbeiter nicht in der Lage sein werden, das Defizit auszugleichen und für eine Entspannung am Arbeitsmarkt zu sorgen.

Auch Rivalitäten und mangelndes Vertrauen zwischen Abteilungen spielen eine wichtige Rolle im Hinblick auf die zentralen Herausforderungen bei der Einführung und Nutzung von Big-Data-Lösungen. Einige Studien identifizieren die mangelnde Bereitschaft von Fachbereichen, »ihre« Daten mit anderen zu teilen, als eines der wesentlichsten Probleme. Im Hinblick auf die Planung gelten einerseits die Identifikation der wichtigsten Investitionsbereiche und andererseits die Organisation von Big-Data-Aktivitäten als Herausforderung. Bei der Durchführung werden die Auswahl der »richtigen« Daten für eine bestimmte Fragestellung und der »richtigen« Verwendung der Analyseergebnisse als besonders herausfordernd angesehen. Eine zentrale betriebliche Anforderung im Zusammenhang mit der Nutzung von Cloud Computing stellt die Tatsache dar, dass die Vertrauenswürdigkeit von Anbietern nur schwer feststellbar ist. Mangelndes Vertrauen stellt für rund ein Viertel der potenziellen Nutzer von Clouddiensten eine wesentliche Barriere dar. Ein Problem wird auch in der unzureichenden Kontrollierbarkeit der Einhaltung von Vereinbarungen gesehen. Damit eng verbunden sind Herausforderungen im Hinblick auf Datensicherheit und Datenschutz. Insbesondere die Angst vor Datenverlust ist im Umfeld von Public-Cloud-Angeboten groß. Darüber hinaus werden die Angst von einem Verlust an IT-Fachwissen aufgrund der Auslagerung und fehlende Unterstützung vor Ort als Barrieren genannt.

RECHTLICHE HERAUSFORDERUNGEN

Neben technischen und betrieblichen Herausforderungen spielen für Unternehmen und staatliche Stellen und Behörden, die sich mit der Durchführung von Big-Data-Analysen bzw. der Nutzung von Cloud Computing beschäftigen, rechtliche Herausforderungen eine zentrale Rolle. Im Kontext von Big Data und Cloud Computing stellen sich vor allem Fragen, die mit dem Datenschutz, einem möglichen Eigentum an Daten, der Ausgestaltung von Verträgen, einer möglichen Haftung und dem Wettbewerb zusammenhängen. Besonders problematisch ist, dass hinsichtlich zahlreicher Fragen Rechtsunsicherheit herrscht bzw. sich das anwendbare Recht als wenig geeignet erweist und einer Anpassung an den

Stand der Technik bedarf. Datenschutz spielt als Herausforderung sowohl im Kontext der Durchführung von Big-Data-Analysen als auch bei der Nutzung von Cloud Computing eine wichtige Rolle. In Deutschland wird der Datenschutz durch das Bundesdatenschutzgesetz (BDSG) geregelt. Nach dem BDSG steht die Verarbeitung von personenbezogenen Daten unter einem Verbot mit Erlaubnisvorbehalt. Cloud Computing stellt einen typischen Fall der Auftragsdatenverarbeitung dar. Dem BDSG zufolge ist der Auftraggeber dafür verantwortlich, dass die Vorschriften des Gesetzes eingehalten werden und muss dies im Rahmen der technischen und rechtlichen Ausgestaltung der Datenverarbeitung sicherstellen. Aus rechtlicher Sicht ist einerseits die Ausgestaltung entsprechender vertraglicher Vereinbarungen schwierig, andererseits sind Art und Umfang der gesetzlichen Kontrollpflichten weitgehend unklar. Noch herausfordernder wird die Nutzung von cloudbasierten Lösungen aus Datenschutzperspektive bei grenzüberschreitendem Bezug. Wann ein ausländischer Anbieter überhaupt deutsches Datenschutzrecht beachten muss, ist noch völlig ungeklärt. Auch wenn sich im Kontext von Cloud Computing die Frage stellt, inwiefern sich der Standort, an dem sich bestimmte Daten befinden oder wo sie erhoben wurden, verlässlich ermittelt werden kann, gilt das Territorialitätsprinzip.

Aus juristischer Sicht kann kein »Eigentum« an Daten oder Informationen begründet werden. Dennoch können Daten oder Informationen unter verschiedenen gesetzlichen Voraussetzungen vom Zugriff oder der Verwertung durch Dritte geschützt sein. In Deutschland sind neben Nutzungsbedingungen vor allem das Urheberrecht und das ebenfalls im Urheberrechtsgesetz (UrhG) zu verortende Datenbankrecht relevant. Da nur der Datenträger, nicht aber die Daten selbst körperlich sind, können Daten nicht als Sachen im Sinne des § 90 des Bürgerlichen Gesetzbuchs (BGB) qualifiziert werden. Nach § 93 BGB sind Bestandteile, die nicht von einer Sache getrennt werden können, wesentliche Bestandteile. So betrachtet würde das Eigentum an einem Datenträger auch das Eigentum an den darauf gespeicherten Daten umfassen. Im Hinblick auf die tatsächlichen Verhältnisse und insbesondere den Bereich Cloud Computing wäre dies kaum praktikabel. Auch könnten viele Probleme, die bei vernetzten Datenbeständen auftreten, nicht gelöst werden. Sofern keine legitimierenden Schranken greifen, hängt die Zulässigkeit einer urheberrechtlich relevanten Handlung von der Zustimmung des Rechteinhabers ab. Eine besondere Herausforderung im Kontext von Big Data stellt die Tatsache dar, dass in vielen Fällen weder legitimierende Schranken greifen noch eine vertragliche Einräumung von Nutzungsrechten mit angemessenem Aufwand möglich ist. Während bei Big-Data-Lösungen, die innerhalb Deutschlands genutzt werden, in der Regel deutsches Urheberrecht zu beachten sein wird, ist die Beurteilung in Fällen mit grenzüberschreitendem Bezug oft schwierig. Im Rahmen der Experteninterviews wurde betont, dass im Kontext von Big Data Urheberrechtsfragen vor allem dann nicht ausreichend beantwortet sind, wenn Daten aggregiert oder wenn neue Daten aus vorhande-



ZUSAMMENFASSUNG

nen Daten abgeleitet werden. Bei Big-Data-Anwendungen ist das allerdings eher die Regel als die Ausnahme. Die Verteilung der Rechte und Pflichten aus einem Vertragsverhältnis liegt nach dem Grundsatz der Vertragsfreiheit in den Händen der beteiligten Parteien. Trotzdem ist die Vertragsgestaltung im Kontext von Big Data aufgrund der zahlreichen technischen Fragen und der Berührungspunkte zum Datenschutz- und Urheberrecht kompliziert. Grundsätzlich ist es nicht abwegig, von einer mietrechtlichen Beurteilung eines Cloud-Computing-Vertrags auszugehen. Unklar ist allerdings, ob auf das gesetzliche Leitbild des Mietrechts geschlossen werden könnte, wenn bestimmte Aspekte nicht vertraglich geregelt wurden. Aufgrund der zunehmenden Bedeutung von Daten in Entscheidungsprozessen und in Anbetracht der daraus entstehenden Gefahren ist Klarheit im Hinblick auf Fragen der Haftung im Kontext von Big Data unerlässlich. Eine besondere Herausforderung in diesem Zusammenhang stellt die Tatsache dar, dass es unklar ist, ob es qualitative Mindeststandards für Informationen gibt und welche das sind. Aus wettbewerbsrechtlicher Sicht stellt die Tatsache, dass Cloudanbieter Hindernisse aufbauen können, die den Wechsel zu einem anderen Anbieter erschweren (sog. Vendor-Lock-in), ein zentrales Problem dar.

GESELLSCHAFTLICHE HERAUSFORDERUNGEN

Die gesellschaftlichen Herausforderungen, die Big Data und Cloud Computing mit sich bringen, sind vielfältig und betreffen Bereiche wie Vertrauen und Verlässlichkeit, die Filterung beziehungsweise Einschränkung von Inhalten, Analysen und Ergebnissen, Autonomie und Beeinflussung, Diskriminierung und Teilhabe, aber auch Vulnerabilität und Abhängigkeit. Die Nutzung von Public-Cloud-Diensten führt in der Regel dazu, dass ein Nutzer nicht nachvollziehen kann, wo und wie seine Daten gespeichert und verarbeitet werden. Das daraus resultierende Gefühl des Kontrollverlustes ist eine wesentliche Ursache für die Vorbehalte gegenüber Cloud Computing. Während bei Cloud Computing das Vertrauen in die Verlässlichkeit des Anbieters und seine technologischen Fähigkeiten im Vordergrund steht, spielt bei Big Data das Vertrauen in die Technologie und die zugrundeliegenden Prinzipien eine besondere Rolle. Von Bedeutung sind dort vor allem statistische Verfahren, die auf Korrelationen als Messgröße für Zusammenhänge setzen. Ergebnisse aus Big-Data-Analysen führen in der Regel zu einer Verengung des Sichtbaren. Das kann einerseits bedeuten, dass durch Filterung primär Informationen angezeigt werden, die für den Nutzer relevant sind und damit häufig bereits bestehende Meinungen unterstützen, andere Blickwinkel aber völlig außen vor gelassen werden. Andererseits kann es aber auch bedeuten, dass es zur Überanpassung (Overfitting) von Modellen an vorhandene Daten kommt und dadurch die Vorhersageeigenschaften des Modells im Hinblick auf die Wirklichkeit beeinträchtigt werden.

Die Verengung des Sichtbaren ist eng verbunden mit einer weiteren zentralen gesellschaftlichen Herausforderung, einem möglichen Verlust der Autonomie

durch umfassende Möglichkeiten der Beeinflussung. Big Data erlaubt die Vorhersage des Verhaltens von Gruppen und Individuen und damit letztlich auch deren gezielte Beeinflussung. Da in der Regel konkrete Verhaltensoptionen vorgeschlagen werden und damit die Auswahlmöglichkeit eingeschränkt wird, geht der Verlust der Autonomie über eine bloße Verengung des Sichtbaren hinaus. Aus gesellschaftlicher Sicht ist dabei aber vor allem problematisch, dass sich die Möglichkeiten der Verhaltensvorhersage nicht nur auf die Autonomie einzelner Individuen auswirken, sondern dass diese Möglichkeiten genutzt werden können, um die gesellschaftliche Entwicklung zu beeinflussen. Eng verwandt mit der Frage nach der Autonomie ist die Möglichkeit, Big-Data-Analysen zur Diskriminierung von Personen zu nutzen. Einerseits werden auf Grundlage der bestehenden Datenlage Vorhersagen für einzelne Personen oder Personengruppen getroffen, die diese nicht beeinflussen können. Im Gegensatz zur bloßen Beeinflussung geht es hierbei darum, dass nicht nur Handlungen der Personen beeinflusst werden sollen, sondern dass aus den Vorhersagen vorbeugende Maßnahmen abgeleitet werden, welche die Entfaltungs- und Handlungsspielräume der Betroffenen einschränken und ohne deren Wissen oder Einspruchsmöglichkeit umgesetzt werden. Andererseits kann es aber auch sein, dass, wenn keine oder nur wenige Daten über Individuen vorhanden sind, ihre Interessen ignoriert werden. Diese Form der Diskriminierung kann in einer zunehmend vernetzten Gesellschaft zu verminderter Teilhabe führen.

Mit der zunehmenden Vernetzung von kritischen Infrastrukturen in Bereichen wie Energie, Telekommunikation, Gesundheit und Wasser nimmt sowohl die Vulnerabilität als auch die Abhängigkeit von komplexen IT-Systemen zu. Gezielte Angriffe mit Schadprogrammen wie Stuxnet und Duqu zeigten, wie anfällig IT-Systeme in der Industrie oder Infrastrukturen für Angriffe sind. Durch die zunehmende Automatisierung und Vernetzung dieser Systeme entstehen neue Angriffsmöglichkeiten, bei denen Big Data und Cloud Computing eine zentrale Rolle spielen können. Eng verbunden ist damit die Diskussion um die »technologische Souveränität« eines Staates, die als eine mögliche Ursache von Abhängigkeiten und damit auch Vulnerabilität gesehen wird.

STRUKTURELLE HERAUSFORDERUNGEN

Strukturelle Herausforderungen betreffen in der Regel die Rahmenbedingungen, die einen direkten oder indirekten Einfluss auf die Möglichkeiten haben, die Potenziale von Big Data und Cloud Computing zu nutzen. Die mangelnde Verfügbarkeit von Humankapital, d.h. Experten und Fachwissen auf Entwickler- wie Anwenderseite, stellt vor allem im Hinblick auf die Durchführung von Big-Data-Analysen eine zentrale strukturelle Herausforderung dar. Durch die Nutzung von Cloud Computing kommen Fragestellungen in den Bereichen Infrastruktur/Breitbandversorgung und Internetgovernance hinzu. So ist die Verfügbarkeit von ausreichender Bandbreite sowie deren Finanzierung ein Faktor, der die Aus-



nutzung von möglichen Potenzialen beeinflusst. Die Enthüllungen von Edward Snowden haben in Verbindung mit Entwicklungen in den Bereichen Big Data und Cloud Computing dazu geführt, dass die Debatte über die Steuerung (Governance) des Internets wieder stark an Bedeutung gewonnen hat. Wesentlicher Grund dafür ist, dass für den Erfolg von Big Data in der Cloud ein hohes Vertrauen in die Technologie und die Anbieter notwendig ist. Gerade die Abgabe von Kontrolle über eigene Daten und die Möglichkeiten des Missbrauchs sind wesentliche Aspekte, die als ernsthafte Barrieren identifiziert wurden. In letzter Zeit wird zunehmend Kritik an der Dominanz der USA bei Fragen der Internetgovernance geäußert. Unter digitaler Wettbewerbsfähigkeit wird die Fähigkeit eines Standortes verstanden, in neuen Märkten wie beispielsweise in jenen für Cloud Computing und Big Data relevante, globale Akteure hervorzubringen. Im Zusammenhang mit dem Mangel an innovativen, schnell wachsenden Unternehmen in Europa werden verschiedene Einflussfaktoren diskutiert. Ein erster Faktor ist die Marktfragmentierung, welche die Wachstumsmöglichkeiten von Unternehmen beschränkt. Internationalisierung ist deshalb für europäische Unternehmen viel schneller ein Thema als für US-amerikanische Unternehmen. Neben rechtlichen Herausforderungen spielen auch andere Faktoren wie Sprachbarrieren oder kulturelle Unterschiede eine wesentliche Rolle. Ein weiterer Faktor ist das geringe Niveau der unternehmerischen Aktivitäten in Europa im Vergleich zu den USA. Gründe hierfür sind sowohl mangelnde Kommunikation zwischen den Akteuren in Wirtschaft, Politik und Forschung als auch mangelndes Finanzkapital, welches für die Gründung und das Wachstum von Unternehmen in größerem Ausmaß notwendig wäre.

HANDLUNGS- UND FORSCHUNGSBEDARF

Auch wenn Big-Data-Analysen ohne Cloud Computing durchgeführt werden können und Cloud Computing neben Big-Data-Analysen zahlreiche andere Anwendungsfelder hat, wird der Kombination von Big Data und Cloud Computing großes Potenzial zugeschrieben. Während sich bei einigen Big-Data-Anwendungen die Nutzung von Clouddiensten anbietet, ist das bei anderen nicht der Fall. Ein wesentlicher Punkt ist, dass es sich insbesondere für große Datenmengen empfiehlt, die Daten dort zu analysieren, wo sie anfallen, da der Transport großer Datenmengen ein massives technisches Problem darstellt. Aus diesem Grund ist die cloudbasierte Analyse großer Datenmengen insbesondere dann sinnvoll, wenn die Daten auch im Netz entstehen oder dort bereits aggregiert wurden. Abgesehen davon bringt die Nutzung von Clouddiensten im Zusammenhang mit der Durchführung von Big-Data-Analysen vor allem dann Vorteile, wenn im Vorfeld eines Big-Data-Vorhabens vorübergehend mit den vorhandenen Daten experimentiert werden soll oder wenn bei einem Analysevorhaben mit vorübergehenden Anforderungsspitzen zu rechnen ist. Bei den folgenden sieben Themenbereichen mit dazugehörigen Fragestellungen ist davon auszugehen, dass sie zu-

künftig im Kontext von Big Data in der Cloud für Wirtschaft, Politik, Forschung und Gesellschaft besonders relevant sein werden. Sich daraus ergebende zentrale Handlungs- und Forschungsbedarfe sowie mögliche Optionen werden im Folgenden erläutert und skizziert.

Für Big Data und Big Data in der Cloud sind vielfältige *Anwendungsfälle und Einsatzszenarien* denkbar. Allerdings sind die konkreteren Chancen und Risiken von Big Data in der Cloud genauso wie die Beschaffenheit geeigneter Einsatzszenarien in vielen Fällen unbekannt. Handlungs- und Forschungsbedarf besteht insbesondere im Hinblick auf die Förderung des Erfahrungsaustausches, den Einsatz in der öffentlichen Verwaltung und die Berücksichtigung der Interessen von Privatpersonen.

Obwohl die für Big Data und Big Data in der Cloud erforderliche *technische Infrastruktur* im Prinzip vorhanden ist, gibt es in der Praxis oft Probleme, die aufwendige individuelle Anpassungen erforderlich machen. Daraus ergibt sich mit Blick auf die technische Infrastruktur ein Handlungs- und Forschungsbedarf bei der Unterstützung fortschrittlicher Analysen, wie sie z. B. für Cluster- oder Netzwerkanalysen erforderlich sind, bei der Erweiterung des möglichen Nutzerkreises durch die Reduktion der Anforderungen an Nutzer und beim Ausbau der Netzwerkinfrastruktur.

Auch wenn häufig von einer *Ökonomie der Daten* die Rede ist, ist der Wert von Big Data für Organisationen schwer quantifizierbar, unabhängig davon, ob entsprechende Analysen auf Basis von Cloud Computing genutzt werden sollen oder nicht. Besonderer Handlungs- und Forschungsbedarf existiert in Bezug auf die Entwicklung von Ansätzen zur Bewertung von Investitionen in Daten und Analysearchitekturen, die Wiederverwendung von Daten und die Folgen der Ökonomisierung von Privatheit, also der Bereitschaft von Privatpersonen, für eine Gegenleistung auf Privatheit zu verzichten.

Die Auswertung umfassender Datenbestände wird in Zukunft immer mehr die *Wettbewerbsfähigkeit* von Organisationen bestimmen. Wer den Anschluss verliert, passt sich langsamer an die zunehmend dynamischen Märkte an und hat dadurch einen klaren Nachteil. Im Hinblick auf die technische Infrastruktur besteht Handlungs- und Forschungsbedarf insbesondere im Hinblick auf Entwicklungen in anderen Ländern und Regionen sowie die Gründungs- und Wachstumsmöglichkeiten neuer IT-Unternehmen.

Die Vielfalt der *Kompetenzen*, die *Fachkräfte* wie Data Scientists haben müssen, macht es insbesondere für KMU und die öffentliche Verwaltung schwierig, geeignetes Personal zu finden. Mit der zunehmenden Bedeutung von Big Data spielt vor allem die Informationskompetenz, also der selbstbestimmte, souveräne, verantwortliche und zielgerichtete Umgang mit Informationen, eine immer wichtigere Rolle. Handlungs- und Forschungsbedarf ergibt sich daher bei der



ZUSAMMENFASSUNG

Entwicklung einer benutzerfreundlichen Abfragesprache, der Überarbeitung der Curricula für die Ausbildung von Data Scientists und der Erhöhung der Informationskompetenz insgesamt in der Gesellschaft.

In verschiedenen Bereichen sind die *rechtlichen Rahmenbedingungen* den Anforderungen von Big Data und Cloud Computing nicht gewachsen. Besonders relevant sind beispielsweise Urheber-, Datenschutz- und Haftungsrecht. Aus diesem Grund besteht ein Handlungs- und Forschungsbedarf vor allem im Zusammenhang mit der Untersuchung des Anpassungsbedarfs auf Ebene der Gesetzgebung, der Untersuchung der Auswirkungen der EU-Datenschutzgrundverordnung, der Schaffung von Klarheit im Hinblick auf Eigentum an Daten und dem Vorantreiben von Harmonisierungen auf EU-Ebene.

Bei Überlegungen im Bereich Big Data in der Cloud spielen *Sicherheit und technischer Datenschutz* eine zentrale Rolle. Für an sich lohnend erscheinende Anwendungen wird das Thema Sicherheit mitunter zu einem Ausschlusskriterium. Besonderer Handlungs- und Forschungsbedarf ergibt sich daher bei der Umsetzung von Maßnahmen zur Gewährleistung der Informationssicherheit sowie der Festlegung von Sicherheit und Datenschutz als explizite Anforderungen bei der Entwicklung von Lösungen.



EINLEITUNG

I.

DIGITALE GESELLSCHAFT UND KONVERGENZ VON TECHNOLOGIEN

1.

Themen wie die »Digitalisierung unseres Lebens« oder die »digitale Revolution« sind in den letzten 15 Jahren zum festen Bestandteil der öffentlichen Diskussion geworden. Die Ansichten über die Auswirkungen neuer Informations- und Kommunikationstechnologien (IKT) auf Wirtschaft, Politik, Forschung und Gesellschaft waren immer wieder von deutlichen Gegensätzen geprägt. Vielfach waren die in neue Plattformen, Dienste oder Geräte gesetzten Erwartungen zunächst sehr groß, bevor dann rasch Ernüchterung einsetzte. Häufig ging es zwei Schritte vor und einen zurück. Ein eindrucksvoller Beleg für dieses Phänomen ist beispielsweise das Platzen der Dotcom-Blase nach der ersten Euphorie rund um die »New Economy« Ende der 1990er Jahre. Weitgehend unabhängig von der öffentlichen Diskussion und trotz aller Rückschläge schritt die Digitalisierung des Alltags jedoch immer weiter voran. Neue IKT spielen für die Bewältigung alltäglicher Aufgaben eine immer bedeutendere Rolle. Gleichzeitig wachsen die Technologien durch Entwicklungen wie intelligente Netze oder Maschinenkommunikation zunehmend zusammen. Big Data ist neben Mobile Computing, Social Media und Cloud Computing ein aktueller Kulminationspunkt dieser Entwicklungen.

Die Menge an produzierten bzw. verfügbaren Daten hat in den letzten Jahren rasant zugenommen. Häufig ist gar von einer Daten- oder Informationsexplosion die Rede (z.B. Beath et al. 2012; Schrader 2011). Die Analyse von großen Datenmengen führt nicht zwangsläufig zu besseren Entscheidungen. Um den erhofften Erfolg zu erbringen, müssen bei der Analyse zahlreiche Aspekte berücksichtigt werden. Dies betrifft sowohl die zu analysierenden Daten als auch den Zweck der Auswertung. Dennoch wird Big Data als ein nicht aufzuhaltender Treiber eines revolutionären Wandels in der Geschäftswelt gesehen (Schrage 2012). Immer öfter ist von »Data-Driven Companies« (EIU 2012, S.6) oder »Digital Enterprises« (Carr 2001) die Rede. Big Data spielt allerdings auch über die Geschäftswelt hinaus eine nicht unbedeutende Rolle. Insbesondere in der Wissenschaft eröffnet die effiziente Analyse großer Datenmengen ungeahnte Möglichkeiten. Beeindruckend sind in diesem Zusammenhang beispielsweise die Leistungen von IBMs Watson, einem System aus dem Bereich der künstlichen Intelligenz, in der Fernsehquizshow »Jeopardy!« im Jahr 2011 (Best 2013). Watson wurde in der Folge weiterentwickelt und insbesondere im Gesundheitswesen in den USA eingesetzt. Bekannte Anwender des Systems sind der Verwaltungsdienstleister WellPoint (IBM 2013b) und das Memorial Sloan-Kettering Cancer Center (IBM 2013a).



I. EINLEITUNG

Zunächst wurden die Möglichkeiten, die sich durch die Analyse großer Datenmengen ergeben, vor allem positiv wahrgenommen. Ein Beispiel dafür liefert Viktor Mayer-Schönberger, Professor für Internetgovernance am Oxford Internet Institute, der sich Ende 2013 im Rahmen eines Interviews mit der Süddeutschen Zeitung von der Tatsache, dass große Datenmengen zunehmend gesammelt und ausgewertet werden, fasziniert zeigte (Hauck 2013): »Das Beeindruckende daran ist, dass die Daten, anders als Goldminen, nie versiegen, sie wachsen quasi nach. Und sie verändern die Welt, so wie wir sie sehen. Durch die Masse an Daten können wir besser verstehen, wie die Wirklichkeit in all ihrer Komplexität tatsächlich ist.«

Insgesamt entspricht dies den positiven Konnotationen zu Big Data, die etwa seit dem Jahreswechsel 2010/2011 einen großen Teil der öffentlichen Diskussion ausmachen. Ähnliche Kommentare waren auch von zahlreichen anderen Vertretern aus Wirtschaft, Politik und Forschung zu hören und fanden sich auch in Stellungnahmen von Branchenverbänden wieder. Vor allem seit den Enthüllungen rund um die Aktivitäten der in den USA ansässigen National Security Agency (NSA) und anderer Geheimdienste im Sommer 2013, aber auch aufgrund zahlreicher anderer Zwischenfälle der jüngeren Vergangenheit, bei denen Unternehmen allzu sorglos mit dem Thema Datenschutz umgingen, finden kritische Stimmen, wie sie beispielsweise in den Arbeiten von Frank Schirrmacher zu finden sind (Schirrmacher 2013), vermehrt Zuspruch: »Das neue Zeitalter von Big Data erschafft die größte Überwachungsmaschine, die es jemals gab.«

Während diese Kritik noch vorwiegend auf den Überwachungscharakter abzielt, gehen andere – wie der Publizist Evgeny Morozov – noch weiter, indem sie darauf hinweisen, dass diese Entwicklung nicht nur Überwachung beinhaltet, sondern viel tiefgreifender ist. Aus ihrer Sicht geben die hinter Big-Data-Analysen stehenden Technologien die Möglichkeit zum »präemptiven Steuern«, wo gewünschtes Verhalten des Bürgers durch bestimmte, politisch gesteuerte Anreizsetzung erreicht werden soll. Dies entspräche Morozovs Ansicht nach einer Entmündigung des Bürgers (Morozov 2011).

Auch wenn Morozovs Ansicht auf den ersten Blick vielleicht etwas weit hergeholt wirkt, geht die öffentliche Diskussion im Zusammenhang mit den Möglichkeiten zur Analyse großer Datenmengen durchaus in diese Richtung. Während beispielsweise der Einsatz von IBMs Watson im Gesundheitswesen zweifelsohne als positiv zu beurteilen ist, sind durchaus auch Anwendungsszenarien denkbar, bei denen das nicht uneingeschränkt der Fall ist. Die Sicherstellung des Datenschutzes ist eine der zentralen Herausforderungen im Umfeld von Big Data. In welchem Ausmaß z. B. Persönlichkeitsmerkmale wie sexuelle Orientierung oder Drogenkonsum auf Grundlage von Facebook-Profilen rekonstruiert werden können, zeigt eine Studie von Kosinski et al. (2013). Über den Einsatz von Big-Data-Analysen in bestimmten Bereichen gehen die Meinungen auseinander.

Während z. B. die Prävention von Verbrechen durch erhöhte Polizeipräsenz an bestimmten Punkten durchaus sinnvoll erscheint, wird beispielsweise die Berechnung der Rückfallwahrscheinlichkeit von einzelnen Straftätern und damit der Möglichkeit auf vorzeitige Entlassung aus dem Vollzug auf Grundlage von Daten über andere Straftäter eher kritisch gesehen.

Diese Beispiele zeigen das große Spannungsfeld auf, in dem sich die aktuelle Diskussion um Big Data abspielt. Ähnliche Muster, wenn auch teilweise weniger scharf, zeigen sich bei Trends wie Mobile Computing, Social Media und Cloud Computing, wo ebenso Themen wie Privatsphäre und Informationssicherheit im Vordergrund stehen. Im Rahmen dieser Vorstudie liegt das Hauptaugenmerk auf Big Data und Cloud Computing. Während Big Data umfassend behandelt wird, liegt der Fokus bei Cloud Computing primär auf seiner Funktion als Treiber und Lösungstechnologie (Abolhassan 2013) zur Durchführung von Big-Data-Analysen. Cloud Computing erlaubt die flexible Bereitstellung einer skalierbaren Infrastruktur, wie sie unter anderem für die rasche Analyse großer Datenmengen benötigt wird. Durch die Kombination von Big Data und Cloud Computing können sich sowohl deren Potenziale als auch die mit den Trends verbundenen Herausforderungen gegenseitig verstärken oder abschwächen. Mobile Computing und Social Media spielen in dem Sinne eine Rolle, als dass sie gegenwärtig zu den wichtigsten Datenquellen für Big-Data-Analysen gehören. Aufgrund der Interdependenzen der hinter den Trends stehenden Technologien erscheint eine gemeinsame Betrachtung sinnvoll. Im Rahmen der Studie sollen unterschiedliche Positionen aufgegriffen, wissenschaftlich bewertet und auf Grundlage neuester Erkenntnisse diskutiert werden.

Grundsätzlich werden in der Kombination von Big Data und Cloud Computing verschiedene und große Potenziale gesehen. Am augenscheinlichsten ist, dass Organisationen, insbesondere kleinere und mittlere Unternehmen (KMU), aber auch Privatpersonen durch Cloud Computing die Möglichkeit bekommen, effizient große Datenmengen zu analysieren, ohne vorher eine größere IT-Infrastruktur aufbauen zu müssen. Eine Studie des Branchenverbands der deutschen Informations- und Telekommunikationsbranche (BITKOM 2012, S.47) belegt, dass die Durchführung von Big-Data-Analysen, vor allem in den Bereichen Hardware und Infrastruktur, eine starke Investitionskomponente hat.

ZIEL UND AUFBAU DES BERICHTS

2.

Ziel dieser Vorstudie ist die Bewertung der Potenziale und Herausforderungen, die von Big Data und Cloud Computing ausgehen. Im Rahmen einer detaillierten Analyse und Bewertung setzt sich diese Studie intensiv mit den teils widersprüchlichen Positionen auseinander. Dabei steht eine ausgewogene, wissen-



I. EINLEITUNG

schaftliche Betrachtung von Potenzialen und Herausforderungen im Vordergrund, die im Hinblick auf eine weiterführende Diskussion eine möglichst objektive Betrachtung von Chancen und Risiken ermöglichen soll. Darauf aufbauend sollen Handlungs- und Forschungsbedarfe identifiziert sowie Optionen entwickelt werden, die es Wirtschaft, Politik, Forschung und Gesellschaft ermöglichen, die sich bietenden Chancen zu nutzen.

Um dieses Ziel zu erreichen, werden in einem ersten Schritt die Grundlagen von Big Data und Cloud Computing erläutert (Kap. II). Die Erläuterung umfasst eine Klärung des Verständnisses der oft nur sehr vage definierten und unpräzise verwendeten Begriffe »Big Data« und »Cloud Computing«, eine kurze Darstellung der technischen Grundlagen und eine Diskussion der technologischen Interdependenzen. Ein grundlegendes Verständnis der Technologien ist als Basis für die sich anschließende detaillierte Analyse und Bewertung notwendig. Darüber hinaus werden wesentliche Anwendungsmöglichkeiten der jeweiligen Technologien anhand von ausgewählten Beispielen aufgezeigt.

Im Anschluss daran liegt der Fokus auf der Angebotsseite (Kap. III). Näher betrachtet und diskutiert werden die Märkte für Big Data und Cloud Computing sowie die jeweilige Anbieterlandschaft und deren Produkte und Dienstleistungen. Der Schwerpunkt liegt auf der Situation in Deutschland, jedoch wird auch die Situation in anderen Staaten Europas, den USA und Asien beleuchtet. Ebenso wird die Nachfrageseite eingehend betrachtet (Kap. IV); die Diffusion und Nutzung von Big Data und Cloud Computing in den verschiedenen Anwendungsbereichen wird auf Grundlage aktueller Studien untersucht und diskutiert. Zu den Anwendungsbereichen zählen Unternehmen, öffentliche Verwaltung, Wissenschaft und Privatpersonen. Wie bei der Betrachtung der Angebotsseite liegt der Fokus auch bei der Nachfrageseite auf der Situation in Deutschland, jedoch werden auch andere Staaten und Regionen sowie die Entwicklungen auf Ebene der Europäischen Union (EU) miteinbezogen. Diese vertiefte Betrachtung der Angebots- und Nachfrageseite zielt darauf ab, den gegenwärtigen Stand der Marktentwicklung für die jeweiligen Technologien herauszuarbeiten und realistische Entwicklungstendenzen für die nächsten Jahre aufzuzeigen. Ein besonderer Schwerpunkt liegt dabei auf der Frage, inwieweit die Marktentwicklung der Technologien interdependent ist.

Während die Grundlagen von Big Data und Cloud Computing weitgehend separat betrachtet werden, erfolgt die Analyse der Potenziale und Herausforderungen in einer gemeinsamen Betrachtung. In einem ersten Schritt werden die Potenziale anhand von zwei Dimensionen analysiert (Kap. V): Erwartete und bereits realisierte ökonomische Potenziale werden sowohl auf der Mikroebene von Unternehmen einerseits und staatlichen Stellen und Behörden andererseits (z. B. Kostenreduzierung, Produktivitätssteigerungen) als auch auf der gesamtgesellschaftlichen Makroebene (z. B. Wachstum, Arbeitsplätze) betrachtet. Da es sich hier

vor allem um quantifizierbare Größen handelt, wird diese Analyse um eine zweite, mehr auf die konkreten und möglichen Anwendungsbereiche abzielende Dimension ergänzt. Diese Dimension orientiert sich an den großen gesellschaftlichen Bedarfefeldern der Hightech-Strategie der deutschen Bundesregierung. Auf Grundlage von geplanten und teilweise schon realisierten zukünftigen Anwendungsmöglichkeiten der sich weiterentwickelnden Technologien werden zentrale Anwendungspotenziale aufgezeigt.

Die Diskussion der Anwendungspotenziale dient als Ausgangspunkt für die Betrachtung der zentralen Herausforderungen (Kap. VI). Technische, betriebliche, rechtliche, gesellschaftliche und strukturelle Herausforderungen werden eingehend erläutert und analysiert. Abschließend werden Handlungsbedarfe identifiziert und Optionen entsprechend der eingangs genannten Zielsetzung entwickelt.

Methodisch basiert der vorliegende Bericht auf einer umfassenden Literaturanalyse. Berücksichtigt wurden in erster Linie Marktforschungsergebnisse und wissenschaftliche Veröffentlichungen in Zeitschriften und Tagungsbänden, aber auch Beiträge in anwendungsorientierten Zeitschriften sowie weitere einschlägige Veröffentlichungen. Herausforderungen bei der Durchführung von Big-Data-Analysen sowie Auswirkungen der Nutzung von Cloudinfrastrukturen auf die Herausforderungen wurden mit Experten aus Unternehmen und der Wissenschaft sowie mit Vertretern von Verbänden im Rahmen von Interviews erörtert. An den Interviews nahmen Prof. Dr. Franziska Boehm (Institut für Informations-, Telekommunikations- und Medienrecht, Universität Münster), Dr. Stefan Fischer (SAP), Marit Hansen (Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein), Prof. Dr. Volker Markl (Institut für Softwaretechnik und Theoretische Informatik, TU Berlin), Prof. Michael Rotert (Verband der deutschen Internetwirtschaft e.V.), Dr. Michael Schermann (Lehrstuhl für Wirtschaftsinformatik, TU München) und Achim Schlosser (ParStream) teil. Zusätzlich werden im Rahmen der Analyse von Anwendungspotenzialen Miniszenarien eingesetzt, um mögliche zukünftige Entwicklungen zu antizipieren.

Abschließend sei noch darauf hingewiesen, dass für die Betrachtungen zu Cloud Computing zum Teil auf Ergebnisse aus einer vor Kurzem abgeschlossenen Studie für das Science and Technology Options Assessment Panel (STOA Panel) des Europäischen Parlaments zurückgegriffen wurde, die vom Fraunhofer-Institut für System- und Innovationsforschung ISI in Kooperation mit dem Institut für Technikfolgenabschätzung und Systemanalyse (ITAS) des Karlsruher Instituts für Technologie, dem Institut für Technikfolgen-Abschätzung (ITA) der Österreichischen Akademie der Wissenschaften und dem Danish Board of Technology (DBT) erstellt worden ist (Leimbach et al. 2014).

ÜBERBLICK ÜBER DIE FORSCHUNGSLAGE

3.

Sowohl im Bereich Big Data als auch im Bereich Cloud Computing wurde in den letzten Jahren intensiv geforscht. Die Kombination von Big Data und Cloud Computing bekam vergleichsweise wenig Aufmerksamkeit. Aus diesem Grund werden im vorliegenden Bericht Erkenntnisse aus Veröffentlichungen, die sich mit einem der Bereiche beschäftigen, zusammengeführt. Diskutiert wird diese Zusammenführung von Erkenntnissen vor dem Hintergrund von Veröffentlichungen, die sich ausdrücklich mit beiden Bereichen befassen, und von Experteninterviews, im Rahmen derer die Kombination von Big Data und Cloud Computing thematisiert wurde.

Im Hinblick auf Big Data fließen Ergebnisse aus folgenden Studien an verschiedenen Stellen in diesen Bericht ein:

- › Der Fokus der Studie »The Emerging Big Returns of Big Data«, die von TCS (2013) durchgeführt wurde, lag auf Großunternehmen. An der Studie nahmen insgesamt 1.217 Vertreter von Unternehmen teil. Davon wurden 643 Personen, die bereits Erfahrungen im Big-Data-Kontext gesammelt hatten, ausführlicher befragt. Befragt wurden überwiegend IT-Experten und Analysten, aber auch Mitarbeiter aus den Fachbereichen.
- › Bange et al. (2013) konzentrierten sich im Rahmen der für BARC durchgeführten Studie »Big Data Survey Europe« auf die Situation in Europa. Mittels Onlinebefragung wurden 274 Vertreter überwiegend großer Unternehmen befragt. Es handelte sich dabei sowohl um Mitarbeiter aus der IT als auch um Mitarbeiter aus den Fachbereichen.
- › Bei der von Interxion (2013) durchgeführten Studie »Big Data – Beyond the Hype« wurde der Fokus auf die Einschätzungen von IT-Verantwortlichen von Unternehmen mit über 500 Mitarbeitern gelegt. Die 750 Studienteilnehmer vertraten Unternehmen aus dem Vereinigten Königreich, Irland, Deutschland, Frankreich, Spanien, Österreich, Schweiz, Belgien, den Niederlanden, Dänemark und Schweden.
- › Im Rahmen der von Schäfer et al. (2012) für das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS durchgeführten Studie »Big Data – Vorsprung durch Wissen« wurden sowohl eine Onlineumfrage als auch ergänzende qualitative Expertenworkshops durchgeführt. An der Studie nahmen Fachbereichsverantwortliche von KMU teil.
- › Im Rahmen der von Schroeck et al. (2012) für IBM durchgeführten Studie »Analytics: Big Data in der Praxis« wurden 1.144 Fachbereichs- und IT-Verantwortliche befragt. Die Datenerhebung erfolgte mittels Onlinebefragung.
- › Russom (2011) führte für TDWI die Studie »Big Data Analytics« durch. Die Studie basiert auf den Einschätzungen von 325 Vertretern überwiegend großer Unternehmen, die im Rahmen einer Onlinebefragung erhoben wurden.



Ergänzend zur Onlinebefragung wurden auch Interviews durchgeführt. An der Studie nahmen überwiegend IT-Experten teil, aber auch Mitarbeiter aus den Fachbereichen und Berater.

Auch im Hinblick auf Cloud Computing fließen die Ergebnisse ausgewählter Studien besonders stark in den vorliegenden Bericht ein:

- › Für die von KPMG/BITKOM (2013) in Auftrag gegebene Studie »Cloud-Monitor 2013« wurden Telefoninterviews durchgeführt. Kontaktiert wurden IT-Verantwortliche und Mitglieder der Geschäftsführung von Unternehmen mit mehr als 20 Mitarbeitern über alle Branchen hinweg. Gearbeitet wurde mit einer geschichteten Zufallsstichprobe. Die öffentliche Verwaltung wurde im Rahmen der Studie nicht berücksichtigt.
- › Bradshaw et al. (2012) führten die Studie »Quantitative Estimates of the Demand for Cloud Computing in Europe and the Likely Barriers to Take-up« für IDC durch. Im Rahmen der Studie wurden Vertreter von 1.056 europäischen Unternehmen mittels Telefoninterviews befragt. Auch hier wurde mit einer geschichteten Zufallsstichprobe gearbeitet, um sicherzustellen, dass Unternehmen unterschiedlicher Größen und Branchen ausgewogen in der Stichprobe vertreten sind. Die öffentliche Verwaltung wurde im Rahmen der Studie berücksichtigt.
- › Das Cloud Industry Forum gab 2012 mehrere Studien zum Thema Cloud Computing mit unterschiedlicher regionaler Ausrichtung bei Vason Bourne in Auftrag. Auf die Studien »USA Cloud Adoption & Trends 2012« und »UK Cloud« wird im Rahmen dieses Berichts verwiesen. Während bei der Studie mit USA-Fokus 400 Fachbereichs- und IT-Verantwortliche aus Unternehmen und der öffentlichen Verwaltung befragt wurden (Cloud Industry Forum 2012b), wurden bei der Studie mit Fokus auf das Vereinigte Königreich 250 IT-Verantwortliche befragt (Cloud Industry Forum 2012a).

Alle angeführten Studien befassten sich mit Potenzialen und Herausforderungen im Kontext von Big Data bzw. Cloud Computing, setzten jedoch unterschiedliche Schwerpunkte, beispielsweise im Hinblick auf technische oder betriebliche Fragestellungen. Weitere Studien wie beispielsweise die Arbeiten von Mainzer (2014), Mayer-Schönberger/Cukier (2013) oder andere, die sich eher theoretisch oder anekdotisch mit den Potenzialen und Herausforderungen vor allem von Big Data beschäftigen, sowie Studien, auf deren Ergebnisse nur an einer oder an wenigen Stellen im Bericht verwiesen wird, werden direkt im Bericht genannt und gegebenenfalls kurz beschrieben.





GRUNDLAGEN

II.

BIG DATA

1.

Im Kontext wachsender Datenmengen stellt die Tatsache, dass zwischen 80 und 85 % der Daten unstrukturiert und damit schwer zu verwalten und analysieren sind, eine besondere Herausforderung dar (Behörden Spiegel 2013; TNS Infratest 2012, S.5). Im Jahr 2012 wurden nur 3 % der erzeugten Daten erfasst und nur 0,05 % analysiert. In diesem Zusammenhang wird von einem »Big Data Gap« gesprochen. Die wachsenden Datenmengen führen jedoch nicht nur zu mehr Datenmüll, sondern auch zu neuen Nutzungsmöglichkeiten, vorausgesetzt es stehen geeignete Analysemöglichkeiten zur Verfügung (Behörden Spiegel 2013).

In diesem Teilapitel wird zunächst diskutiert, was unter Big Data verstanden wird (Kap. II.1.1) und wie sich dieses Verständnis im Laufe der Zeit weiterentwickelt hat (Kap. II.1.2). Anschließend werden die technischen Grundlagen von Big Data erklärt (Kap. II.1.3) und Anwendungsperspektiven aufgezeigt (Kap. II.1.4).

BEGRIFFSDEFINITION

1.1

Der Begriff »Big Data« scheint auf den ersten Blick vor allem den Umfang von Daten in den Mittelpunkt zu stellen. Hierbei handelt es sich jedoch nicht um die einzige relevante Dimension. Beispielsweise verarbeiten Großbanken bereits seit vielen Jahren jede Nacht umfangreiche Mengen an Transaktionsdaten in Stapelverarbeitung auf Mainframes, ohne dass man dabei von Big Data sprechen würde. Die Transaktionsdaten sind zwar mengenmäßig sehr umfangreich, in ihrer Struktur aber homogen. Außerdem sind gewisse Verzögerungen bei ihrer Verarbeitung akzeptabel. Neben der Menge an Daten (Volume) charakterisieren vor allem die Geschwindigkeit ihrer Erzeugung, die hohen Anforderungen an die Geschwindigkeit ihrer Analyse (Velocity) sowie die Heterogenität der Daten (Variety) Big Data:

- › Menge: Unternehmen und Einrichtungen der öffentlichen Verwaltung verfügen über immer größere Datenmengen. Häufig geht es dabei um Größenordnungen von einigen Terabytes bis hin zu Petabytes (BITKOM 2012, S. 21). Eine IDC-Studie zum Datenwachstum deutet an, dass sich das weltweite Datenvolumen alle zwei Jahre verdoppelt (Gantz/Reinsel 2012). Diese Entwicklung wird vor allem durch die zunehmende maschinelle Erzeugung von Daten, z. B. über Protokolle von Telekommunikationsverbindungen und Webzugriffen sowie automatische Erfassungen von Radio-frequency-Identification-Lesern (RFID-Lesern), Kameras, Mikrofonen und sonstigen Sensoren, getrieben.

- › Geschwindigkeit: Die maschinelle Erzeugung von Daten trägt nicht nur zu einer sehr großen Datenmenge bei, sondern dies geschieht auch in immer kürzeren Zeiträumen (Russom 2011, S. 7). Aufgrund wachsender Datenmengen, einer engen Verknüpfung der Prozesse mit IT-Systemen und steigender Komplexität sind die Logistik und die Pflege von Kundenbeziehungen Beispiele für Bereiche, in denen Big Data besonders relevant ist. Auch in der Wissenschaft werden Daten mit sehr hoher Geschwindigkeit erzeugt. Man denke nur an den Large Hadron Collider (LHC), einen Teilchenbeschleuniger am Europäischen Kernforschungszentrum CERN bei Genf, bei dem nur 0,001 % der Sensordaten erfasst, aber trotzdem jährlich Daten im Umfang von rund 25 Petabyte analysiert werden. Eine hohe Geschwindigkeit im Hinblick auf die Analyse vorhandener Daten ist in vielen Bereichen kritisch. Nicht selten ist gar eine Auswertung in Echtzeit erforderlich (BITKOM 2012, S. 21).
- › Heterogenität: Die Vielfalt von Datenquellen und Datenformaten nimmt kontinuierlich zu. Organisationsinterne Daten werden zunehmend durch externe Daten ergänzt. Neben strukturierten Daten (z. B. Kennzahlen, Sensordaten) spielen auch semistrukturierte (z. B. XML, HTML, JSON) und unstrukturierte Daten (z. B. Texte, Bilder) eine immer größere Rolle (BITKOM 2012, S. 21). Im Zusammenhang mit Big Data ist daher häufig von der Analyse polystrukturierter Daten die Rede. Das Nutzbarmachen von Daten unterschiedlicher Strukturen ist eine Kernherausforderung im Kontext von Big Data.

Die Dimensionen Menge, Geschwindigkeit und Heterogenität gehen auf einen Beitrag von Laney (2001) über die Grenzen des klassischen Datenmanagements zurück und werden im Rahmen von Definitionen von Big Data gerne aufgegriffen. Einige Begriffsdefinitionen beschreiben allerdings noch weitere Dimensionen, wie z.B. die Richtigkeit der Daten (Veracity) (Schroeck et al. 2012, S.4), den Wert der Datenauswertung (Value) (Tibco 2012) oder die Anzahl der Freiheitsgrade bei der Interpretation der Daten (Variability) (Dumbill 2012).

Dass diese Dimensionen nicht nur in der Theorie, sondern auch in der Praxis mit Big Data in Verbindung gebracht werden, bestätigt eine Studie von IBM (Schroeck et al. 2012). Im Rahmen der Studie wurden 1.144 Fachleute aus 93 Staaten zum Thema Big Data befragt. Unter den sechs am häufigsten genannten Assoziationen mit Big Data waren nicht nur »größere Bandbreite an Informationen«, »größere Datenmengen« und »neue Arten von Daten und Analysen«, sondern auch »Echtzeitinformationen«, »Zustrom von Daten aus neuen Technologien« und »moderne Medienarten« (Schroeck et al. 2012, S.3).

Einer Studie der Experton Group zufolge wird Big Data überwiegend als neues IT-Paradigma verstanden, das durch immenses Datenwachstum hervorgerufen wurde (Velten/Janata 2012b, S.9). Auch die Ansicht, dass Big Data eine neue Generation von Datenbank- und Analytiktechnologien beschreibt, wird von relativ vielen Studienteilnehmern geteilt. Entsprechenden Aussagen stimmten 49 %



bzw. 41 % ohne Einschränkungen zu. Auffällig ist auch die Einschätzung, dass es sich bei Big Data nur um einen neuen Marketinghype handelt. Mit 32 % stimmte immerhin knapp ein Drittel der Befragten einer entsprechenden Aussage uneingeschränkt zu. An der Studie nahmen 100 Vertreter von deutschen Unternehmen mit mehr als 500 Mitarbeitern teil. Überwiegend wurden IT-Verantwortliche sowie Mitglieder der Geschäftsführung befragt.

Über die Kriterien, denen zufolge Daten zu Big Data werden, bestand lange Uneinigkeit. Mittlerweile hat sich die Ansicht durchgesetzt, dass es sich dann um Big Data handelt, wenn die Daten mit Standardtechnologien nicht mehr effizient verarbeitet werden können (BITKOM 2012, S.21; EITO 2012a, S.14; Jacobs 2009, S.11). Was Big Data ist und was nicht, ändert sich daher mit der Weiterentwicklung der verfügbaren Technologien. Hilbert/Lopez (2011) zufolge wuchs die Leistung von handelsüblichen Rechnern zwischen 1986 und 2007 um durchschnittlich 58 % pro Jahr. Der analysierte Datenbestand des US-amerikanischen Handelsriesen Walmart galt mit einem Umfang von rund einem Terabyte im Jahr 1992 als unglaublich groß (Harris 2013). Heute verarbeiten viele Organisationen Daten im Umfang von mehreren Petabyte. Laut einer IDC-Studie wächst der auf der Welt verfügbare Speicherplatz langsamer als das Datenaufkommen (The Economist 2010c).

Bei der Durchführung von Analysen großer Datenmengen, die häufig allgemein als Big-Data-Analytik bezeichnet wird, geht es vor allem um die Erkennung von Zusammenhängen, Bedeutungen und Mustern (BITKOM 2012, S.21). Zum Einsatz kommen neben statistischen Verfahren Vorhersagemodelle, Optimierungsalgorithmen, Data Mining sowie Text- und Bildanalytik.

Auffällig ist, dass zahlreiche Definitionen von Big Data und Big-Data-Analytik grundsätzlich von einer Unterstützung von wettbewerbskritischen Entscheidungen in einem betriebswirtschaftlichen Kontext ausgehen. Auf die Tatsache, dass Big-Data-Analytik z.B. auch in der Wissenschaft oder in der öffentlichen Verwaltung eine bedeutende Rolle spielt, wird nur selten explizit eingegangen. Auch wenn Big-Data-Analytik eine Weiterentwicklung der Methoden zur Datenanalyse im Allgemeinen darstellt, scheint doch das geschäftliche Umfeld und die Nähe zum etablierten Thema Business Intelligence (BI) vor allem für Anbieter im Bereich Big Data zentral zu sein. Eine präzisere Abgrenzung von Big-Data-Analytik und BI erfolgt im Kapitel II.1.2. Ein Überblick über Anbieter findet sich im Kapitel III.1.2.

ENTWICKLUNG UND ABGRENZUNG

1.2

Aus der Definition von Big Data geht klar hervor, dass Big-Data-Analytik nicht etwas gänzlich Neues ist, sondern vielmehr eine Weiterentwicklung von existie-



renden Ansätzen zur Analyse von Daten. Die sich ändernden Anforderungen im Hinblick auf die zu analysierende Datenmenge, die Analysegeschwindigkeit und die Heterogenität der zu analysierenden Daten führten zu Anpassungen und Neuentwicklungen bei den zur Datenanalyse verwendeten Ansätzen und Technologien. Daten wurden im Laufe der Zeit von einem Nebenprodukt immer mehr zu einem zentralen Faktor (TNS Infratest 2012, S. 5).

Press (2013a) stellte für Forbes.com einen historischen Überblick über den zunehmenden Umfang an Daten und den Umgang damit zusammen. Die Darstellung beginnt mit einer in Buchform veröffentlichten Einschätzung aus dem Jahr 1944, dass sich der Bestand der US-amerikanischen Universitätsbibliotheken alle 16 Jahre verdoppeln wird (Rider 1944) und endet mit dem Hinweis auf einen 2012 erschienenen Artikel, in dem Big Data als ein Phänomen bezeichnet wird, das auf dem Zusammenspiel von leistungsfähiger IKT, der Analyse großer Datenbestände und der Annahme, dass umfangreichere Daten zu besseren Ergebnissen führen, beruht (Boyd/Crawford 2012). Der Überblick zeigt nicht nur, dass große Datenmengen bereits seit vielen Jahren ein Thema sind, sondern auch, dass das Wachstum des Datenbestands in der Vergangenheit eher unter- als überschätzt wurde.

Big-Data-Analytik hat ihren Ursprung im Bereich der Entscheidungsunterstützungssysteme (Watson/Wixom 2007, S. 96) und wird als Weiterentwicklung von BI und Business Analytics gesehen. Während der Begriff »Business Intelligence« in den späten 1980er Jahren entstand und in den 1990er Jahren schon weit verbreitet war, gewann der Begriff »Business Analytics« erst in den späten 2000er Jahren an Bedeutung (Chen et al. 2012, S. 1166; Davenport 2006). Es existieren verschiedene, zum Teil widersprüchliche Versuche zur Abgrenzung von BI und Business Analytics. Sowohl unter BI als auch unter Business Analytics wird die Unterstützung von strategischen und operativen Entscheidungen auf Grundlage der systematischen Analyse von Daten verstanden.

Russom (2011, S. 9) untersuchte im Rahmen einer Studie für TDWI unter anderem, welche Begriffe in der Praxis für die Analyse großer Datenmengen verwendet werden. Im Rahmen der Studie wurden die Antworten von 92 Fachleuten für Datenmanagement, die überwiegend im nordamerikanischen Raum tätig sind, analysiert. Neben dem Begriff »Big Data Analytics« wurden die Begriffe »Advanced Analytics« und »Analytics« besonders häufig genannt, aber auch Begriffe wie »Data Warehousing«, »Data Mining« oder »Predictive Analytics«.

Big-Data-Analytik lässt sich von verwandten Themenfeldern wie BI, Advanced Analytics und Predictive Analytics, Data Warehousing und Data Mining unterschiedlich gut abgrenzen.

- › Business Intelligence beschreibt Verfahren und Prozesse zur systematischen Analyse von Daten in elektronischer Form. Ziel ist die Gewinnung von Er-



kenntnissen zur Unterstützung von Entscheidungen. Der Begriff wird dem Fachgebiet der Wirtschaftsinformatik zugerechnet und hat eine klare betriebswirtschaftliche Ausrichtung. Unternehmen sollen mit den gewonnenen Erkenntnissen ihre Geschäftsabläufe optimieren, Kosten senken, Risiken minimieren und die Wertschöpfung vergrößern. Data-Warehouse-Datenbanken sowie Analysemethoden wie Online Analytical Processing (OLAP), Data Mining und fallbasiertes Schließen (Case-Based Reasoning) stellen die Basis für BI dar.

- › Advanced Analytics und Predictive Analytics werden als in die Zukunft gerichtete Erweiterungen von BI verstanden, bei denen Prognosen der zukünftigen Entwicklung stärker im Vordergrund stehen als die Analyse der Istsituation und der Vergangenheit. Genauso wie bei BI handelt es sich bei Advanced Analytics und Predictive Analytics nicht um Technologien, sondern vielmehr um den kombinierten Einsatz von verschiedenen Werkzeugen zur Sammlung, Auswertung und Interpretation von Daten (Bose 2009, S. 156).
- › Data Warehousing oder der Betrieb einer Data-Warehouse-Datenbank erhöht den Komfort beim Zugang zu Daten aus unterschiedlichen Quellen, die in einem einheitlichen Format zusammengefasst wurden. Außerdem soll damit verhindert werden, dass die Leistungsfähigkeit von transaktionsorientierten Datenbanken beeinträchtigt wird (Inmon 2005). Der Prozess der Vereinigung von Daten mehrerer Quellen in einer Datenbank, der bei Data Warehouses eine wichtige Rolle spielt, wird als Extraktion-Transformation-Laden-Prozess (ETL-Prozess) bezeichnet. Anwendungen arbeiten in der Regel mit anwendungsspezifisch erstellten Auszügen aus einem Data Warehouse, den sogenannten Data Marts. Big-Data-Anwendungen nutzen zur Datenhaltung Ansätze, die sich von klassischen Data Warehouses unterscheiden.
- › Data Mining beschreibt die systematische Anwendung von statistischen Methoden auf einen Datenbestand mit dem Ziel, Muster zu erkennen. Data Mining bezeichnet »die Extraktion von Wissen aus Datenbanken, das gültig (im statistischen Sinne), bisher unbekannt und prinzipiell nützlich (für eine gegebene Anwendung) ist« (Ester/Sander 2000, S. 1). Data Mining ist ein Schritt des Knowledge-Discovery-in-Databases-Prozesses (KDD-Prozesses). Text Mining, Web Mining und Zeitreihenanalysen sind Spezialisierungen von Data Mining auf spezielle Typen von Daten. Auch wenn sich die Verfahren in der Regel von denen, die im Data-Warehouse-Umfeld eingesetzt werden, unterscheiden, kommen auch im Big-Data-Kontext statistische Methoden zur Erkennung von Mustern und Trends zum Einsatz.

Da im Kontext der Untersuchung von großen Datenmengen auch die Visualisierung eine bedeutende Rolle spielt, erscheint auch eine Abgrenzung zum eigenständigen, aber komplementären Feld Visual Analytics notwendig. Ziel von Visual Analytics ist es, Erkenntnisse aus extrem großen und komplexen Datensätzen zu gewinnen. Der Ansatz kombiniert die Stärken der automatischen Datenanalyse mit den Fähigkeiten des Menschen, schnell Muster und Trends visuell



zu erfassen. Dem Menschen werden nicht nur Resultate präsentiert, sondern es wird ihm die Möglichkeit gegeben, in die Analyse einzugreifen und die Algorithmen zu beeinflussen (Thomas/Cook 2005 u. 2006). Zu den Anwendungsgebieten zählen neben BI (SAS 2013; Tilton 2011) auch Bereiche wie Physik, Astronomie, Biologie, Medizin und der Katastrophenschutz (Keim et al. 2010; Thomas/Cook 2005). Software muss zur optimalen Unterstützung von Visual Analytics bestimmte Anforderungen erfüllen. Die Herausforderung besteht darin, eine intuitive Bedienung, die ein zügiges Arbeiten mit unterschiedlichsten Daten ermöglicht, mit umfassenden Analysefunktionen zu kombinieren (Few 2012; SAS 2013).

Laut Ortega (2013) sprechen große Datenmengen, Feinkörnigkeit und kurze Latenzzeiten dafür, Big-Data-Lösungen gegenüber BI-Lösungen auf Basis eines Data Warehouse den Vorzug zu geben. Abgesehen davon lassen sich viele Anwendungsfälle, die auf den ersten Blick Big-Data-Analytik zu erfordern scheinen, durchaus auch mit einem Data Warehouse realisieren.

TECHNISCHE GRUNDLAGEN

1.3

Im Hinblick auf die Diskussion der technischen Grundlagen von Big Data erscheint es sinnvoll, die Aspekte Quelldaten, Integration, Speicherung, Analytik, Zugriff und Auswertung (Kap. II.1.3.1 bis II.1.3.6) getrennt voneinander zu betrachten und jeweils die Ansätze der klassischen BI mit jenen aus dem Big-Data-Umfeld zu vergleichen (Singh 2011) (Tab. II.1).

TAB. II.1 TECHNISCHE GRUNDLAGEN VON BUSINESS INTELLIGENCE (BI) UND BIG DATA

	Big-Data-Analyse	traditionelle BI
Auswertung	BI-Anwenderwerkzeuge analytische Anwendungen	
Zugriff	Pig, Hive, API	SQL, MDX
Analytik	MapReduce, R	Data Mining, OLAP
Speicherung	HDFS, NoSQL	analytische Datenbanken
Datenintegration	Pig, Sqoop, ETL/ELT	ETL/ELT
Datenquellen	polystrukturierte Daten	strukturierte Daten

Quelle: nach Bange 2012

Neben Performanz und dem Umgang mit heterogenen Daten spielen auch Sicherheit, einfache Handhabung und Offenheit für die Integration neuer Datenquellen, Analysetools und Funktionen eine wichtige Rolle (Küll o.J.). Die Analysearchitektur für BI unterscheidet sich von jener für Big Data.

QUELLEDATEN**1.3.1**

Einen wesentlichen Unterschied zwischen klassischer BI und Big-Data-Analytik stellt die Tatsache dar, dass BI primär auf strukturierten Daten beruht, während im Kontext von Big Data in der Regel polystrukturierte Daten verarbeitet werden. Darüber hinaus werden im Rahmen von Big-Data-Analysen neben internen verstärkt auch externe Daten verwendet.

Im Hinblick auf externe Daten wird nicht nur in Social-Media-Daten und Daten, die im Rahmen von Newsfeeds bereitgestellt werden, besonderes Potenzial gesehen, sondern auch in Open Data. Bei Open Data handelt es sich vor allem um Daten aus der öffentlichen Verwaltung (auch Open Government Data), die weitgehend frei von Schutzrechten in maschinenlesbarer Form zur Verfügung gestellt werden. Das zentrale Datenportal für Open Government Data in Deutschland »GovData«¹ wird durch Fraunhofer FOKUS im Rahmen eines Forschungsprojekts für das Bundesministerium des Inneren betrieben und weiterentwickelt. Vergleichbare Plattformen werden von zahlreichen Regierungen angeboten. Welche Daten derzeit bei Big-Data-Analysen in welchem Umfang genutzt werden, wird im Kapitel IV.1.1.3 erläutert.

INTEGRATION**1.3.2**

Obwohl es auch vorkommt, dass ETL-Werkzeuge neuentwickelt werden, werden für die Integration von Daten überwiegend bestehende Standardwerkzeuge verwendet. Gründe für den Einsatz von Standardwerkzeugen sind beispielsweise der einfache Zugriff auf gängige Datenbank-, Enterprise-Resource-Planning-Systeme (ERP-Systeme) und Dateisysteme sowie die durch Visualisierungen unterstützte Entwicklung und Wartung von ETL-Prozessen. Immer mehr ETL-Werkzeuge bieten auch Konnektoren zu Hadoop und NoSQL-Datenbanken. Sowohl Hadoop als auch NoSQL-Datenbanken kommen vor allem im Kontext von Big-Data-Lösungen zum Einsatz und werden im Kapitel II.1.3.3 genauer beschrieben. Zur Integration von Big Data wird auch die ursprünglich von Yahoo! entwickelte Skriptsprache Pig verwendet. Ein populäres Werkzeug für den raschen Import von großen Datenmengen in Hadoop ist Sqoop.

SPEICHERUNG**1.3.3**

Die Speicherung von Big Data erfolgt häufig in Hadoop. Hadoop ist ein Open-Source-Java-Framework, das auf dem Dateisystem Hadoop Distributed File System (HDFS) basiert. Bei Hadoop handelt es sich um eine massiv parallele Archi-

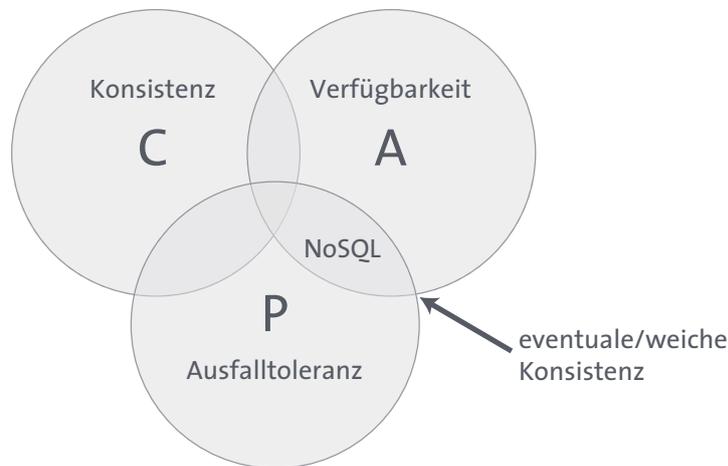
1 www.govdata.de

tektur, die Daten automatisch verteilt und redundant ablegt. Neben Hadoop kommen zur Speicherung von Big Data auch NoSQL-Datenbanken zum Einsatz.

»NoSQL« steht heute für eine Vielzahl von unterschiedlichen Datenbankarchitekturen, die teilweise auf SQL²-Strukturen zurückgehen und wird daher als »not only SQL« verstanden (North 2010). Der Begriff, für den es bis heute keine eindeutige Definition gibt, geht auf die späten 1990er Jahre zurück, als erste Datenbanken im Umfeld von Webanwendungen entstanden, die nicht dem AKID-Prinzip klassischer SQL-Datenbanken folgten. »AKID« steht dabei für Atomarität, Konsistenz, Isolation und Dauerhaftigkeit und beschreibt Eigenschaften, die als Voraussetzung für die Verlässlichkeit eines Systems gelten. Laut CAP-Theorem kann ein verteiltes System nur zwei der drei Eigenschaften Konsistenz (Consistency), Verfügbarkeit (Availability) und Ausfalltoleranz (Partition tolerance) gleichzeitig erfüllen (Brewer 2000; Gilbert/Lynch 2002). Konsistenz wird bei vielen NoSQL-Datenbanken, die auf mehrere Systeme aufgeteilt sein können, zugunsten von Verfügbarkeit und Ausfalltoleranz geopfert. Man spricht in diesem Zusammenhang auch vom BASE-Prinzip (Basically Available, Soft state, Eventual consistency) (Abb. II.1).

ABB. II.1

VERANSCHAULICHUNG DES CAP-THEOREMS



Quelle: nach Brewer 2000

NoSQL-Datenbanken haben keine festen Tabellen und unterstützen folglich auch keine Abfragen über mehrere Tabellen (SQL-Joins). Alternativen zu relationalen Datenbanken wurden vor allem deshalb entwickelt, weil mit relationalen Datenbanken bestimmte Arten von Daten wie rekursive Daten, Grafen, Zeitreihen oder Genomdaten nicht zufriedenstellend modelliert werden konnten.

2 SQL wird im allgemeinen Sprachgebrauch als Abkürzung für Structured Query Language aufgefasst, obwohl sie laut Standard ein eigenständiger Name ist.



Stonebraker et al. (2007) betonen, dass durch spezialisierte Datenbanken in vielen Bereichen deutliche Verbesserungen im Hinblick auf die Performanz möglich sind. Während relationale Datenbanken für die Datenverarbeitung sehr gut geeignet sind, sind sie für den Einsatz im Rahmen von Data Warehouses, die Internetsuche, Untersuchungen in Echtzeit sowie für die Untersuchung von teil- und unstrukturierten Daten nicht ideal. Während sich Programmiersprachen im Laufe der Jahre weiterentwickelt haben, war das bei SQL nicht der Fall. Nichtsdestotrotz muss im Einzelfall entschieden werden, ob eine relationale Datenbank oder eine nichtrelationale Datenbank die jeweiligen Anforderungen besser erfüllt (Olofson 2012).

Alternative Datenbankstrukturen sind vor allem für die performante Indizierung großer Datenmengen, eine Schwäche relationaler Datenbanken, sowie für große, skalierbare, horizontale Anwendungen geeignet. Da unter NoSQL-Datenbanken ganz allgemein Alternativen zu SQL-Datenbanken verstanden werden, subsumieren sie eine Vielzahl von technischen Lösungen, die teilweise auch untereinander konkurrieren. Walker-Morgan (2010) beschreibt vier Hauptkategorien von NoSQL-Datenbanken, die auch als NoSQL-Kernsysteme bezeichnet werden:

- › Dokumentenorientierte Datenbanken: Diese speichern textuelle Daten von beliebigem Umfang mit unstrukturierten Informationen, die das Suchen auf Basis von Dokumentinhalten ermöglichen. Zu bekannten Vertretern dieser Kategorie gehören CouchDB und MongoDB.
- › Key-Value-Datenbanken: Die Speicherung von Daten erfolgt hier auf Basis von Schlüsseln, die auf bestimmte Werte verweisen. Key-Value-Datenbanken wie dbm, gdbm und Berkley DB sind seit jeher Bestandteil von diversen Anwendungen, vor allem in der Unix-Welt.
- › Spaltenorientierte Datenbanken: Daten werden so gespeichert, dass bestimmte Operationen besonders schnell durchgeführt werden können. Sie werden daher gerne für Data-Mining- und Analyseanwendungen eingesetzt. Bei Cassandra und HBase handelt es sich um bekannte Vertreter dieser Kategorie.
- › Grafendatenbanken: Der Fokus wird auf die Speicherung von Daten in Form von Knotenpunkten und Beziehungen gelegt. Informationen über Knoten und Beziehungen werden als Eigenschaften gespeichert. Neo4j ist eine der bekannteren Grafendatenbanken.

Neben den NoSQL-Kernsystemen gibt es auch Soft-NoSQL-Systeme, zu denen objektorientierte Datenbanken, XML-Datenbanken, Griddatenbanken und andere nichtrelationale Datenbanken zählen (Pürner 2013). Das NoSQL-Archiv³ enthält Angaben zu über 150 nichtrelationalen Datenbanken.

Während Datenbankmanagementsysteme in der Vergangenheit in der Regel Festplattenlaufwerke zur Speicherung von Daten nutzten (On-Disk-Datenban-

3 <http://nosql-database.org>



ken), nahm das Interesse an In-Memory-Datenbanken (auch Hauptspeicherdatenbanken oder Hauptspeicherresistente Datenbanken) aufgrund höherer Performance, gesunkener Hardwarepreise und erweiterter Adressierungsmöglichkeiten der 64-Bit-Technologie zuletzt stark zu (Pürner 2013). In-Memory-Datenbanken sind Datenbankmanagementsysteme, die den Arbeitsspeicher als Datenspeicher nutzen. Neben den reinen In-Memory-Datenbanken, die ohne Festplatten im Hintergrund auskommen, gibt es auch Datenbankmanagementsysteme, die den Arbeitsspeicher nur zum Caching verwenden. Sowohl alternative als auch relationale Datenbanken können als In-Memory-Datenbanken betrieben werden. Ein bekanntes Beispiel für eine In-Memory-Datenbank ist SAP HANA.

Im BI-Kontext dienen primär analytische, aber auch relationale Datenbanken als Grundlage für das Data Warehousing. Analytische Datenbanken sind speziell für analytische Aufgaben optimiert und bauen in der Regel auf einer massiv parallelen Architektur auf (Wolfgang Martin Team & IT Research 2013). Analytische Datenbanken kommen häufig in Form von Appliances zum Einsatz.

ANALYTIK

1.3.4

Im Hinblick auf Big-Data-Analytik ist MapReduce ein zentrales, ursprünglich von Google eingeführtes Programmiermodell, das die Zerlegung von Aufgaben in kleine Teile und deren parallele Ausführung auf Clustern erlaubt. MapReduce wurde speziell für Berechnungen im Kontext von umfangreichen Daten entwickelt, die über mehrere Cluster verteilt sind. Der Ablauf des MapReduce-Algorithmus lässt sich grob in die Phasen Map und Reduce unterteilen:

- › Map-Phase: Daten werden auf Prozesse verteilt, die parallel ausgeführt werden. Zwischenergebnisse werden in verschiedenen Zwischenergebnisspeichern abgelegt.
- › Reduce-Phase: Für jeden Satz an Zwischenergebnissen werden von genau einem Reduce-Prozess die Ausgabedaten berechnet.

Für das Data Mining kommt im Kontext von Big Data neben der Hadoop-eigenen Bibliothek Mahout auch die populäre Open-Source-Bibliothek R zum Einsatz. Während MapReduce-Implementierungen wie Hadoop für einfachere Beispiele wie das Zählen von oder die Suche nach bestimmten Wörtern sehr gut geeignet sind, stoßen sie bei komplexeren Aufgaben schnell an ihre Grenzen. Neben einer massiv parallelen Ausführung ist beispielsweise für Netzwerkanalysen, Analysen von Videomaterial oder die Beantwortung von Fragen, die in natürlicher Sprache formuliert wurden, auch die Verfügbarkeit der notwendigen Analysewerkzeuge essenziell (Rajan 2012).

UIMA (Unstructured Information Management Architecture) ist eine frei verfügbare, plattformunabhängige Architektur und Softwareumgebung zur Verwal-



tung von unstrukturierten Daten, mit deren Hilfe verschiedene Verarbeitungskomponenten und Suchtechnologien miteinander verbunden werden können (Ferrucci/Lally 1999). Ursprünglich wurde UIMA von IBM entwickelt, ist aber seit 2006 als Open-Source-Projekt bei der Apache Software Foundation angesiedelt. Das Ziel von UIMA ist es, die Zusammenarbeit bei der Entwicklung von Technologien für Knowledge Discovery voranzutreiben und zu erleichtern. UIMA bietet die Möglichkeit, eine Pipeline mit Analyse- und Annotationswerkzeugen aufzubauen, die ursprünglich nicht dafür entwickelt wurden, miteinander zu interagieren. Grundlage hierfür ist, dass alle Komponenten der Pipeline auf eine gemeinsame Datenstruktur zugreifen, die sogenannte Common Analysis Structure (CAS). Die Komponenten können in Java oder C++ geschrieben sein und müssen eine Schnittstelle bereitstellen, die durch das UIMA-Framework definiert ist und ebenso wie die obligatorischen, selbstbeschreibenden Metainformationen mithilfe einer XML-Descriptor-Datei implementiert wird. Jede Komponente besteht also aus einem deklarativen Teil und dem Programmcode, wobei der deklarative Teil zum einen die Metadaten enthält, wie Namen der Komponente, Autor und Version, zum anderen eine Beschreibung des Typensystems. Im Programmcode wird der eigentliche Algorithmus implementiert.

Die Datenanalyse auf Basis einer UIMA-Pipeline lässt sich in drei wesentliche Phasen unterteilen:

- › Der Collection Reader liest die Daten ein und erstellt für jedes Inputdokument ein CAS-Objekt.
- › Die Analysis Engines und Annotatoren analysieren und annotieren die Daten.
- › Der CAS Consumer führt die abschließende Verarbeitung des CAS-Objekts durch (Extraktion, Visualisierung, Aufbau eines Indexes, Evaluierung mit Goldstandard).

In Bereich der klassischen BI werden neben freien Datenrecherchen vor allem modellgestützte Analysen und Ad-hoc-Analysen durchgeführt. Während für Ad-hoc-Analysen in der Regel OLAP verwendet wird, spielt Data Mining bei modellgestützten Analysen eine zentrale Rolle. Die Suche nach versteckter Information mithilfe von Data Mining spielt auch im Big-Data-Kontext eine wichtige Rolle. Typische Aufgabenstellungen sind Ausreißerererkennung, Clusteranalyse, Klassifikation, Assoziationsanalyse, Regressionsanalyse und Zusammenfassung (OnRamps o. J.).

ZUGRIFF

1.3.5

Zum Zugriff auf die Analytik kommt unter anderem die Skriptsprache Pig zum Einsatz. Beispielsweise lassen sich mithilfe von Pig MapReduce-Prozesse erzeugen und auf einem Hadoop-Cluster ausführen. Bei HiveQL handelt es sich um eine Abfragesprache für Hadoop, die ähnlich aufgebaut ist wie SQL. Pig und



HiveQL sind sehr flexibel einsetzbar, da sich zusätzliche Funktionen in Java und zum Teil auch in anderen Programmiersprachen formulieren lassen. Viele Anwendungen zur Auswertung von Daten, die auf Hadoop zugreifen, machen das mittels Hive. Anwendungen können oft auch direkt über eine Programmierschnittstelle (Application Programming Interface, API) auf die Analytik zugreifen. In BI-Kontext kommen beispielsweise SQL und MDX (Multidimensional Expressions) zum Einsatz. MDX ist eine von Microsoft entwickelte Datenbanksprache für OLAP-Datenbanken.

AUSWERTUNG UND VERARBEITUNG

1.3.6

Die Ergebnisse von Big-Data-Analysen werden entweder mithilfe von analytischen Anwendungen ausgewertet (Entscheidungsunterstützung) oder direkt weiterverarbeitet (Entscheidungsautomatisierung). Klassische Anwendungen zur Auswertung von Daten im BI-Kontext umfassen Dashboards sowie Reporting-, Analyse-, und Planungswerkzeuge. Im Big-Data-Umfeld kommen noch spezielle Anwendungen für Suche, Textanalyse, Netzwerkanalyse, Pfadanalyse und Mustererkennung dazu. Wie bereits im Kapitel II.1.2 angedeutet, können auch Visualisierungen bei der Auswertung umfangreicher Datenmengen eine zentrale Rolle spielen (SAS 2013). Sowohl bei der Verwendung von analytischen Anwendungen als auch bei der Exploration von Daten mithilfe von Visual-Analytics-Lösungen ist bei der Entscheidungsunterstützung die Berücksichtigung der Bedürfnisse der jeweiligen Nutzergruppe zentral. Complex Event Processing (CEP) spielt eine wichtige Rolle, wenn Analyseergebnisse kontinuierlich und zeitnah weiterverarbeitet werden müssen (Eckert/Bry 2009). Eine Entscheidungsautomatisierung auf Basis von Big-Data-Analytik und CEP ist beispielsweise im Zusammenhang mit Sensornetzwerken, die zur Überwachung industrieller Anlagen eingesetzt werden, denkbar. Bei der frühzeitigen Erkennung von Trends mithilfe von Marktdaten und der automatischen Reaktion darauf handelt es sich um ein weiteres Anwendungsgebiet von Big-Data-Analytik und CEP.

EINSATZMÖGLICHKEITEN

1.4

Sowohl für Unternehmen im Dienstleistungsbereich und in der Industrie (Schäfer et al. 2012) als auch für die öffentliche Verwaltung (Kommune21 2013) und die Wissenschaft gibt es zahlreiche Anwendungsmöglichkeiten von Big-Data-Lösungen (The Economist 2010c). Aber auch Privatpersonen profitieren von den Erkenntnissen aus der Untersuchung großer Datenmengen. Beispielsweise kann die rasche Untersuchung von Daten aus unterschiedlichen Quellen nicht nur die Zuverlässigkeit von Wettervorhersagen verbessern, sondern auch dabei helfen, Verkehrsstaus zu vermeiden. Auch computergesteuerte Autos sind keine Utopie mehr. Sebastian Thrun von der Universität Stanford nimmt an, dass in zehn Jah-



ren die Technik so weit sein wird, dass autonome Autos definitiv zuverlässiger und damit sicherer fahren als der Mensch (Glockner/Beyers 2012, S.33). Die Echtzeitanalyse großer Datenmengen aus verschiedenen Quellen spielt dabei eine entscheidende Rolle.

Für Unternehmen bietet die Untersuchung großer Datenmengen die Möglichkeit, Wettbewerbsvorteile zu erzielen, Einsparungen und Umsatzsteigerungen zu realisieren und innovative, neue Geschäftsfelder zu schaffen (Schäfer et al. 2012). Es wird angenommen, dass Big-Data-Lösungen in der Privatwirtschaft zukünftig insbesondere in den Bereichen Controlling, Logistik, Risikomanagement und Wettbewerbsanalyse eine Rolle spielen werden (Kommune21 2013). Welche Anwendungsperspektiven in Unternehmen am vielversprechendsten sind, unterscheidet sich von Branche zu Branche (Schäfer et al. 2012). Durch die Analyse bereits vorhandener Datenbestände können z.B. Banken und Versicherungen finanzielle Risiken besser einschätzen und Betrugsversuche schneller erkennen. Im Handel können Big-Data-Lösungen beispielsweise zur Verbesserung personalisierter Produktempfehlungen beitragen. Es wird angenommen, dass auf Grundlage von Erkenntnissen aus der Analyse großer Datenmengen ausgewählte Zielgruppen mit relevanten Inhalten über den richtigen Kanal und zur richtigen Zeit adressiert werden können (Helios 2013, S. 15).

Auch für die öffentliche Verwaltung stellt das Management großer Datenmengen eine neue Herausforderung dar (Kommune21 2013). Daten über Telekommunikationsverbindungen, Georeferenzierungsdaten, Energieverbrauchsdaten aus smarten Messgeräten, Daten aus dem Finanzbereich oder aus dem Gesundheitswesen, aber auch Daten aus sozialen Medien werden künftig eine bedeutende Rolle in der öffentlichen Verwaltung spielen. Die Nutzung von Big-Data-Lösungen erscheint beispielsweise beim Zoll und bei der Steueranalyse interessant. In diesen Bereichen werden bereits jetzt große Datenmengen analysiert. Allerdings stellt die Untersuchung der vorhandenen Daten in Echtzeit häufig noch ein Problem dar (Behörden Spiegel 2013). Ein weiteres, vielversprechendes Anwendungsgebiet von Big-Data-Lösungen im Umfeld der öffentlichen Verwaltung liegt in der Koordination von Polizeieinheiten, beispielsweise bei Großereignissen. Dabei müssen Daten aus der Überwachung, dem Mobilfunk und auch aus sozialen Medien in Echtzeit untersucht werden.

Wie nützlich die Untersuchung großer Datenmengen in der Wissenschaft sein kann, zeigen Beispiele wie der bereits erwähnte LHC am CERN eindrucklich. Aber auch darüber hinaus kann Big Data in unzähligen Forschungsfeldern von großer Bedeutung sein. Im Gesundheitsbereich bietet die Untersuchung großer Datenmengen beispielsweise neue Möglichkeiten im Rahmen der Bekämpfung von Krankheiten und Epidemien (Krüger-Brand 2013). Mithilfe von Big-Data-Lösungen können unabhängig vom Forschungsfeld automatisiert Interdependenzen erkannt und Hypothesen gebildet werden.

CLOUD COMPUTING

2.

Im Zusammenhang mit Cloud Computing ist häufig von der Industrialisierung von IT-Ressourcen die Rede. Ähnlich wie es bei der Bereitstellung von Wasser oder Elektrizität mittlerweile selbstverständlich ist, werden beim Cloud Computing IT-Ressourcen nicht in eigenen Rechenzentren betrieben, sondern können bedarfsorientiert und flexibel über ein Netzwerk bezogen werden. Unternehmen sowie Einrichtungen der öffentlichen Verwaltung erwarten sich vom Einsatz von Cloud Computing vor allem die Verminderung von Kosten, Privatpersonen einen Komfortgewinn. Cloud Computing beinhaltet sowohl Technologien als auch Geschäftsmodelle, um IT-Ressourcen dynamisch zur Verfügung zu stellen und ihre Nutzung nach flexiblen Bezahlmodellen abzurechnen.

In diesem Teilkapitel wird zunächst diskutiert, was unter Cloud Computing verstanden wird und wie sich dieses Verständnis im Laufe der Zeit weiterentwickelt hat, anschließend werden die technischen Grundlagen von Cloud Computing erklärt und Anwendungsperspektiven aufgezeigt.

BEGRIFFSDEFINITION

2.1

Für den Begriff »Cloud Computing« konnte sich lange Zeit keine einheitliche Definition durchsetzen. Im Jahr 2009 veröffentlichte das in den USA ansässige National Institute of Standards and Technology (NIST) einen ersten Entwurf für die Definition des Begriffs, der unterschiedliche Ansätze bündelt und in der Folge häufig für Publikationen und Vorträge herangezogen wurde (z. B. Marston et al. 2011; Zhang et al. 2010). Zwei Jahre später wurde dann ohne wesentliche Änderungen die endgültige Version der Definition veröffentlicht (Liu et al. 2011). Im Rahmen der Definition wird Cloud Computing als ein Modell beschrieben, das es erlaubt, bei Bedarf jederzeit und überall bequem über ein Netzwerk auf einen geteilten Pool von konfigurierbaren IT-Ressourcen zuzugreifen, die schnell und mit minimalem Managementaufwand oder geringer Interaktion mit dem Anbieter zur Verfügung gestellt werden können. Das NIST konkretisiert sein Verständnis von Cloud Computing anhand von fünf wesentlichen Eigenschaften und beschreibt drei Kategorien von Diensten, die im Rahmen von Cloud Computing bereitgestellt werden (Service Models) sowie vier Formen der Bereitstellung (Deployment Models). Fünf wesentliche Eigenschaften charakterisieren im Rahmen von Cloud Computing bereitgestellte Dienste (Liu et al. 2011):

- > On-demand Self-Service: Die Bereitstellung von IT-Ressourcen läuft automatisch und ohne oder mit geringer Interaktion mit dem Anbieter ab.
- > Broad Network Access: Die IT-Ressourcen sind auf Basis von Standardmechanismen über ein Netzwerk verfügbar und nicht an eine bestimmte Clientplattform gebunden.



- › Resource Pooling: Die IT-Ressourcen eines Anbieters liegen in einem Pool vor, aus dem sich mehrere Nutzer bedienen können (Multi-Tenant-Modell). Physische und virtuelle Ressourcen werden den Nutzern abhängig vom jeweiligen Bedarf dynamisch zugewiesen. Für den Nutzer ist nicht ohne Weiteres ersichtlich, wo sich die verwendeten Ressourcen befinden.
- › Rapid Elasticity: Die IT-Ressourcen können bei Bedarf schnell und elastisch zur Verfügung gestellt werden, in manchen Fällen auch automatisch. Der Nutzer kann Ressourcen jederzeit und in beliebigem Ausmaß in Anspruch nehmen.
- › Measured Service: Die Nutzung von IT-Ressourcen kann gemessen und überwacht werden.

Die meisten anderen Begriffsdefinitionen charakterisieren Cloud Computing anhand von ähnlichen Eigenschaften (z. B. Gartner 2009; Schubert et al. 2010; Vaquero et al. 2008; Youseff et al. 2008). Gartner (2009) hebt beispielsweise die fünf Eigenschaften Service-based, Scalable and Elastic, Shared, Metered by Use und Uses Internet Technologies im Hinblick auf Cloud Computing hervor. Es gibt allerdings auch Ansätze, z. B. von einer Expertengruppe der Europäischen Kommission (Schubert et al. 2010), die Cloud Computing deutlich breiter definieren und als die Nutzung unterschiedlicher Ressourcen über eine Plattform verstehen (Schubert et al. 2010, S. 8).

Das NIST beschreibt drei Kategorien von Diensten, die im Rahmen von Cloud Computing bereitgestellt werden (Liu et al. 2011):

- › Software as a Service (SaaS): SaaS bezeichnet die Bereitstellung von Software. Anbieter betreiben ausgewählte Software auf ihrer Cloudinfrastruktur und erlauben Nutzern deren Verwendung. Nutzer können die Verbindung zur Software entweder über eine Thin-Client-Schnittstelle (z. B. einen Webbrowser) oder über eine API herstellen. Die Konfiguration der Software durch den Nutzer ist in der Regel nur in sehr beschränktem Umfang möglich. Die vom Anbieter verwendete Infrastruktur bleibt dem Nutzer verborgen.
- › Platform as a Service (PaaS): Unter PaaS wird die Bereitstellung von Programmier- und Laufzeitumgebungen mit flexiblen Rechen- und Datenkapazitäten verstanden. Nutzer können in der vom Anbieter bereitgestellten Programmier- und Laufzeitumgebung eigene Software entwickeln oder vorhandene Software ausführen. Der Anbieter stellt unter anderem Entwicklungswerkzeuge und Programmbibliotheken zur Verfügung. Während dem Nutzer die vom Anbieter verwendete Infrastruktur verborgen bleibt, kann er die Umgebung seinen Anforderungen entsprechend konfigurieren.
- › Infrastructure as a Service (IaaS): Bei IaaS werden Hardwareressourcen (z. B. Rechner, Netzwerke, Speicher) im Rahmen von Cloudangeboten bereitgestellt. Der Nutzer kann beliebige Software ausführen und ist für den Betrieb der Software selbst verantwortlich. Die verwendete Infrastruktur bleibt dem Nutzer weitgehend verborgen.



II. GRUNDLAGEN

Neben diesen Hauptkategorien werden teils weitere Bezeichnungen wie z. B. Storage as a Service, Backup as a Service oder auch Big Data as a Service (BDaaS) verwendet. Diese Bezeichnungen dienen Anbietern von Clouddiensten zur Differenzierung ihrer Angebote. Von den Kategorien, auf die das NIST im Rahmen seiner Definition nicht eingeht, hat vor allem Business Process as a Service (BPaaS) größere Aufmerksamkeit erlangt (z. B. Gartner 2012c; Graumann/Speich 2011). Bei BPaaS wird die Durchführung ganzer Geschäftsprozesse zu einem Anbieter ausgelagert. Der Anbieter stellt nicht nur alle IT-Ressourcen, sondern auch alle nicht IT-basierten Dienstleistungen, die für die Durchführung des Geschäftsprozesses benötigt werden, bereit. Der Unterschied zwischen BPaaS und Business Process Outsourcing (BPO) liegt in den Bereichen Vertragsgestaltung und Abrechnung. Während bei BPaaS keine längerfristige vertragliche Bindung besteht und die Abrechnung für einen bestimmten Anlassfall erfolgt, werden BPO-Verträge oft über viele Jahre abgeschlossen und unabhängig von der tatsächlichen Nutzung abgerechnet. Die Grenzen zwischen BPaaS und BPO sind nicht klar definiert. BPaaS zeichnet sich allerdings durch Cloudcharakteristika wie Elastizität, geteilte Ressourcen und von der tatsächlichen Nutzung abhängige Bezahlung aus (Ried et al. 2010). Dennoch geht das Verständnis von BPaaS als Cloud Computing klar über die engere Definition des NIST hinaus und unterstreicht den Trend, immer komplexere Prozesse als Services anzubieten.

Das NIST beschreibt vier Formen der Bereitstellung von Diensten (Liu et al. 2011), die sich vor allem im Hinblick auf den Betreiber der Cloudinfrastruktur und den Nutzerkreis unterscheiden:

- › Private Cloud: Die Cloudinfrastruktur steht nur einer Organisation, die normalerweise mehrere Nutzer umfasst, zur Verfügung. Die Infrastruktur kann von der Organisation selbst oder einem Dritten betrieben werden und sich im Rechenzentrum der eigenen oder einer fremden Organisation befinden.
- › Community Cloud: Die Cloudinfrastruktur steht nur einem bestimmten Nutzerkreis von Organisationen mit gemeinsamen Anliegen zur Verfügung. Die Infrastruktur kann von den beteiligten Organisationen selbst oder einem Dritten betrieben werden und sich in einem Rechenzentrum der beteiligten Organisationen oder einer fremden Organisation befinden.
- › Public Cloud: Die Cloudinfrastruktur steht der breiten Öffentlichkeit zur Verfügung und wird im Rechenzentrum des Anbieters betrieben.
- › Hybrid Cloud: Die Cloudinfrastruktur setzt sich aus mehreren Infrastrukturen zusammen, die voneinander unabhängig bleiben aber durch standardisierte oder proprietäre Technologien verbunden wurden, um die Portabilität und Interoperabilität von Diensten zu erlauben. Obwohl in der NIST-Definition nicht explizit die Rede davon ist, wird häufig die Verbindung von Private-Cloud- und Public-Cloud-Infrastrukturen als wesentliche Eigenschaft von Hybrid-Cloud-Infrastrukturen angesehen (z. B. Dillon et al. 2010; Velten/Janata 2012a). Auch die Nutzung von Public-Cloud-Diensten in Kombination



mit eigener IT-Infrastruktur wird gelegentlich als Hybrid Cloud verstanden. Dabei ist nicht unbedingt erforderlich, dass es sich bei der eigenen IT-Infrastruktur um eine Cloudinfrastruktur handelt.

Community Clouds und Hybrid Clouds sind Spezialfälle der primären Bereitstellungsformen Private Cloud und Public Cloud. Im Gegensatz zu Community Clouds, die in der Praxis eine eher untergeordnete Rolle spielen, ist die Verwendung von Hybrid Clouds durchaus üblich. Dies ist vor allem deshalb der Fall, da die Verfolgung von Multi-Sourcing-Strategien – also die Kombination unterschiedlicher Anbieter und Formen der Ressourcenbereitstellung – als ein vielversprechender Ansatz angesehen wird. Ein Überblick über die Bedeutung der einzelnen Bereitstellungsformen in der Praxis findet sich im Kapitel IV.2.

Manchmal ist zusätzlich von Managed Private Clouds und Managed Cloud/Cloud Hosting die Rede (Velten/Janata 2012a). Wenn zwischen Private Clouds und Managed Private Clouds unterschieden wird, stellt der Betreiber das entscheidende Kriterium dar. Während eine Private Cloud von der Organisation, die sie nutzt, selbst im eigenen Rechenzentrum betrieben wird, wird eine Managed Private Cloud von einem Dritten betrieben. Ob eine Managed Private Cloud in einem Rechenzentrum des Anbieters oder des Kunden betrieben wird, spielt dabei keine Rolle. Unter Managed Cloud/Cloud Hosting wird eine Cloudinfrastruktur verstanden, die von einem Dritten betrieben wird und mehreren Nutzern zur Verfügung steht.

Eigenschaften wie Measured Service oder Metered by Use erlauben eine Abrechnung, die sehr eng mit der tatsächlichen Nutzung eines Dienstes zusammenhängt. Dadurch wird neben der technischen auch eine wirtschaftliche Flexibilität erreicht. Grundsätzlich werden drei verschiedene Abrechnungsmodelle unterschieden:

- › Abonnementbasierte Modelle: In der Praxis, vor allem im Kontext von SaaS-Diensten, ist es beispielsweise üblich, dass bestimmte Dienste im Rahmen eines Abonnements genutzt werden, das monatlich oder jährlich verlängert werden kann. Salesforce.com oder Adobe sind bekannte Anbieter, die SaaS-Dienste auf Grundlage dieses Modells bereitstellen.
- › Zeitbasierte Modelle: Ein anderes weit verbreitetes Modell basiert auf der Zeitspanne zwischen Anfang und Ende der Dienstbereitstellung. In welchem Ausmaß der Dienst innerhalb dieser Zeitspanne genutzt wird, ist unerheblich. Amazon ist ein bekannter Anbieter, der auf Grundlage dieses Modells vor allem IaaS-Dienste bereitstellt.
- › Nutzungsbasierte Modelle: Alternativ dazu kann auch die tatsächliche Nutzungsdauer als Grundlage für die Abrechnung herangezogen werden. Wiederrum spielt es keine Rolle, in welchem Ausmaß der Dienst genutzt wird. Letztlich ist auch eine Abrechnung auf Grundlage des tatsächlichen Verbrauchs möglich. Dabei werden Kriterien wie die beanspruchte Menge an Speicher,



die benötigten CPU-Zyklen (Central-Processor-Unit-Zyklen) oder der erzeugte Netzverkehr herangezogen.

Nutzungsbasierte Modelle spielen in der Praxis eine untergeordnete Rolle, da eine höhere Genauigkeit bei der Messung des Verbrauchs zwangsläufig zu einem höheren Ressourcenverbrauch durch die Messung selbst führt. Die Gebühren für die Nutzung von IT-Ressourcen sind in der Regel statisch, können aber auch dynamisch mithilfe von Auktionen oder Spotmärkten festgelegt werden. Die dynamische Bestimmung von Gebühren gilt als komplex und ist in der Praxis nur selten anzutreffen.

Im Umfeld von Cloud Computing gibt es neben gebührenfinanzierten Diensten auch Dienste, die sich über Werbung finanzieren. Werbefinanzierte Dienste spielen vor allem dann eine große Rolle, wenn Privatpersonen die primäre Zielgruppe eines Dienstes sind. Alternativ zu oder in Kombination mit Werbeeinblendungen ist es auch möglich, dass das Nutzungsverhalten ausgewertet und gesammelte Daten verkauft werden.

Im Hinblick auf die Abrechnungsmodelle und Finanzierungsansätze ist eine Tendenz zur Hybridisierung erkennbar. Eine Form der Hybridisierung sind sogenannte Freemiumdienste, wo vor allem Privatpersonen bestimmte Dienste werbefinanziert zur Verfügung gestellt werden. Bei steigenden Anforderungen können nach Umstellung auf einen gebührenfinanzierten Ansatz zusätzliche Funktionen werbefrei genutzt werden. Hybride Abrechnungsmodelle und Finanzierungsansätze bieten vor allem im Zusammenhang mit der Entwicklung neuer Marktteilnehmer, wie beispielsweise den Serviceaggregatoren, die verschiedene Dienste zu neuen Angeboten zusammenstellen, oder Brokern, die Ressourcen vermitteln, Potenzial (Leimeister et al. 2010).

ENTWICKLUNG UND ABGRENZUNG

2.2

Die Idee, zentral bereitgestellte IT-Ressourcen über eine Art Versorgungsnetz zu nutzen, reicht in die frühen 1960er Jahre zurück. Normalerweise wird die Idee John McCarthy zugeschrieben, der sie bereits 1961 in einem Vortrag im Rahmen des Massachusetts Institute of Technology (MIT) Centennial ausführte (Garfinkel 1999). Entsprechende Überlegungen waren vor allem im Kontext der Entwicklung von Mehrbenutzerbetriebssystemen für die damaligen Großcomputer relevant. Mit der Miniaturisierung und Dezentralisierung verloren diese Überlegungen aber wieder an Bedeutung. Erst mit dem Erfolg des Internets in den 1990er Jahren wurden sie wieder aufgegriffen. Der Begriff »Cloud Computing« selbst wurde im Jahr 1997 zum ersten Mal im Rahmen eines Fachvortrags anlässlich der INFORMS-Tagung verwendet. Chellappa (1997) bezeichnete damit ein IT-Paradigma, in dem die genutzte Rechenleistung vielmehr auf Grund-

lage wirtschaftlicher Überlegungen festgelegt als durch technische Grenzen beschränkt wird. Seit 1999 sind erste Dienste auf dem Markt verfügbar, die im Rahmen von Cloud Computing angeboten werden.

Das Konzept hinter Cloud Computing ist nicht völlig neu, sondern stellt eine Weiterentwicklung im Umfeld von verschiedenen Ansätzen dar:

- › Application Service Provisioning (ASP): ASP beschreibt die Bereitstellung von Software, die auf die Bedürfnisse eines Nutzers oder einer Nutzergruppe zugeschnitten ist, über ein Netzwerk. Meist wird auf die Anwendungen mittels Webbrowser zugegriffen, es kann aber auch sein, dass nur die Server-Komponenten einer Software vom Anbieter betrieben werden und die Client-Komponente bei den Nutzern läuft. SaaS-Anbieter bieten im Gegensatz dazu Software, die in standardisierter Form und mit vom Anbieter definierten Service Level Agreements (SLAs) einem breiten Publikum angeboten wird. Charakteristisch für SaaS ist auch eine nutzungsabhängige Abrechnung. Cloud Computing verallgemeinert das Prinzip von SaaS dahingehend, dass neben Software auch andere IT-Ressourcen genutzt werden können. Cloud Computing gilt als Weiterentwicklung von ASP.
- › Distributed Computing: Distributed Computing erlaubt das Aufteilen sowie das parallele Verarbeiten von Aufgaben. Mehrere voneinander unabhängige Rechner werden über ein Netzwerk miteinander verbunden und bilden ein verteiltes System, das Rechenleistung zur Verfügung stellt (Attiya/Welch 2004).
- › Grid Computing: Grid Computing wird vor allem durch die drei Eigenschaften Decentralized Resource Control, Standardization und Non-trivial Qualities of Service charakterisiert (Foster 2002). Verteilte Rechner stellen gemeinsam und nach Bedarf Rechenleistung zur Verarbeitung umfangreicher Aufgaben über standardisierte Protokolle zur Verfügung (Foster et al. 2008). Aus technischer Sicht unterscheiden sich Cloud Computing und Grid Computing vor allem darin, dass beim Cloud Computing sämtliche Ressourcen zentral gesteuert werden, während es beim Grid Computing keine zentrale Steuerung der Ressourcen gibt. Während Clouddienste normalerweise für eine große Anzahl an Nutzern bereitgestellt werden, werden Grids meist für sehr spezielle Anwendungsfälle in Organisationen aufgebaut. Grid Computing wird meist als Form von Distributed Computing gesehen (Buyya et al. 2008; Foster et al. 2008).
- › Utility Computing: Unter Utility Computing wird die nach Verbrauch abgerechnete Bereitstellung von IT-Ressourcen in Anlehnung an die Bereitstellung von Wasser oder Elektrizität verstanden (Brynjolfsson et al. 2010). Bedeutung erlangte der Begriff Utility Computing 2005 durch einen Fachartikel von Carr (2005). Da die Grundprinzipien von Utility Computing wie Virtualisierung und Serviceorientierung heute zentrale Elemente des Verständnisses von Cloud Computing sind, verlor der Begriff Utility Computing zunehmend zugunsten von Cloud Computing an Bedeutung.

TECHNISCHE GRUNDLAGEN
2.3

Das NIST beschreibt im Hinblick auf die Architektur von Cloudinfrastrukturen drei Ebenen, die für die Bereitstellung von IT-Ressourcen im Rahmen von Cloud Computing zentral sind (Liu et al. 2011, S. 12 ff.) (Abb. II.2).

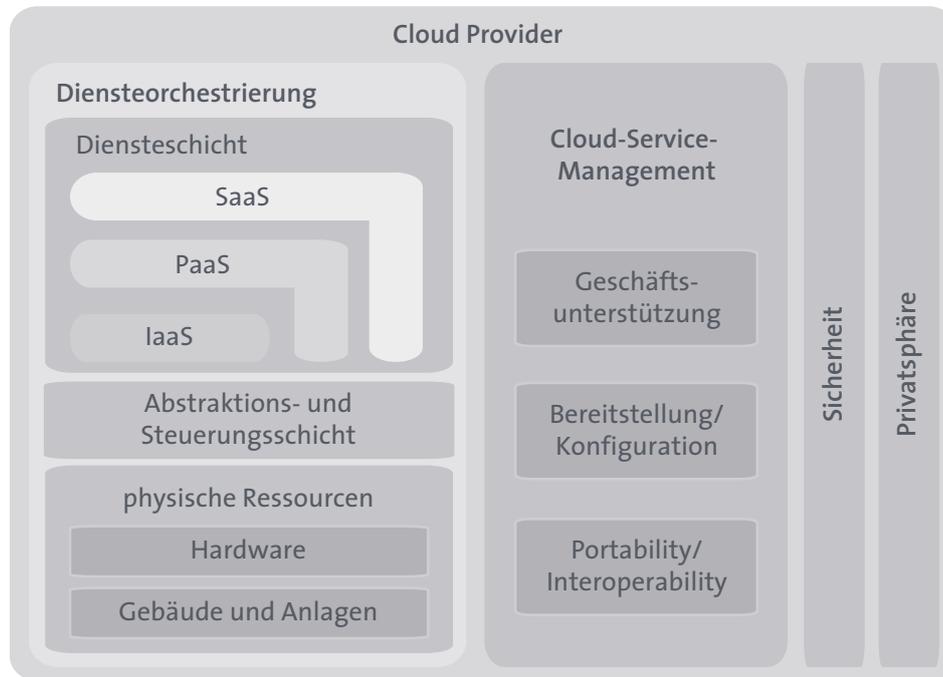
- › Service Layer: Es werden die Schnittstellen definiert, über die Nutzer auf Dienste aller Kategorien zugreifen können. Im Rahmen von Cloud Computing bereitgestellte Dienste können von anderen Diensten abhängig sein. Zum Beispiel können SaaS-Dienste auf PaaS-Diensten aufbauen und PaaS-Dienste auf IaaS-Diensten.
- › Resource Abstraction and Control Layer: Die Abstraktion von IT-Ressourcen erlaubt die Nutzung der Cloudinfrastruktur eines Anbieters durch mehrere Nutzer und wird in der Regel mithilfe spezieller Software realisiert, die die effiziente, sichere und zuverlässige Nutzung der Ressourcen gewährleisten soll. Die Steuerung von Ressourcen erfolgt ebenfalls mithilfe spezieller Software. Der Fokus liegt hier auf der Zuteilung von Ressourcen, der Zugriffskontrolle und der Überwachung der Nutzung.
- › Physical Resource Layer: Umfasst die physischen Ressourcen wie z. B. Rechner, Netzwerke und Speicher, aber auch Klimaanlage und die Stromversorgung in Rechenzentren.

Die Komponenten jeder Ebene sind von den Komponenten der darunterliegenden Ebenen abhängig. Die Aufgaben im Hinblick auf die Verwaltung der Dienste sind in der NIST-Architektur auf die Bereiche Business Support, Provisioning/Configuration und Portability/Interoperability aufgeteilt (Liu et al. 2011, S. 14 f.):

- › Business Support: Umfasst betriebswirtschaftliche Aufgaben, die den Umgang mit dem Kunden betreffen, wie die Verwaltung von Kundenkonten, Verträgen und Leistungskatalogen sowie die Buchführung, die Fakturierung, die Preisbestimmung und die Überwachung der Nutzung.
- › Provisioning/Configuration: Umfasst technische Aufgaben, die mit der Verwaltung der Cloudinfrastruktur zusammenhängen, wie die Bereitstellung benötigter IT-Ressourcen, der Austausch von Ressourcen, die Überwachung und Berichterstattung, die Messung der Nutzung und die Verwaltung von SLAs.
- › Portability/Interoperability: Mechanismen zur Förderung der Portabilität von Daten und Diensten von einem Anbieter zu einem anderen und der Interoperabilität von Diensten verschiedener Anbieter. Die Anforderungen in diesem Bereich hängen von den Kategorien der genutzten Dienste ab.

Im Rahmen der Beschreibung der Architektur von Cloudinfrastrukturen weist das NIST auch ausdrücklich auf die Bedeutung von Security und Privacy im Kontext von Cloud Computing hin (Liu et al. 2011, S. 15 ff.).

ABB. II.2 GÄNGIGE ARCHITEKTUR VON CLOUDINFRASTRUKTUREN AUF ANBIETERSEITE



Quelle: nach Liu et al. 2011

Es gibt auch andere Ansätze zur Beschreibung der Architektur von Cloudinfrastrukturen. Zhang et al. (2010, S. 9) unterscheiden z. B. vier Ebenen. Bei diesem Ansatz wird die mittlere Ebene der drei NIST-Ebenen auf zwei Ebenen aufgeteilt.

Im Rahmen von Cloud Computing werden Dienste normalerweise in Form von Webservices angeboten. Aus diesem Grund spielt das Architekturmuster der Serviceorientierten Architektur (SOA) im Hinblick auf Cloud Computing eine wichtige Rolle (Wei/Blake 2010). Charakteristisch für SOA sind die Verteilung der genutzten Services, die Verwendung offener Standards und die lose Kopplung von Services. Sicherheit und Einfachheit werden als zentrale Aspekte einer SOA angesehen.

Der Zugriff auf einen geteilten Pool von Ressourcen durch mehrere Nutzer (Multi-Tenant-Modell) stellt eine wesentliche Eigenschaft von Cloudangeboten dar. Jede Ressource kann mehreren Nutzern zugewiesen werden. Die Nutzer teilen sich eine physische Infrastruktur und können Teile der genutzten Anwendungen, aber nicht deren Quellcode an ihre Bedürfnisse anpassen. Virtualisierung stellt einen Ansatz zur Realisierung des Multi-Tenant-Modells dar.

Virtualisierung ist eine wesentliche Grundlage von Cloud Computing. Von physischen Ressourcen entkoppelte virtuelle Maschinen (VM) werden im Rahmen von Clouddiensten Nutzern zur Verfügung gestellt. Virtualisierung erlaubt die Reduktion der Komplexität für den Nutzer und vereinfacht die Steuerung für



den Anbieter (Schubert/Jeffery 2012). Eine VM ist die softwarebasierte Implementierung einer physischen Ressource (Smith/Nair 2005). Auf VM können je nach Bedarf vollständige Betriebssysteme (System Virtual Machine) oder nur einzelne Anwendungen (Process Virtual Machine) ausgeführt werden. Im Kontext von Cloud Computing werden verschiedene Arten von Virtualisierung unterschieden: Storage Virtualization, Platform Virtualization, Network Virtualization und Application Virtualization.

Für die Nutzung von Diensten, die im Rahmen von Cloud Computing bereitgestellt werden, wird eine Reihe von technischen Anforderungen als besonders wichtig angesehen (Leimbach et al. 2013):

- › **Netzwerkinfrastruktur:** Die Verwendung von Cloudangeboten setzt neben einer funktionierenden Netzwerkinfrastruktur auch eine stabile Verbindung zum Internet voraus. Netzwerkausfälle und Kapazitätsengpässe können im Kontext von Cloud Computing beachtliche Probleme darstellen. Berücksichtigt werden muss auch, dass (unsichere) Netzwerkverbindungen unter Umständen ein attraktives Ziel für Angreifer darstellen.
- › **Zuverlässigkeit und Fehlertoleranz:** Um zuverlässig zu sein, ist ein gewisser Grad an Fehlertoleranz erforderlich. Im Rahmen von Cloud Computing bereitgestellte Dienste müssen beispielsweise mit vorübergehenden Netzwerkausfällen und Fehlern auf einzelnen Systemen zurechtkommen. Zur Erhöhung der Zuverlässigkeit werden Daten normalerweise repliziert und in mehreren Rechenzentren abgelegt. Maßnahmen zur Erhöhung der Zuverlässigkeit von Diensten werden im Rahmen von SLAs zwischen Anbieter und Nutzer vereinbart (Wieder et al. 2011).
- › **Pooling und Sicherheit:** Die Verwendung eines Pools von Ressourcen, auf den mehrere Nutzer zugreifen können, ist ein wesentlicher Aspekt bei der Beurteilung der Sicherheit von Diensten. Schubert/Jeffery (2012) weisen vor allem auf die Wichtigkeit der isolierten Verarbeitung von Nutzungsdaten sowie die Gewährleistung der Sicherheit und Konsistenz von Daten hin.

EINSATZMÖGLICHKEITEN

2.4

Cloud Computing erlaubt die dynamisch an den Bedarf angepasste Nutzung von IT-Ressourcen (z. B. Rechenkapazität, Datenspeicher oder auch Software) über ein Netzwerk. Unternehmen, die öffentliche Verwaltung, die Wissenschaft und Privatpersonen können mithilfe von Cloud Computing IT-Ressourcen in beliebigem Ausmaß nutzen, ohne eine eigene IT-Infrastruktur, die möglicherweise nur zu Spitzenzeiten ausgelastet ist, betreiben zu müssen.

Public-Cloud-Dienste sind über das Internet prinzipiell öffentlich zugänglich. Jeder – egal ob Unternehmen oder Privatperson – hat im Normalfall gegen Ge-



büht Zugriff auf Dienste, die über eine Public Cloud bereitgestellt werden. Dennoch unterscheiden sich die Anwendungsperspektiven für die jeweilige Anwendergruppe teils recht deutlich. Für Unternehmen sowie für Stellen und Behörden der öffentlichen Verwaltung erscheint vor allem eine Unterstützung von relativ stark standardisierten Aufgabenfeldern wie beispielsweise des Customer-Relationship-Managements (CRM) oder des Enterprise Resource Planning (ERP) sinnvoll. Für Privatpersonen spielen hingegen eher Blogging- und Streamingdienste eine große Rolle. Allen Anwendergruppen bieten Public-Cloud-Dienste vor allem in den Bereichen Zusammenarbeit und Kommunikation interessante Anwendungsmöglichkeiten. Während bei Privatpersonen z. B. auch Fotospeicherdienste eine große Rolle spielen, erscheinen ansonsten vor allem auch Kalender- und Social-Network-Dienste sowie Telefoniedienste interessant. Im Hinblick auf Private-Cloud-Dienste ist noch deutlich mehr Individualität möglich.

FAZIT – CLOUD COMPUTING ALS BASIS VON BIG DATA 3.

Die Phänomene Big Data und Cloud Computing unterscheiden sich in vielerlei Hinsicht. Während der Begriff »Big Data« normalerweise im Zusammenhang mit Ansätzen zur effizienten Analyse besonders großer Datenmengen verwendet wird, beschreibt »Cloud Computing« einen Ansatz zur bedarfsgerechten Bereitstellung von IT-Ressourcen über ein Netzwerk. Auch wenn Big-Data-Analysen ohne Cloud Computing durchgeführt werden können und Cloud Computing neben Big-Data-Analysen zahlreiche andere Anwendungsfelder hat, wird der Kombination von Big Data und Cloud Computing ein größeres Potenzial zugeschrieben. Am augenscheinlichsten ist, dass Organisationen, aber auch Privatpersonen durch Cloud Computing die Möglichkeit bekommen, effizient große Datenmengen zu analysieren, ohne vorher eine entsprechende IT-Infrastruktur aufbauen zu müssen.

Einer Studie zufolge, die vom TDWI im Rahmen eines Best Practices Reports veröffentlicht wurde, liegen zu analysierende Daten bei den meisten Studienteilnehmern derzeit in einem Data Warehouse oder in einer traditionellen relationalen Datenbank (Russom 2011, S. 17). Den Studienergebnissen zufolge soll die Nutzung von Datenbanken zukünftig aber stark zurückgehen, während die Nutzung von verteilten Dateisystemen wie Hadoop und insbesondere von cloudbasierten Analyseplattformen weiter zunehmen wird. Im Rahmen der Studie wurden die Antworten von 325 Teilnehmern ausgewertet. Bei den Teilnehmern handelte es sich überwiegend um IT-Fachleute (58 %), aber auch um Nutzer aus den Geschäftsbereichen und Berater. Als Gründe für den Wechsel der Analyseplattform werden wachsende Datenmengen, zunehmende Anforderungen an die Geschwindigkeit und neue Modellierungs- und Analyseverfahren genannt (Russom 2011, S. 20). Gut 10 % der 700 Befragten, die an der »IDC European



II. GRUNDLAGEN

Software Survey 2013« teilnahmen, stimmten der Aussage, dass sie Big-Data-Analysen auf der Basis von Cloud Computing testen oder bereits produktiv nutzen, entweder zu oder stark zu (EITO 2013). An der Studie nahmen Vertreter von Organisationen aus Frankreich, Deutschland, Italien, den Niederlanden, Schweden, Finnland, Norwegen, Spanien und aus dem Vereinigten Königreich teil. EITO (2013, S. 13) hebt vier Treiber für die Durchführung von Big-Data-Analysen auf der Basis von Cloud Computing hervor:

- › Da sich die Daten von Organisationen zunehmend in der Cloud befinden, ist es naheliegend, dass die Daten auch in der Cloud analysiert werden (Data Gravity).
- › Cloud Computing erlaubt das Experimentieren mit Daten und die vorübergehende Durchführung von Analysen mit überschaubarem Aufwand.
- › Big-Data-Analysen auf der Basis von Cloud Computing bieten sich vor allem dann an, wenn mit temporären Anforderungsspitzen zu rechnen ist.
- › Externe Daten, z. B. aus sozialen Netzwerken, die im Rahmen einer Analyse berücksichtigt werden sollen, sind für lokal durchgeführte Analysen oft zu umfangreich und müssen gefiltert werden.

Die Durchführung von Big-Data-Analysen auf der Basis von Cloud Computing ist technisch ohne Weiteres möglich. Distributed Computing wird als Teil einer Strategie für Big-Data-Analysen gesehen (Jacobs 2009, S. 9). Beachtet werden muss allerdings, dass klassisches Distributed Computing den Fokus auf Rechenleistung legt und nicht auf die Verteilung von großen Datenmengen (Singh 2011). Während Ansätze und Techniken im Umfeld von Cloud Computing bereits eine gewisse Reife erreicht haben, wird im Kontext von Big-Data-Lösungen vielfach auf Schwächen aufgrund mangelnder Reife hingewiesen. Nichtsdestotrotz zeigen Organisationen ein wachsendes Interesse an der Kombination von Big-Data-Analysen und Cloud Computing. Organisationen können sich dabei grob zwischen IaaS- und SaaS-Angeboten entscheiden. Bei der Nutzung von IaaS-Angeboten stellt der Cloudanbieter nur die Infrastruktur zur Verfügung, und der Nutzer betreibt die zur Analyse großer Datenmengen nötige Software selbst. Bei der Nutzung von SaaS-Angeboten, die im Datenanalyse- bzw. Big-Data-Kontext auch als Analytics as a Service (AaaS) (SAS 2012) oder als BDaaS (EMC 2012; McKendrick 2013) bezeichnet werden, wird die gesamte Big-Data-Lösung vom Cloudanbieter bereitgestellt.

Technik allein reicht allerdings nicht aus. Der Schlüssel für den ökonomischen Erfolg ist die Entwicklung passgenauer Kundenlösungen und nachhaltiger Geschäftsmodelle (Glockner/Beyers 2012, S. 33). Mit den Märkten und Anbietern im Kontext von Big Data und Cloud Computing beschäftigt sich Kapitel III.



MÄRKTE UND ANBIETER

III.

BIG DATA

1.

Den großen Marktforschungsunternehmen zufolge gehören sowohl Big Data als auch Cloud Computing zu den am schnellsten wachsenden Teilmärkten des Marktsegments für Software und IT-Services. Es wird erwartet, dass beide Märkte in den kommenden Jahren beachtliche Auswirkungen auf das Marktsegment und die Nutzung von IT im Allgemeinen haben werden (EITO 2012b).

In diesem Teilkapitel wird der Teilmarkt für Big Data des Marktsegments für Software und IT-Services diskutiert und die entsprechende Anbieterlandschaft betrachtet.

MARKT

1.1

Es wird angenommen, dass das Interesse an Big Data in den nächsten Jahren stark zunehmen wird. Nach Ansicht des EITO (2012a, S. 25) gilt dies vor allem für die Bereiche In-Memory-Datenbanken, BI-Lösungen, Appliances, workload-optimierte Systeme und Datenbankanwendungen mit vorkonfigurierten Analysefunktionen. Ganz ähnlich sehen Wolfgang Martin Team & IT Research (2012) in Big Data vor allem einen Treiber für Innovationen in den Bereichen Datenhaltungstechnologien sowie Zugriff- und Speichermanagement. Darüber hinaus wird erwartet, dass Big Data ein wachsendes Interesse an geeigneten Datenintegrationswerkzeugen (vor allem im Hinblick auf Web und Cloud) und Lösungen zum Datenmanagement (Quellenidentifikation und -extraktion) bedingt.

Es wird zunächst das Volumen des Marktes verdeutlicht, dann werden die zentralen Investitionsbereiche betrachtet und anschließend regionale Unterschiede diskutiert. Im Rahmen der Diskussion der regionalen Unterschiede wird besonderes Augenmerk auf den deutschen Markt gelegt.

MARKTVOLUMEN

1.1.1

Die Angaben zum Big-Data-Markt des BITKOM basieren auf den Ergebnissen einer von der Experton Group im Jahr 2012 veröffentlichten Studie. Die globalen Umsätze am Markt für Big Data lagen laut BITKOM (2012) 2011 bei rund 3,3 Mrd. Euro. Für das Jahr 2012 wurde mit einem Umsatz von 4,5 Mrd. Euro gerechnet. Bis 2016 sollen die globalen Umsätze auf 15,7 Mrd. Euro wachsen. Das entspricht einer mittleren Wachstumsrate von 36 % pro Jahr. Der Markt für Big Data zählt somit zu den wachstumstärksten Teilmärkten des gesamten IT-



III. MÄRKTE UND ANBIETER

Marktes. Laut BITKOM wird die Wachstumsspitze in den Jahren 2013 und 2014 erwartet. Zum Abflachen des Wachstums soll einerseits der Preisverfall auf der Technologieseite und andererseits das zunehmende Fachwissen bei den Anwendern beitragen. IDC errechnete globale Umsätze von 4,8 Mrd. US-Dollar für das Jahr 2011 (Vesset et al. 2012, S.3) und weicht damit nur unwesentlich von der Berechnung des BITKOM ab. Für den Zeitraum von 2010 bis 2015 geht IDC von einer mittleren Wachstumsrate von 39 % pro Jahr aus (Tab. III.1). Transparency Market Research (2012) geht davon aus, dass die globalen Umsätze am Markt für Big Data von knapp 6,4 Mrd. US-Dollar im Jahr 2012 auf rund 48,3 Mrd. im Jahr 2018 anwachsen werden. Dies würde einer mittleren jährlichen Wachstumsrate von 40,5 % entsprechen.

TAB. III.1 ERWARTETE UMSÄTZE AM MARKT FÜR BIG DATA

Jahr	Umsatz (Mrd. US-Dollar)
2010	3,2
2011	4,7
2012	6,8
2013	9,7
2014	12,9
2015	16,9

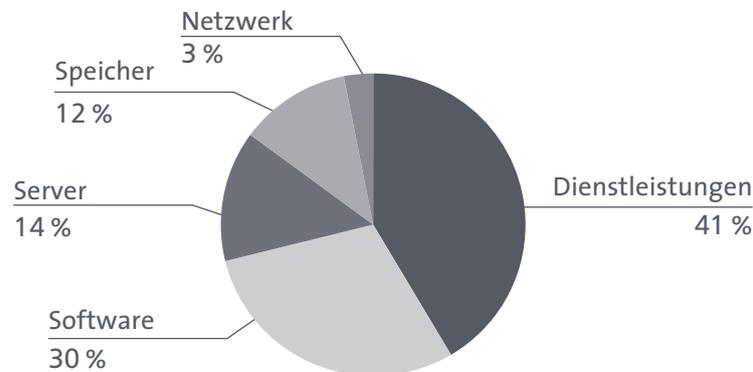
Quelle: Vesset et al. 2012, S.8

INVESTITIONSBEREICHE

1.1.2

Die Verteilung der globalen Umsätze auf die Kategorien »Dienstleistungen«, »Software« und »Hardware« ist ausgewogen. Auf jede der drei Kategorien entfällt ungefähr ein Drittel des Gesamtumsatzes, wobei laut BITKOM (2012, S.49) für Software und Hardware 2011 jeweils etwas mehr ausgegeben wurde als für Dienstleistungen. Mit zunehmender Marktreife soll es in Zukunft zu einer Verschiebung der Umsätze in Richtung Dienstleistungen kommen. Einer Studie von IDC zufolge trug die Kategorie »Dienstleistungen« bereits 2011 gut 41 % zum Gesamtumsatz bei (Vesset et al. 2012, S.3). IDC schlüsselte im Rahmen seiner Analyse die Kategorie »Hardware« weiter in die Unterkategorien »Server«, »Speicher« und »Netzwerk« auf. Mit 14 % Anteil am Gesamtumsatz ist die Kategorie »Server« noch vor der Kategorie »Speicher« mit 12 % die größte Unterkategorie. Die Unterkategorie »Netzwerk« ist mit einem Anteil von nur gut 3 % am Gesamtumsatz vergleichsweise klein (Abb. III.1). Bis 2015 erwartet IDC ein starkes Wachstum, vor allem im Hinblick auf die Unterkategorien »Speicher« und »Netzwerk«, während im Bereich »Server« nur ein moderates Wachstum erwartet wird (Vesset et al. 2012, S.8).

ABB. III.1

 UMSÄTZE MIT DIENSTLEISTUNGEN, SOFT-UND HARDWARE
 AM MARKT FÜR BIG DATA


Quelle: nach Vesset et al. 2012, S. 3

REGIONALE UNTERSCHIEDE

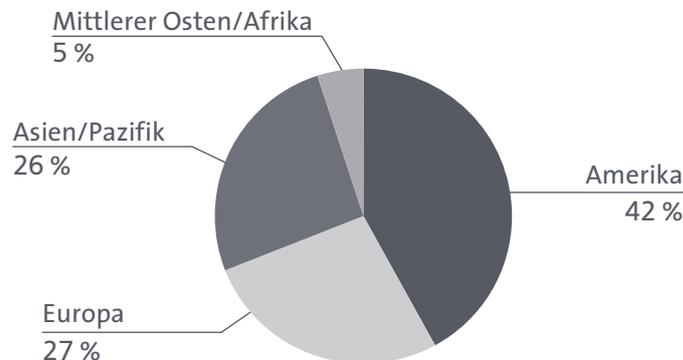
1.1.3

Ein wesentlicher Teil der Umsätze entfiel im Jahr 2011 mit 42 % auf den amerikanischen Raum (BITKOM 2012, S.49). Die USA sind derzeit der zentrale Big-Data-Entwicklungsstandort. Allerdings nimmt auch Europa mit 27 % der globalen Umsätze einen wichtigen Platz ein. Der BITKOM führt die starke Position des amerikanischen Raums unter anderem auf den volkswirtschaftlichen Stellenwert jener Branchen in den USA zurück, die federführend beim Einsatz von Big-Data-Analysen sind. Allerdings muss auch berücksichtigt werden, dass sich in den USA Regierungsbehörden und Nachrichtendienste für einen wesentlichen Teil der Umsätze verantwortlich zeigen. Mit 26 % fällt ebenfalls gut ein Viertel der globalen Umsätze im asiatischen Raum an. Laut BITKOM (2012, S.49) liegt das einerseits am hohen Einsatzgrad von Big Data in Japan und andererseits an den weitgehenden Befugnissen von Behörden und Unternehmen, vor allem in China, im Umgang mit personen- und firmenbezogenen Daten. Der mittlere Osten und Afrika spielen mit 5 % der weltweiten Umsätze nur eine untergeordnete Rolle (Abb. III.2).

Organisationen in Deutschland halten sich derzeit bei der Nutzung von Big-Data-Lösungen noch zurück. Auf deutsche Unternehmen entfällt derzeit nur ein Fünftel der Big-Data-Umsätze in Europa (BITKOM 2012, S.50). Der BITKOM geht allerdings davon aus, dass Deutschland im Jahr 2016 für rund die Hälfte der europäischen Big-Data-Umsätze verantwortlich sein wird. Es wird erwartet, dass der Markt für Big-Data-Lösungen in Deutschland bis 2016 von derzeit knapp 200 Mio. Euro auf rund 1,6 Mrd. Euro anwächst (BITKOM 2012, S.50). Das würde ein mittleres jährliches Wachstum von rund 48 % bedeuten.

ABB. III.2

UMSÄTZE AM MARKT FÜR BIG DATA NACH REGIONEN



Quelle: nach BITKOM 2012, S. 49

ANBIETER

1.2

Viele der jungen Unternehmen, deren Unternehmensstrategie und Portfolio ausschließlich auf Big Data ausgerichtet ist, agieren als Technologie- und Know-how-Partner etablierter und weltweit operierender Technologie- und IT-Serviceanbieter (BITKOM 2012, S. 48; Landrock 2013a, S. 4 f.). Die auf Big Data spezialisierten Unternehmen erwirtschaften immerhin rund 8 % des globalen Big-Data-Umsatzes. Dem BITKOM zufolge belegt diese Tatsache die These, dass Big Data ein neues Marktsegment bildet und nicht ein »recyceltes Business Intelligence« darstellt. Allerdings werden für den Big-Data-Markt, wie für einen jungen Markt nicht unüblich, zahlreiche Übernahmen und Zusammenschlüsse erwartet (Wolfgang Martin Team & IT Research 2012, S. 38). Wolfgang Martin Team & IT Research (2012, S. 38) gehen außerdem davon aus, dass Open-Source-Angebote sowie cloudbasierte Angebote im Big-Data-Markt eine größere Rolle spielen werden als im Markt für BI.

Zunächst wird die Anbieterlandschaft allgemein beschrieben, dann werden die wichtigsten etablierten Anbieter und Start-ups betrachtet sowie daran anschließend die Ergebnisse einer Studie der Experton Group zur Situation in Deutschland zusammengefasst dargestellt (Kap. III.1.2.1 bis III.1.2.4).

ANBIETERLANDSCHAFT

1.2.1

Bereits im Jahr 2012 gab es weltweit mehr als 80 Anbieter mit konkreten Big-Data-Lösungen (BITKOM 2012, S. 47). Die Anbieterlandschaft kann im Hinblick auf Big Data in zwei große Gruppen unterteilt werden (Korolov 2013). Zum einen gibt es Start-ups, die mit dem Ziel, den Big-Data-Markt zu bedienen, entstanden sind, und zum anderen etablierte Anbieter von Datenbanken und



Data-Warehouse-Lösungen, die versuchen ihr Geschäftsfeld auf den Markt für Big Data auszudehnen. Vertreter der zweiten Gruppe haben nicht nur den Vorteil einer bestehenden Kundenbasis, sondern auch den einer bereits erprobten Produktpalette.

Analysten um Jeff Kelly veröffentlichten Anfang 2013 auf Wikibon unter dem Titel »Big Data Vendor Revenue and Market Forecast 2012–2017«⁴ eine umfassende Marktstudie, die im Laufe des Jahres mehrfach aktualisiert wurde. Obwohl die Angaben zu Umsatzzahlen teilweise auf Schätzungen beruhen, stellt die im Rahmen der Studie veröffentlichte Liste der Anbieter auf dem Markt für Big Data einen guten Anhaltspunkt für eine Diskussion der Anbieterlandschaft dar. Zum Verständnis der Anbieterlandschaft spielen weniger die konkreten Zahlen, sondern vielmehr die Größenverhältnisse eine Rolle. Die Liste vermittelt nicht nur einen guten Eindruck von den Größenverhältnissen zwischen den Start-ups und den etablierten Anbietern, sondern auch von der unterschiedlichen Gewichtung der Bereiche »Hardware«, »Software« und »Dienstleistungen« bei den einzelnen Anbietern.

Im Rahmen einer von BARC durchgeführten Studie wurde erhoben, wie bekannt einzelne Anbieter im Bereich Big Data sind (Bange et al. 2013, S. 37 f.). An der Studie nahmen 274 Fachbereichs- und IT-Verantwortliche aus Deutschland, Österreich der Schweiz, Frankreich und dem Vereinigten Königreich teil. Etablierte Anbieter wie IBM, Oracle, SAP, Microsoft und Teradata wurden im Hinblick auf die Bereiche Hardware und Software von besonders vielen Studienteilnehmern genannt. Aber auch jüngere Unternehmen wie QlikTech und Cloudera haben es geschafft, mit Big Data in Verbindung gebracht zu werden. Im Hinblick auf Dienstleister im Bereich Big Data wurden von den Studienteilnehmern neben IBM und Microsoft auch Capgemini und Accenture besonders häufig genannt.

ETABLIERTE ANBIETER

1.2.2

Im Jahr 2012 war IBM mit einem Umsatz von rund 1,3 Mrd. US-Dollar im Bereich Big Data der weltgrößte Anbieter auf dem Markt. Bei einem Gesamtumsatz von rund 104 Mrd. US-Dollar trägt der Bereich Big Data derzeit nur gut 1 % zum Gesamtumsatz von IBM bei. IBM bietet laut Korolov (2013) nicht nur in den Bereichen Server und Storage an, sondern ist mit Produkten wie DB2, Informix und InfoSphere auch im Bereich Datenbanken aktiv. Darüber hinaus bietet IBM mit Software wie Cognos und SPSS auch im Bereich Analyticslösungen an. IBM ist außerdem an der Weiterentwicklung von Hadoop beteiligt.

4 http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2012-2017



HP war 2012 mit einem Umsatz von rund 664 Mio. US-Dollar im Bereich Big Data der zweitgrößte Anbieter. Obwohl HP genauso wie IBM Hardware, Software und Dienstleistungen für den Big-Data-Markt anbietet, ist das Unternehmen im Big-Data-Kontext vor allem für die Analyticslösung Vertica bekannt (Korolov 2013). Mit einem Umsatz von rund 435 Mio. US-Dollar folgt Teradata auf Platz drei der größten Big-Data-Anbieter. Der Anteil des Umsatzes mit Big Data am Gesamtumsatz ist bei Teradata deutlich größer als bei IBM oder HP. Bei Teradata machte der Umsatz im Bereich Big Data im Jahr 2012 immerhin gut 16 % des Gesamtumsatzes von rund 2,7 Mrd. US-Dollar aus. Teradata ist vor allem für seine Hardwarelösungen sowie für Software in den Bereichen Datenbanken und Analytics bekannt. Das Unternehmen bietet unter anderem spezielle Analyticslösungen für den Einzelhandel und das Transportwesen an (Korolov 2013). Knapp dahinter folgen Dell und Oracle mit einem Umsatz von 425 bzw. 415 Mio. US-Dollar. Dell ist im Hinblick auf den Big-Data-Markt insbesondere im Bereich »Hardware« aktiv. Die von Oracle angebotene Big-Data-Appliance kombiniert einen Intel-Server, Clouderas Hadoop-Distribution und die hausinterne NoSQL-Datenbank (Korolov 2013).

SAP erzielte mit Big-Data-Lösungen 2012 einen Umsatz von rund 368 Mio. US-Dollar. Damit ist SAP der weltweit größte Anbieter am Markt für Big Data mit Zentrale außerhalb der USA (Korolov 2013). SAP bedient den Markt mit einer Reihe von Lösungen aus dem Analytikbereich, hat jedoch vor allem durch seine In-Memory-Datenbank HANA im Kontext von Big Data Bekanntheit erlangt (Korolov 2013). Mit einem Umsatz von rund 336 Mio. US-Dollar ist EMC auch eines der großen Unternehmen am Markt für Big Data. Bekanntheit erlangte EMC im Kontext von Big Data vor allem durch das Spin-off Pivotal, das Hadoop mit der Datenbank Greenplum von EMC und HAWQ-Abfragerwerkzeugen kombiniert (Korolov 2013). Darüber hinaus gehört das Marketing Science Lab, ein Big-Data-Think-Tank, der sich auf die Analyse von Marketingdaten spezialisiert hat, zu EMC.

Weitere etablierte Anbieter, die ihr Geschäftsfeld auf den Markt für Big Data ausgedehnt haben, sind laut Korolov (2013) Amazon, Microsoft, Google und VMware. Amazon, Google und VMware sind vor allem auch im Bereich der cloudbasierten Big-Data-Lösungen aktiv.

START-UPS

1.2.3

Eines der bekanntesten Unternehmen in der Gruppe der Start-ups ist Cloudera, das schon mehr als 141 Mio. US-Dollar an Risikokapital erhalten hat und bereits 2008 Hadoop ersten Unternehmen näher brachte. Hortonworks, ein Yahoo!-Spin-off, hat seit 2011 bereits mehr als 70 Mio. US-Dollar an Risikokapital erhalten. Hortonworks ist vor allem aufgrund seiner strategischen Partner-

schaften mit Unternehmen wie Microsoft, Rackspace, Red Hat und Teradata bekannt. Splunk, ein Unternehmen, das sich auf die Analyse von Maschinendaten spezialisiert hat, hat von allen Unternehmen, die ausschließlich den Big-Data-Markt bedienen, den größten Marktanteil (Korolov 2013). Im Jahr 2012 erzielte Splunk einen Umsatz von rund 186 Mio. US-Dollar. Zwei weitere Start-ups, die mit dem Ziel den Big-Data-Markt zu bedienen gegründet wurden, sind 10Gen und MapR.

Henschen (2012) bestätigt das Potenzial der von Korolov (2013) genannten Start-ups und benennt mit CouchBase, Datameer, DataStax, Hadapt, Karmasphere, Neo Technology und Platfora eine ganze Reihe weiterer. Es fällt auf, dass cloudbasierte Big-Data-Lösungen bei Start-ups, die mit dem Ziel den Big-Data-Markt zu bedienen entstanden sind, eine deutlich größere Rolle spielen als bei etablierten Anbietern, die versuchen ihr Geschäftsfeld auf den Markt für Big Data auszudehnen. Eine Ausnahme stellt Amazon dar, das mit Elastic MapReduce bereits vor mehreren Jahren eine entsprechende Lösung auf den Markt brachte (Henschen 2012).

Einschätzungen im Hinblick auf die Bedeutung und das Potenzial der zahlreichen Start-ups gehen zum Teil weit auseinander. Zum Beispiel enthält eine von Philip (2013) veröffentlichte Liste der 20 vielversprechendsten Unternehmen am Markt für Big Data mit Hadapt nur ein Unternehmen, das auch von Korolov (2013) oder Henschen (2012) genannt wurde. Die Liste, die ausschließlich US-amerikanische Unternehmen umfasst, basiert auf den Einschätzungen von Geschäftsführern, IT-Verantwortlichen, Risikokapitalgebern und Analysten. Furrier (2012) und Press (2013b) verwenden die auf Wikibon veröffentlichte Studie als Grundlage für ihre Diskussion von Start-ups. Eine besonders umfangreiche Liste von Anbietern auf dem Markt für Big Data stellt Feinleib (2013) im Rahmen der »Big Data 100« vor. Die Abweichungen sind vermutlich primär auf unterschiedliche Bewertungskriterien und die Schwierigkeiten bei deren Einschätzung zurückzuführen.

SITUATION IN DEUTSCHLAND

1.2.4

Die Experton Group (2013) hat 2013 eine Anbieterbewertung für Deutschland durchgeführt. Für den »Big Data Vendor Benchmark« wurden 68 Anbieter untersucht, deren Angebote auf dem deutschen Markt erhältlich sind. Die Anbieter wurden in drei Gruppen eingeteilt: Anbieter von Big-Data-Software, Anbieter von Technologien für Big-Data-Infrastrukturen (Hardware) und Dienstleister im Bereich Big Data. Im Hinblick auf Software wurde weiter zwischen Lösungen für Analytics, Aggregation und Visualisierung unterschieden. Die Lösungskategorie »Visualisierung« umfasst auch Lösungen für Syndizierung und Reporting. Im Hinblick auf Hardware wurde noch zwischen Lösungen für Storage, Appliances



und Datenbanken unterschieden. Der Bereich Dienstleistungen umfasst die Kategorien »Projekte und Beratung« sowie »IT-Operations«.

Die Kategorie »Appliances« nimmt in gewisser Weise eine Sonderstellung ein, da sie Angebote im Sinne von Komplettsystemen mit Server, Storage, Netzwerk, Datenbank, Middleware und Applikationen bis hin zum Accesslevel beinhaltet. Die Experton Group rechnet damit, dass auch weitere Anbieter dem Wunsch von Anwendern nach Komplettlösungen nachkommen und in diesen Markt eintreten werden. Die Unternehmen Teradata, IBM, EMC, HP und Exasol werden im »Big Data Vendor Benchmark« aufgrund ihrer Marktpräsenz in Deutschland sowie ihres Marktangebots dem Leader-Quadranten zugeordnet.

Aufgrund des relativ großen Anteils am Gesamtumsatz spielen auch die dem Bereich Dienstleistungen zugeordneten Kategorien »Projekte und Beratung« sowie »IT-Operations« eine besondere Rolle. Im Hinblick auf Projekte und Beratung gehören die Unternehmen IBM, ATOS, Capgemini, Informatica, HP, T-Systems, Fujitsu/TDS und Steria Mummert dem Leader-Quadranten an (Landrock 2013a, S. 5). Was »IT-Operations« betrifft, sind es die Unternehmen IBM, T-Systems, Atos, HP und Fujitsu/TDS (Landrock 2013a, S. 5). Es fällt auf, dass alle Leader in der Kategorie »IT-Operations« auch im Hinblick auf die Kategorie »Projekte und Beratung« eine führende Position einnehmen.

Da der Markt für Big Data noch relativ jung ist, bewegen sich die Unternehmen, die ihn bedienen, in einem schmalen Wettbewerbsfeld. Unternehmen im Leader-Quadranten zeichnen sich daher insbesondere dadurch aus, dass sie bereits größere und komplexere Projekte mit zahlreichen Datenquellen erfolgreich abgewickelt haben. Die Experton Group betont allerdings, dass für Anwender keinesfalls nur die Leader interessant sind. Auch im Bereich der »Product Challenger« und »Follower« gibt es Unternehmen, deren Ansätze vielversprechend sind. Vor allem im Hinblick auf Software zur Aggregation und Aufbereitung großer Datenmengen aus unterschiedlichen Quellen gibt es aktuell nur sehr wenige Anbieter auf dem deutschen Markt. Im Gegensatz dazu befassen sich sehr viele Anbieter mit Software, die der Kategorie »Analytics« zuzuordnen ist. Dies führt die Experton Group vor allem auf die Erfahrung vieler Anbieter im BI-Umfeld und die entsprechende Produkthistorie zurück.

CLOUD COMPUTING

2.

Aufgrund von Unterschieden im Hinblick auf den gewählten methodischen Ansatz, die Abgrenzung von Marktsegmenten und Teilmärkten sowie Annahmen zur wirtschaftlichen Entwicklung unterscheiden sich die Ergebnisse der Studien der großen Marktforschungsunternehmen Gartner, IDC und Forrester zum Teil erheblich. Trotzdem lassen sich bestimmte Entwicklungen des Marktes für Cloud Computing ablesen.

In diesem Teilkapitel wird zunächst der Teilmarkt für Cloud Computing des Marktsegments für Software und IT-Services diskutiert. Anschließend erfolgt eine Betrachtung der Anbieterlandschaft.

MARKT

2.1

Cloud Computing erlaubt Anbietern nicht nur die Befriedigung einer wachsenden Nachfrage nach flexibel nutzbaren IT-Ressourcen sondern auch die Steigerung der Auslastung der eigenen Rechenzentren.

Zunächst wird das Volumen des Marktes verdeutlicht, dann werden die unterschiedlichen Kategorien von Clouddiensten betrachtet und anschließend regionale Unterschiede diskutiert. Im Rahmen der Diskussion der regionalen Unterschiede wird besonderes Augenmerk auf den deutschen Markt gelegt.

MARKTVOLUMEN

2.1.1

Die globalen Umsätze am Markt für Public-Cloud-Computing lagen 2011 laut Studien von IDC und Forrester bei rund 40 Mrd. US-Dollar (Dignan 2011; IDC 2012b). Gartner (2012c) kam gar auf einen Umsatz von 91,4 Mrd. US-Dollar. Berücksichtigt wurden in den drei Studien ausschließlich Umsätze mit Public-Cloud-Diensten. Für das Jahr 2016 rechnet Gartner mit einem Umsatz von 206,6 Mrd. US-Dollar. Forrester sagt für 2020 einen globalen Umsatz im Bereich Cloud Computing von 241 Mrd. US-Dollar voraus. Es fällt auf, dass über die nächsten Jahre alle großen Marktforschungsunternehmen von einer durchschnittlichen jährlichen Wachstumsrate von mehr als 20 % ausgehen (Tab. III.2).

TAB. III.2 ERWARTETE UMSÄTZE AM MARKT FÜR CLOUD COMPUTING

Jahr	Umsatz (Mrd. US-Dollar)		
	Gartner	IDC	Forrester
2011	91,0		40,7
2012		40,0	
2013	131,0	47,4	
2014			
2015			97,0
2016	206,6	100,0	113,9
2017	244,0	107,0	

Quellen: Dignan 2011; Gartner 2012c; IDC 2012b



III. MÄRKTE UND ANBIETER

Neben dem Markt für Public-Cloud-Computing erfassen Marktforschungsunternehmen zum Teil auch andere Marktsegmente. So kommt z. B. IDC für das Jahr 2013 im Hinblick auf cloudbezogene IT-Dienstleistungen auf globale Umsätze in Höhe von 9,6 Mrd. US-Dollar. Den Markt für Managed Private Clouds schätzt IDC auf rund 26 Mrd. US-Dollar im Jahr 2016 (IDC 2013b; Little et al. 2013).

Aufgrund der Tatsache, dass die durchschnittliche jährliche Wachstumsrate deutlich über jener des gesamten Marktsegments für Software und IT-Services liegt, gehen die Marktforschungsunternehmen davon aus, dass der Anteil von Cloud Computing am gesamten Marktsegment kontinuierlich steigen wird. Basierend auf den Angaben von IDC kann man davon ausgehen, dass der Anteil von Public-Cloud-Diensten bei aktuell 3 bis 5 % liegt und sogar bei 6 bis 7 %, wenn weitere Segmente wie cloudbezogene IT-Dienstleistungen oder Managed Private Clouds berücksichtigt werden (EITO 2012a; IDC 2012b, 2013a, 2013b; Little et al. 2013). Es wird angenommen, dass allein der Anteil von Public-Cloud-Diensten im Laufe der nächsten fünf Jahre auf über 10 % und über die nächsten zehn Jahre auf über 20 % steigen wird (Gartner 2012c; IDC 2012b). Insgesamt wird sich Cloud Computing voraussichtlich zu einem wesentlichen Segment des Marktes für Software und IT-Services entwickeln.

Während das Wachstum im Bereich Cloud Computing einerseits zum Wachstum des gesamten Marktsegments beitragen wird, ist andererseits mit dem schrumpfen einiger Bereiche zu rechnen. Vor allem im Hinblick auf Software, die auf Basis von Lizenzgebühren und Wartungsverträgen vertrieben wird, aber auch im Bereich IT-Outsourcing sind Einbußen zu erwarten. Eine Einschätzung der genauen Auswirkungen gestaltet sich schwierig. Aumasson et al. (2010, S. 115 ff.) vermuten, dass Entwicklungen in den Bereichen Mobile Computing und Cloud Computing ab 2016 zu einer Stagnation bei Lizenzgebühren führen werden.

Noch offen ist, wie sich die Enthüllungen zu den Überwachungspraktiken der NSA und anderer Geheimdienste auswirken werden. Die Auswirkungen wurden bisher von Marktforschern noch wenig thematisiert, aber vor allem in den USA ist eine Debatte darüber entbrannt. So sorgen in den USA sowohl Umfragen, die auf eine Zunahme abgebrochener Cloudprojekte vor allem ausländischer Kunden hindeuten (Johnson 2013), sowie Studien, die auf Basis der relativ hohen Gartner-Schätzungen bis 2016 einen kumulierten Verlust für amerikanische Anbieter von bis zu 35 Mrd. US-Dollar vorhersagen (Castro 2013), für Besorgnis. Mit welcher Wahrscheinlichkeit es tatsächlich zu Verlusten in dieser Größenordnung kommt, ist aufgrund fehlender Daten schwer abschätzbar. Erste Marktforschungsergebnisse legen nahe, dass zumindest in Deutschland, insbesondere im Mittelstand, trotz der zunehmenden Enthüllungen der Trend Richtung Cloud Computing weiterhin anhält. Bemerkenswert ist dabei, dass sich der Fokus auf Private Clouds und hybride Nutzungsformen weiter verstärkt zu haben scheint

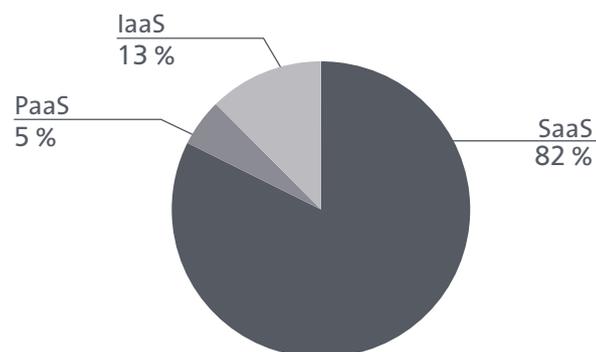
(Böckle 2014; Hottelot 2014). Zudem sehen einige Umfragen gleichzeitig einen Trend in Richtung deutscher und europäischer Anbieter. Zwar lässt sich dieser bisher nicht anhand von konkreten Zahlen belegen, und auch amerikanische Anbieter vermelden weiterhin Umsatzzuwächse, doch weisen die Forderungen der großen Anbieter in den USA nach mehr Transparenz im Hinblick auf die Geheimdienstaktivitäten in den letzten Monaten auf gewisse Befürchtungen hin. Langfristig könnte dies neue Chancen für heimische Anbieter ermöglichen (Hackmann 2014; Hottelot 2014).

DIENSTKATEGORIEN

2.1.2

Auch im Hinblick auf die unterschiedlichen Kategorien von Diensten unterscheiden sich die Studien der großen Marktforschungsunternehmen zum Teil erheblich. Die Marktforschungsunternehmen betonen die große Bedeutung von BPaaS- (Gartner 2012c) und SaaS-Diensten (Dignan 2011; IDC 2012b). Sie bilden nicht nur derzeit die umsatzstärkste Dienstkategorie, sondern werden das auch in Zukunft tun (Abb. III.3). Obwohl angenommen wird, dass der Umsatz mit PaaS- und IaaS-Diensten in den nächsten Jahren schneller wachsen wird, wird mit BPaaS- und SaaS-Diensten auch mittelfristig noch der größte Umsatz erzielt werden (IDC 2012b). Gartner (2012c) hebt im Zusammenhang mit der Bedeutung von BPaaS-Diensten die herausragende Rolle von Cloud Advertising hervor.

ABB. III.3 UMSÄTZE MIT SAAS, PAAS UND IAAS AM MARKT FÜR CLOUD COMPUTING



Quelle: nach Dignan 2011

Studien belegen eine zunehmende Vielfalt an Diensten. Insbesondere die zunehmenden Nutzererfahrungen führen dazu, dass immer komplexere Anwendungen, z.B. im ERP-Bereich, in die Cloud ausgelagert werden. Es gibt allerdings auch zahlreiche völlig neue Angebote, die durch die Verfügbarkeit anderer Clouddienste erst realisierbar wurden. Die zunehmende Vielfalt an Diensten führt natürlich auch zu einer Umverteilung der Umsätze. Während in der Ver-



gangenheit einige wenige SaaS-Dienste, insbesondere im Bereich CRM, das Segment dominierten, führt die Nutzererfahrung zusammen mit neuen Clouddiensten zu einem Trend in Richtung ERP- und BI-Lösungen. Im IaaS-Segment haben sich die in der Vergangenheit häufig genutzten Speicherdienste immer mehr zu umfassenden Backup-Diensten entwickelt.

REGIONALE UNTERSCHIEDE

2.1.3

Derzeit ist der nordamerikanische Markt nicht nur der umsatzstärkste Markt für Cloud Computing, sondern auch der am schnellsten wachsende. Den aufstrebenden asiatischen Märkten, insbesondere denen von Indien, Indonesien und China, wird großes Potenzial im Hinblick auf Cloud Computing zugeschrieben. Der europäische Markt ist derzeit nach dem nordamerikanischen der zweitgrößte vor dem japanischen und den Märkten anderer Regionen in Asien (Gartner 2012c; IDC 2012b).

Das schnelle Wachstum in den aufstrebenden asiatischen Märkten wird darauf zurückgeführt, dass viele Unternehmen in diesen Regionen keine komplexen, über Jahre hinweg gewachsenen IT-Infrastrukturen haben. Daher fällt ihnen die Migration auf neue Modelle wie Cloud Computing deutlich leichter. Clouddienste werden sich laut Gartner (2012a) in Europa, insbesondere im Vergleich zu den USA, langsamer entwickeln. Als Gründe werden neben der allgemeinen Zurückhaltung gegenüber Cloud Computing in Europa unter anderem die Eurokrise und die anhaltende Rezession sowie die zahlreichen Unterschiede zwischen den einzelnen Staaten Europas genannt.

Prognosen von Pierre Audoin Consultants (PAC) deuten laut Fielder et al. (2012, S. 20) jedoch darauf hin, dass auch der europäische Markt für Cloud Computing an Fahrt aufnimmt. Die Ergebnisse von PAC zeigen unter anderem, dass der Umsatz mit IaaS-Diensten (inklusive Speicherdienste) in Europa überraschenderweise größer ist als jener mit SaaS-Diensten. Das könnte bedeuten, dass europäische Unternehmen zu Angeboten tendieren, die ein höheres Maß an Kontrolle erlauben. Dies könnte auch die im Vergleich zu den USA stärkere Nutzung von Private-Cloud-Angeboten in Europa erklären.

Auf Basis aktueller Prognosen der Experton Group berichtet der BITKOM (2013), dass der Umsatz im Bereich Cloud Computing im Jahr 2012 in Deutschland bei rund 5 Mrd. Euro lag, was im Vergleich zu 2011 einer Umsatzsteigerung von 47% entspricht. Für 2013 wird eine Umsatzsteigerung von knapp 50% vorhergesagt und auch in den Folgejahren sollen die Zuwächse im zweistelligen Bereich liegen. Es wird angenommen, dass der deutsche Cloud-Computing-Markt bis 2016 auf rund 20 Mrd. Euro wachsen wird. Besonders stark legt den Prognosen zufolge der Umsatz mit Geschäftskunden zu. Der Hauptteil dieses Umsatzes entfällt auf die Nutzung von Diensten, während auf Technologie, Integration und Beratung kleinere Teile entfallen.

ANBIETER**2.2**

Die Anzahl der Anbieter im Bereich Cloud Computing nimmt rasant zu. Neben zahlreichen kleineren Unternehmen, die auf dem Cloud-Computing-Markt tätig sind, gibt es nur wenige große Anbieter. Die Anbieter lassen sich in der Regel einer der vier Kategorien »Early Movers«, spezialisierte Technologie- und Softwareanbieter, große IT-Dienstleister und Cloud-born-Unternehmen zuordnen.

Es werden zunächst unterschiedliche Kategorien von Anbietern betrachtet. Anschließend werden die Ergebnisse einer Studie der Experton Group zur Situation in Deutschland zusammengefasst dargestellt.

ANBIETERKATEGORIEN**2.2.1**

»*Early Movers*«: Zu den Unternehmen, die früh in den Markt eingetreten sind, gehören insbesondere Anbieter wie Amazon, Google oder Salesforce.com. Der Kerngeschäftsbereich von Amazon ist E-Commerce. Das Engagement im Cloudmarkt war naheliegend, um die umfangreichen Ressourcen, die zur Erbringungen des Kerngeschäfts betrieben werden müssen, besser auszunützen. Bei Google entwickelte sich das angebotene Service-Portfolio, das mittlerweile zahlreiche Cloudservices umfasst, rund um die bekannte Suchmaschine. Salesforce.com begann 1999 zunächst als ASP-Anbieter. Erst Mitte der 2000er Jahre gelang Salesforce.com mit seiner CRM-Lösung der Durchbruch.

Spezialisierte Technologie- und Softwareanbieter: Unternehmen wie VMware, Citrix, Terremark und Rackspace haben sich auf die Bereitstellung von Technologien und Software spezialisiert, die für den Aufbau und Betrieb von Cloudinfrastrukturen benötigt werden. Einige Unternehmen dieser Gruppe haben sich mittlerweile zu Public-Cloud-Anbietern entwickelt. In dieser Kategorie kam es zuletzt verstärkt zu Übernahmen und Zusammenschlüssen. Insbesondere Citrix und Rackspace kauften in den letzten Jahren zahlreiche kleine Anbieter und Technologiespezialisten wie Xen (Citrix) oder Slicehost und JungleDisk (Rackspace). VMware und Terremark wurden selbst Ziel von Übernahmen, als sie von EMC bzw. Verizon übernommen wurden.

Große IT-Dienstleister: Bei IBM, HP, Dell oder auch Cisco handelt es sich nicht nur um große IT-Dienstleister, sondern auch um Unternehmen, die im Cloud-Computing-Markt aktiv sind. Große IT-Dienstleister und auch viele Unternehmen aus dem Telekommunikationsbereich wurden im Laufe der letzten Jahre mit eigenen Entwicklungen oder infolge von Übernahmen auf dem Markt für Cloud Computing aktiv. Während Unternehmen wie HP und Dell bald als Public-Cloud-Anbieter aktiv waren, konzentrierten sich andere auf Private-Cloud-Lösungen. Die bestehende Kundenbasis, eine gute Vernetzung in der Branche und der problemlose Zugriff auf einschlägiges Expertenwissen erleich-



terte das Geschäft mit Cloudservices für diese Unternehmen deutlich. Auch große Softwareunternehmen wie Microsoft, SAP und Oracle sind mittlerweile mit ihren Angeboten in der Cloud vertreten.

Cloud-born-Unternehmen: Cloud-born-Unternehmen bieten ausschließlich Cloudservices an, häufig auf der Basis von Clouddiensten anderer Unternehmen. Die Nachfrage nach Cloudservices stieg mit der immer stärkeren Nutzung von mobilen Endgeräten und dem dadurch entstehenden Synchronisationsaufwand an. Ein bekanntes Beispiel ist Dropbox, das 2008 zur Synchronisation von Dateien zwischen verschiedenen Computern und Personen entwickelt wurde. Der Service stieß auf breites Interesse und wurde bald von einem großen Publikum genutzt. Dropbox hat sich mittlerweile mehr und mehr zu einem Kollaborationsservice entwickelt (Barret 2011). Obwohl der Großteil der Nutzer nur den kostenlos bereitgestellten Speicherplatz nutzt, ist Dropbox mittlerweile eines der wertvollsten Start-ups im Silicon Valley. Trotz dieses sehr erfolgreichen Beispiels ist die Entwicklung tragfähiger Geschäftsmodelle für den Cloud-Computing-Markt herausfordernd.

IDC identifizierte IBM und Accenture als weltweit führende Dienstleister am Markt für Cloud Computing (Little et al. 2013, S. 4). Bei Infosys und PwC handelt es sich IDC zufolge um wichtige Anbieter, die es knapp nicht in die Führungsgruppe geschafft haben. IDC schätzte zur Einordnung von 13 Dienstleistern in die vier Kategorien »Leaders«, »Major Players«, »Contenders« und »Participants« deren aktuelle Leistungsfähigkeit sowie deren Strategie zur Verbesserung ihrer Marktposition ein. Unter Leistungsfähigkeit verstehen Little et al. (2013) den Umfang des Dienstleistungsangebots sowie die Abstimmung des Angebots mit den Kundenbedürfnissen. Die Leistungsfähigkeit wirkt sich daher vor allem auf den aktuellen Erfolg eines Dienstleisters aus. Im Gegensatz dazu beeinflusst die Strategie zur Verbesserung der Marktposition den zukünftigen Erfolg eines Dienstleisters (Little et al. 2013, S. 4). Es geht dabei insbesondere um die Berücksichtigung zukünftiger Kundenbedürfnisse und die Unterscheidung von Wettbewerbern. IDC ordnet die meisten Anbieter in die Kategorie »Major Players« ein. Neben den führenden Anbietern IBM und Accenture haben auch HP, Capgemini, Fujitsu und Microsoft relativ große Marktanteile. Während HP, Fujitsu und Microsoft von IDC in die Kategorie »Major Players« eingeordnet wurden, wird Capgemini eher als den »Contenders« zugehörig gesehen.

SITUATION IN DEUTSCHLAND

2.2.2

Die Experton Group hat seit 2010 jährlich Anbieter, die auf dem deutschen Cloud-Computing-Markt aktiv sind, analysiert und bewertet. Im Jahr 2012 wurden für den »Cloud Vendor Benchmark« (Velten/Janata 2012a) 109 Anbieter untersucht, deren Angebote auf dem deutschen Markt erhältlich sind. Die

Anbieter wurden in drei Gruppen eingeteilt: Anbieter von Cloudservices, Anbieter von Technologien für Cloudinfrastrukturen und Dienstleister im Bereich Cloud Computing.

Anbieter von Cloudservices: Im Hinblick auf SaaS-Dienste legten Velten/Janata (2012a) im Rahmen ihrer Studie den Fokus auf CRM-, ERP-, Webconferencing und Kollaborations-/Kommunikationslösungen. Während CRM-, Webconferencing und Kollaborations-/Kommunikationslösungen besondere Anforderungen an Mobilität und Erreichbarkeit stellen und hochgradig standardisiert sind, sind ERP-Lösungen eher für ihren hohen Individualisierungsgrad bekannt. Es ist daher nicht verwunderlich, dass das Angebot an Clouddiensten im ERP-Bereich sehr überschaubar und im CRM- und Webconferencingbereich reichhaltig ist. Auf dem deutschen Markt sind laut Velten/Janata (2012a) im Hinblick auf Portfolioattraktivität und Wettbewerbsstärke Salesforce.com, Microsoft, SAP und Oracle im CRM-Bereich führend. Für den ERP-Bereich werden im »Cloud Vendor Benchmark« SAP, Scopevisio und myfactory hervorgehoben. Das Angebot im Bereich Webconferencing ist sehr breit. Neben dem Angebot der Deutschen Telekom werden insbesondere die Angebote von Cisco, Citrix und Siemens als attraktiv beschrieben. Im Bereich Kollaboration/Kommunikation sind die Angebote von Microsoft (Office 365), Google (Google Apps for Business) und IBM (Lotus Live Collaboration Suite) laut Velten/Janata (2012a) nicht nur im Hinblick auf die Attraktivität des gebotenen Portfolios führend, sondern auch im Hinblick auf die Wettbewerbsstärke.

Anbieter von Technologien für Cloudinfrastrukturen: Zum Aufbau und Betrieb von Cloudinfrastrukturen sind Technologien aus den Bereichen »Cloud Infrastructure« (Server, Speicher, Cloud Appliances), »Cloud Middleware« (Datenbank, Virtualisierung, Betriebssystem) und »Cloud Management« (Quota Management, Fakturierung) notwendig. Zu den wichtigsten Anbietern im Bereich »Cloud Infrastructure« zählen in Deutschland neben Cisco und IBM, die insbesondere im Hinblick auf die Portfolio-Attraktivität führend sind, auch Fujitsu, EMC, HP, Dell und Oracle. Sowohl im Bereich »Cloud Middleware« als auch im Bereich »Cloud Management« nehmen VMware, Microsoft und IBM die Spitzenplätze ein. Im Bereich »Cloud Middleware« werden außerdem die Portfolios von Citrix, RedHat und Suse für besonders attraktiv gehalten. Im Bereich »Cloud Management« heben Velten/Janata (2012a) auch das Angebot von Zimory hervor.

Dienstleister im Bereich Cloud Computing: Auf dem Cloud-Computing-Markt werden auch Dienstleistungen angeboten, die auf die Unterstützung bei Planung, Aufbau und Integration von Cloudinfrastrukturen abzielen. Die Dienstleistungen lassen sich den Bereichen »Beratung« (Planung und Analyse, TCO/ROI-Modelling, Vendor Selection) und »Integration« (Process and Application Integration, Data Migration, Identity Management) zuordnen. Im Bereich »Beratung« heben



Velten/Janata (2012a) die Angebote von Atos und Accenture besonders hervor. Beide sind in der Lage, eine umfassende Cloudberatung anzubieten. Andere Unternehmen mit attraktivem Portfolio wie PwC, CSC oder COMPAREX werden lediglich in Teildisziplinen als stark eingeschätzt. Im Bereich Integration sind laut »Cloud Vendor Benchmark« IBM, Atos und T-Systems führend, knapp gefolgt von Unternehmen wie Accenture, HP, BT Germany und Fujitsu.

FAZIT – KONVERGENZ DER MÄRKTE

3.

Beide Teilmärkte des Marktsegments für Software und IT-Services, der Markt für Big Data und jener für Cloud Computing, entwickeln sich sehr dynamisch. Für beide wird erwartet, dass die durchschnittlichen jährlichen Wachstumsraten zumindest über die nächsten fünf Jahre im zweistelligen Bereich liegen werden. Studien zur Größe des Marktes für cloudbasierte Big-Data-Lösungen sind rar. Die Analysten von TechNavio (2012) gehen von einem durchschnittlichen jährlichen Wachstum von rund 69 % zwischen 2012 und 2016 aus. Den wesentlichen Treiber für dieses rasante Marktwachstum sieht TechNavio im großen Bedarf an kostengünstigen Big-Data-Lösungen.

Auf beiden Märkten sind neben etablierten Anbietern, die versuchen, ihr Geschäftsfeld auszudehnen, auch zahlreiche Start-ups aktiv, die mit dem Ziel, zumindest einen der Märkte zu bedienen, entstanden sind. Zu den Anbietern, die laut TechNavio (2012) den Markt für cloudbasierte Big-Data-Lösungen dominieren, gehören EMC, Microsoft, IBM und Google. Aber auch Anbieter wie Oracle, HP, SAP, Tresata, 1010data, ClickFox, Amazon, Infochimps und LexisNexis sind in diesem Markt aktiv (TechNavio 2012). Eine umfassende Liste von Anbietern auf dem Markt für cloudbasierte Big-Data-Lösungen, die zwischen Data-as-a-Service- und IaaS-Anbietern unterscheidet, wurde von Feinleib (2013) veröffentlicht.

Google und Amazon sind insofern Spezialfälle, als dass beide Unternehmen sowohl im Bereich Big Data als auch im Bereich Cloud Computing als Pioniere gesehen werden können. Google hat mit BigQuery einen Clouddienst im Angebot, der es erlaubt, große Datenmengen mithilfe von SQL-Kommandos zu untersuchen. Der Dienst nutzt die Cloudinfrastruktur von Google. Die Lösung Elastic MapReduce von Amazon ist ebenfalls ein Clouddienst, mit dem große Datenmengen verarbeitet werden können. Zum Einsatz kommt ein Hadoop-Framework, das auf der Cloudinfrastruktur von Amazon ausgeführt wird.

Während klassische Big-Data-Lösungen durchaus auch von deutschen Unternehmen, bzw. von Unternehmen mit deutschen Wurzeln, angeboten werden, handelt es sich bei den Anbietern von cloudbasierten Big-Data-Lösungen fast ausschließlich um US-amerikanische Unternehmen (Landrock 2013b, S. 27).



DIFFUSION UND NUTZUNG

IV.

BIG DATA

1.

Die Ausführungen zur Diffusion und Nutzung von Big Data spiegeln die Ergebnisse einer Reihe von Studien wider, die seit dem Jahr 2011 veröffentlicht wurden (Bange et al. 2013; Interxion 2013; Russom 2011; Schroeck et al. 2012; TCS 2013). Die Studien beschäftigen sich überwiegend mit der Situation in Unternehmen, nur ansatzweise wird zum Teil auch die Situation in der öffentlichen Verwaltung adressiert. Im Hinblick auf die Nutzung von Big Data in der Wissenschaft durch Privatpersonen sind keine Studien verfügbar.

In diesem Teilkapitel werden die Diffusion und Nutzung von Big-Data-Lösungen in Unternehmen, in der öffentlichen Verwaltung, in der Wissenschaft und von Privatpersonen diskutiert. Diffusion und Nutzung unterscheiden sich von Anwendergruppe zu Anwendergruppe zum Teil erheblich. Außerdem sind innerhalb von bestimmten Gruppen auch regionale Unterschiede erkennbar.

UNTERNEHMEN

1.1

Zunächst wird die Nutzung von Big-Data-Analysen in den verschiedenen Regionen und Fachbereichen diskutiert. Anschließend werden die analysierten Daten, die verwendete Plattform und die eingesetzten Werkzeuge und die Zuständigkeiten für Big Data in Unternehmen betrachtet. Abschließend wird kurz die Zufriedenheit von Unternehmen bezüglich der Durchführung von Big-Data-Analysen adressiert.

REGIONEN

1.1.1

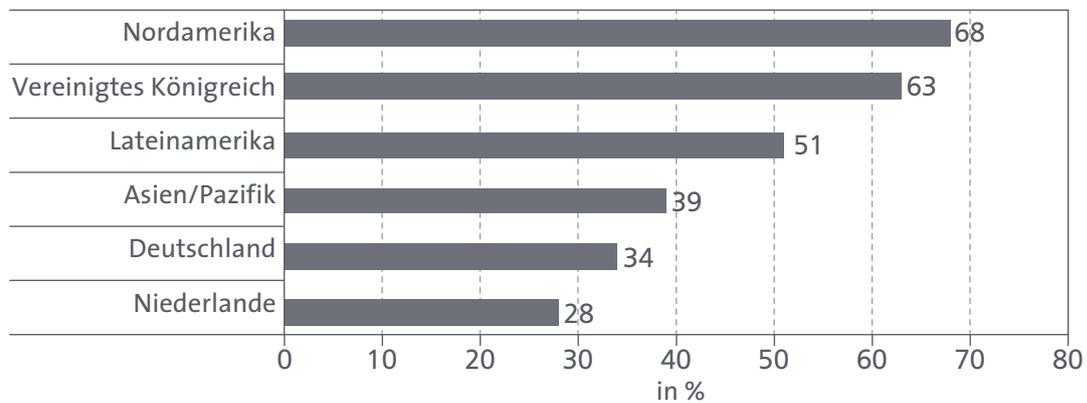
Im Jahr 2012 befragte TCS (2013) 1.217 Vertreter von Unternehmen in den USA, Europa, der Region Asien/Pazifik und Lateinamerika zu ihren Erfahrungen mit Big Data. Insgesamt gaben 643 Teilnehmer (53 %) an, dass ihr jeweiliges Unternehmen große Datenmengen analysiert, um beispielsweise Entscheidungen zu unterstützen oder Geschäftsprozesse zu optimieren. Nur den Vertretern von Unternehmen, die im Bereich Big Data aktiv sind, wurden weitere Fragen zu ihren Erfahrungen gestellt. Die Unternehmen wurden bei der Studie überwiegend von IT-Verantwortlichen (62 %) vertreten, es nahmen aber auch Analysten und Fachbereichsverantwortliche teil. Knapp die Hälfte der befragten Unternehmen (48 %) hatte seinen Hauptsitz in den USA, ein Viertel in Europa.



IV. DIFFUSION UND NUTZUNG

Während 68 % der US-amerikanischen Unternehmen angaben, dass sie im Bereich Big Data aktiv sind, waren es nur 45 % der befragten europäischen Unternehmen und gar nur 39 % aus der Region Asien/Pazifik (TCS 2013). Die Studie von TCS legt die Vermutung nahe, dass US-amerikanische Unternehmen im Hinblick auf Big Data anderen Unternehmen voraus sind. Besonders aktiv sind neben den USA auch Indien und Mexiko. In Europa sind den Ergebnissen der Studie zufolge Unternehmen mit Sitz im Vereinigten Königreich deutlich aktiver als Unternehmen in Deutschland oder den Niederlanden. Im Rahmen der Studie gaben nur 34 % der deutschen Unternehmen an, dass sie im Bereich Big Data aktiv sind (Abb. IV.1).

ABB. IV.1 DURCHFÜHRUNG VON BIG-DATA-ANALYSEN NACH REGIONEN



Quelle: nach TCS 2013, S. 16 f.

In einer ebenfalls 2012 durchgeführten IBM-Studie gaben 28 % der Teilnehmer an, dass sie Pilotprojekte im Bereich Big Data begonnen und erste Lösungen implementiert haben (Schroek et al. 2012). Mit 47 % gab fast die Hälfte der Befragten an, dass Big-Data-Aktivitäten in Planung seien. An der von IBM global durchgeführten Studie nahmen 1.144 Verantwortliche aus den Geschäftsbereichen und der IT teil. Im Jahre 2012 befragte Interxion (2013) 750 europäische IT-Verantwortliche aus Unternehmen mit mehr als 500 Mitarbeitern zu ihren Erfahrungen mit Big Data. Mit 25 % gab genau ein Viertel der Befragten an, sich bereits mit Big Data auseinandergesetzt und ein geeignetes Anwendungsszenario identifiziert zu haben. 22 % der Studienteilnehmer sind noch auf der Suche nach einem geeigneten Anwendungsszenario. In einer von Russom (2011) für TDWI durchgeführten Studie gaben 34 % der Teilnehmer an, dass das von ihnen vertretene Unternehmen große Datenmengen analysiert. 40 % gaben an, dass zwar Daten analysiert werden, aber keine speziellen Big-Data-Werkzeuge zum Einsatz kommen. An der TDWI-Studie nahmen 325 IT-Fachleute, Nutzer aus den Geschäftsbereichen und Berater teil. Sowohl die von IBM durchgeführte Studie als auch die TDWI-Studie hatten einen globalen Fokus.

Die Abweichungen zwischen den Studien sind vermutlich primär auf unterschiedliche Auffassungen im Hinblick auf die Abgrenzung zwischen Datenanalysen im Sinne klassischer BI einerseits und Big-Data-Analysen andererseits zurückzuführen. Relativ klar ist, dass sich Unternehmen intensiv mit dem Thema auseinandersetzen und rund ein Drittel der Unternehmen bereits erste praktische Erfahrungen mit Big Data gesammelt hat. Darüber hinaus gibt es starke regionale Unterschiede im Hinblick auf die Nutzung von Big Data.

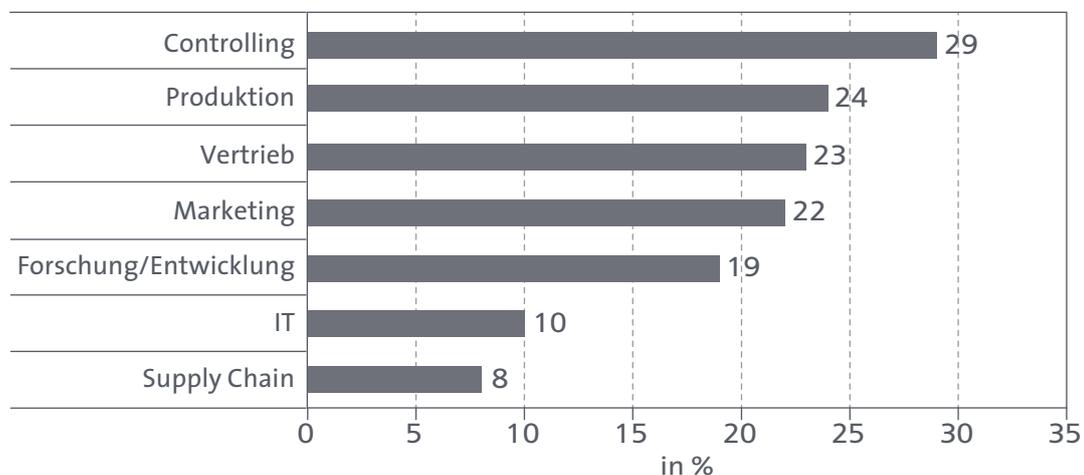
FACHBEREICHE

1.1.2

Im Rahmen einer BARC-Studie wurde unter anderem untersucht, in welchen Fachbereichen Big Data in Europa zum Einsatz kommt (Bange et al. 2013, S.22). An der Studie nahmen 274 Fachbereichs- und IT-Verantwortliche aus Deutschland, Österreich der Schweiz, Frankreich und dem Vereinigten Königreich teil. Die Studienteilnehmer gaben an, dass Big-Data-Analysen vor allem im Bereich Controlling vorgenommen werden. Weitere Bereiche, in denen Big-Data-Analysen relativ häufig durchgeführt werden, sind Marketing, Vertrieb, IT, Produktion sowie Forschung und Entwicklung (Abb. IV.2).

ABB. IV.2

NUTZUNG VON BIG-DATA-ANALYSEN NACH FACHBEREICHEN



Quelle: nach Bange et al. 2013, S.22

Bei der Betrachtung der regionalen Unterschiede fällt auf, dass Big-Data-Analysen im Vereinigten Königreich vor allem in den Bereichen Marketing und Vertrieb häufiger zum Einsatz kommen als in den anderen untersuchten Staaten (Bange et al. 2013, S.24). In Deutschland, Österreich und der Schweiz werden im Bereich Controlling eher Big-Data-Analysen durchgeführt als im Vereinigten Königreich oder in Frankreich. Besonders auffällig ist außerdem, dass Big-Data-Analysen vor allem im Bereich Forschung und Entwicklung in Deutschland, Ös-



terreich und der Schweiz eine deutlich geringere Bedeutung zu haben scheinen als in den anderen beiden Staaten.

TCS (2013) nennt zahlreiche Anwendungsszenarien für Big-Data-Analysen in unterschiedlichen Fachbereichen. Die Szenarien wurden im Rahmen einer umfassenden Studie im Hinblick auf ihr Potenzial bewertet. Die Szenarien sind insbesondere deshalb interessant, da fehlende überzeugende Einsatzszenarien nicht selten als Kernproblem im Kontext von Big Data beschrieben werden (z. B. Bange et al. 2013, S. 6). Bezüglich der Bedeutung von Big-Data-Analysen in den Fachbereichen stimmen die Ergebnisse von TATA Consultancy Services (TCS) weitgehend mit jenen von BARC überein. Im Hinblick auf den Verkauf unterstrichen die TCS-Studienteilnehmer z. B. die Bedeutung von Big-Data-Analysen für die Identifizierung der wertvollsten Kunden, die Identifizierung von Cross-Selling-Potenzialen oder die Bestimmung der optimalen Verkaufstechniken und Verkaufsangebote. Im Bereich Marketing wird die Bedeutung von Big-Data-Analysen beispielsweise für die Bestimmung der Wirksamkeit von Werbekampagnen und Vertriebswegen betont (TCS 2013). Auch im Bereich Controlling werden verschiedene Anwendungsszenarien aufgezeigt.

DATEN

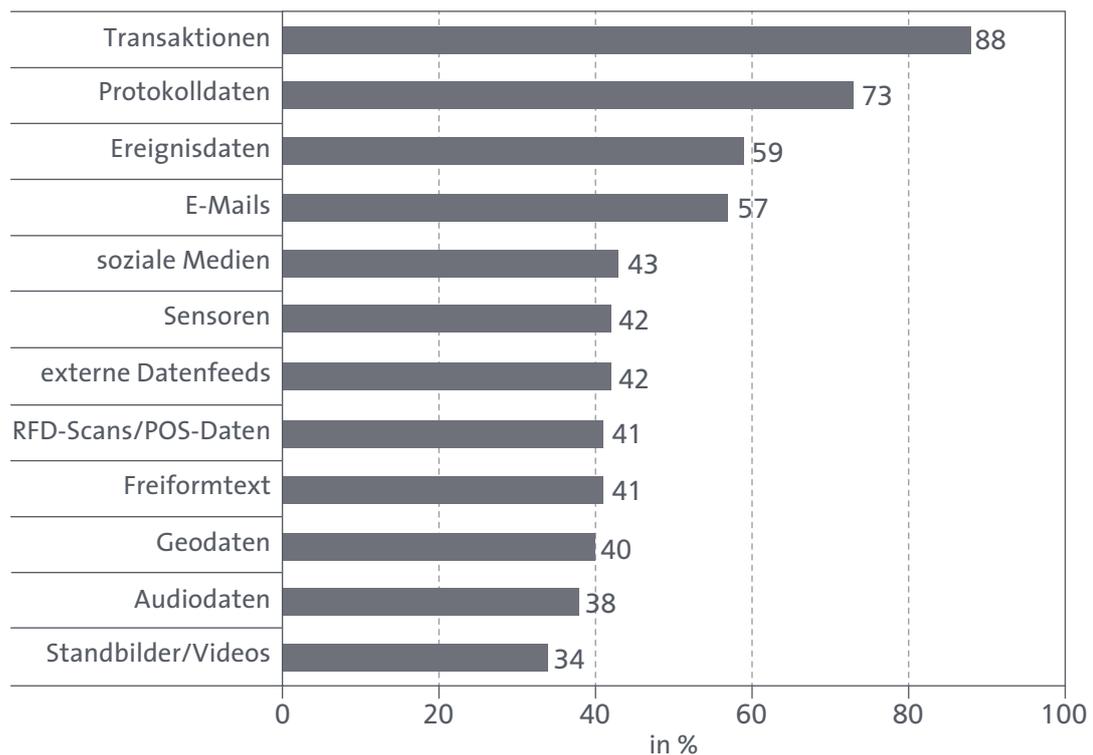
1.1.3

Mit 51 % sind gut die Hälfte der Daten, die in den von TCS befragten Unternehmen analysiert werden können, strukturiert, 27 % unstrukturiert und 21 % teilstrukturiert (TCS 2013). Gut zwei Drittel der Daten fallen intern an und knapp ein Drittel stammt von Quellen außerhalb der Unternehmen. Laut TCS (2013, S. 41) gibt es vor allem im Hinblick auf die Strukturiertheit der zu analysierenden Daten große Unterschiede von Branche zu Branche. Während z. B. im Einzelhandel bis zu zwei Drittel der Daten (67 %) strukturiert sind, ist beispielsweise in der Medien- und Unterhaltungsbranche deutlich mehr als die Hälfte der Daten (58 %) unstrukturiert. Schroeck et al. (2012) fragten in der IBM-Studie ausdrücklich nach den Daten, die für Big-Data-Analysen genutzt werden. In den meisten der befragten Unternehmen werden Transaktionsdaten (88 %), Protokolldaten (73 %), Ereignisdaten (59 %) und E-Mails (57 %) genutzt. Rund 40 % der Unternehmen gaben an, dass sie Daten aus sozialen Medien, Sensordaten, externe Datenfeeds, RFID-Scans/Point-of-Sale-Daten (POS-Daten), Freiformtext, Geodaten oder Audiodaten verwenden. Mit 34 % der Unternehmen analysieren immerhin noch ein gutes Drittel Standbilder und Videos (Schroeck et al. 2012). In Übereinstimmung mit TCS weist IBM darauf hin, dass hauptsächlich interne Datenquellen für Big-Data-Analysen verwendet werden (Abb. IV.3).

Die Ergebnisse der BARC- sowie der TDWI-Studie unterscheiden sich im Hinblick auf die verwendeten Datentypen nicht wesentlich von den Ergebnissen der IBM-Studie (Bange et al. 2013; Russom 2011). Bange et al. (2013) fragten im

Rahmen der BARC-Studie die Teilnehmer auch danach, in welchen Zeitintervallen Daten für die Analyse bereitgestellt werden. Die von den Studienteilnehmern vertretenen Unternehmen stellen nur 4 % ihrer Daten innerhalb von fünf Sekunden bereit und können diese somit in Echtzeit analysieren. Den Ergebnissen der Studie zufolge ist geplant, diesen Wert auf 10 % zu erhöhen. Mit 45 % wird derzeit fast die Hälfte der Daten nur einmal am Tag bereitgestellt. Russom (2011) kommt bei der Untersuchung der Analysehäufigkeit zum Ergebnis, dass tatsächlich derzeit nur 4 % der Daten in Echtzeit analysiert werden. Russom untersuchte darüber hinaus das Volumen der analysierten Daten. 37 % der im Rahmen der TDWI-Studie befragten Unternehmen gaben an, dass sie derzeit Daten im Umfang von 10 bis 100 Terabyte analysieren. Nur 11 % der Unternehmen analysieren schon heute Daten im Umfang von mehr als 100 Terabyte. Den Ergebnissen der Studie zufolge soll der Anteil der Unternehmen in dieser Kategorie in absehbarer Zeit bei über 40 % liegen (Russom 2011).

ABB. IV.3 BEI BIG-DATA-ANALYSEN VERWENDETE DATENTYPEN



Quelle: nach Schroeck et al. 2012

PLATTFORM UND WERKZEUGE

1.1.4

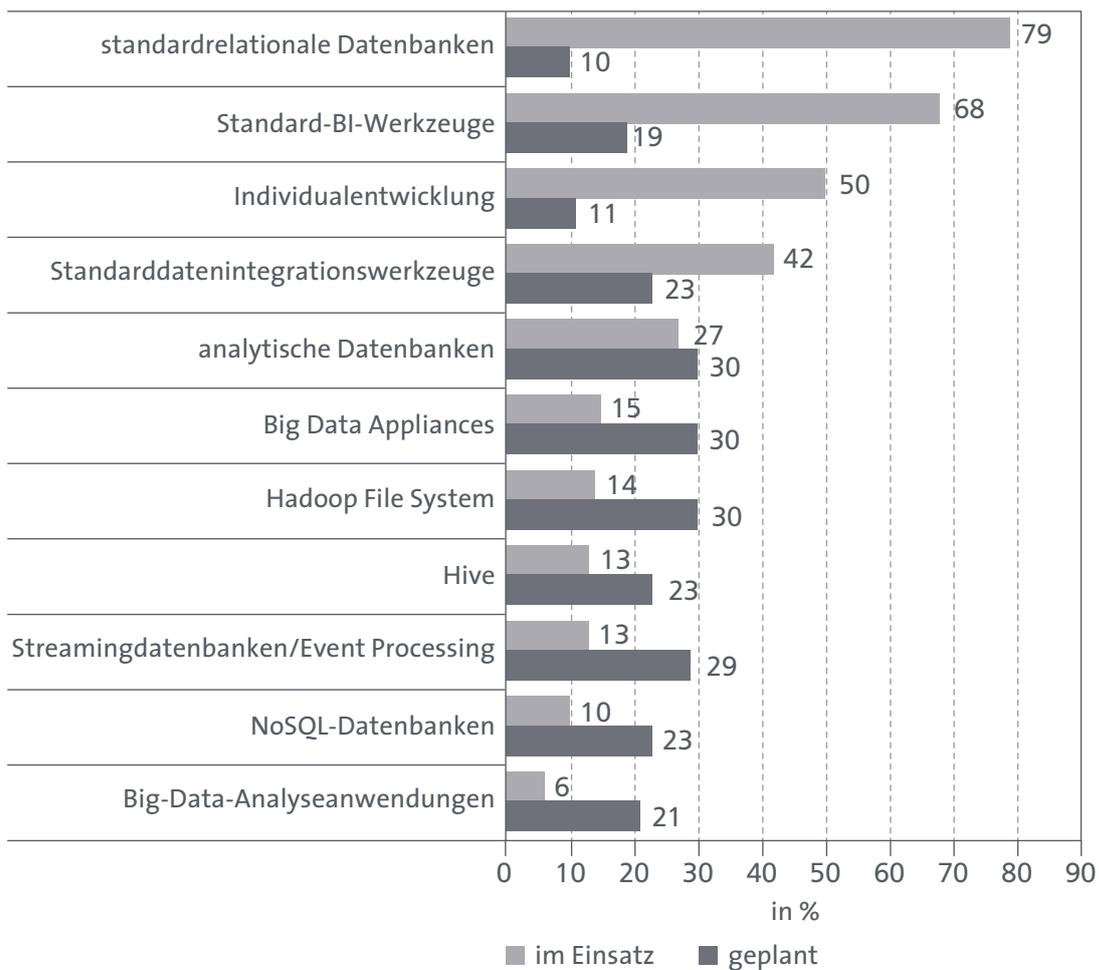
Im Hinblick auf die zur Durchführung von Big-Data-Analysen verwendete Plattform spielen die Fähigkeit zur Datenintegration, eine skalierbare Speicherinfra-



IV. DIFFUSION UND NUTZUNG

struktur, ein leistungsfähiges Data Warehouse sowie Sicherheit und Governance eine besonders wichtige Rolle (Schroek et al. 2012). In der von IBM durchgeführten Studie hielten zwischen 58 und 65 % der Befragten diese Aspekte für zentral. Abfragen und Reporting, Data Mining, Datenvisualisierung, Vorhersagemodelle und Optimierung zählen zu den kritischen Aspekten im Zusammenhang mit Analysewerkzeugen für Big Data. Zwischen 65 und 91 % der Studienteilnehmer maßen diesen Aspekten besondere Bedeutung bei. Bange et al. (2013) fragten nicht nach den zentralen Aspekten, sondern nach den verwendeten Werkzeugen. Während 86 % der befragten Unternehmen Standardwerkzeuge zur Analyse großer Datenmengen verwenden, sind bei 49 % Individualentwicklungen im Einsatz. Spezielle Big-Data-Werkzeuge werden derzeit von immerhin 44 % der befragten Unternehmen eingesetzt. Bei weiteren 36 % ist der Einsatz solcher Werkzeuge geplant (Abb. IV.4).

ABB. IV.4 BEI BIG-DATA-ANALYSEN VERWENDETE TECHNOLOGIEN



Quelle: nach Bange et al. 2013



Bei den speziellen Big-Data-Werkzeugen spielen neben analytischen Datenbanken, vor allem Big-Data-Appliances, HDFS, Hive, Streamingdatenbanken/Event Processing und NoSQL-Datenbanken eine wichtige Rolle. Spezielle Big-Data-Analyse-Anwendungen sind noch kaum verbreitet.

Auch Russom (2011) kommt zum Ergebnis, dass Data Warehouses (64 %) und relationale Datenbanken (38 %) bei der Analyse großer Datenmengen eine wichtige Rolle spielen. Auf die Analyse von großen Datenmengen spezialisierte Datenbanken folgen mit 28 % vor verteilten Dateisystemen wie HDFS/Hadoop mit 24 %. Den Einsatz von cloudbasierten Analyseplattformen bestätigten im Rahmen der Studie nur 12 % der Befragten. Genau in diesem Bereich besteht allerdings den Ergebnissen der TDWI-Studie zufolge sehr großes Interesse. Während ein Rückgang vor allem bei der Verwendung von relationalen Datenbanken erwartet wird, soll die Nutzung von verteilten Dateisystemen und cloudbasierten Analyseplattformen deutlich zunehmen. Mit 30 % würde knapp ein Drittel der Befragten die Nutzung einer cloudbasierten Analyseplattform der aktuellen Lösung vorziehen. 43 % der von Russom im Rahmen der TDWI-Studie befragten Studienteilnehmer gaben an, dass sie ihre Datenanalyse bis 2015 ersetzen wollen. Den Ergebnissen der Studie zufolge haben 9 % bereits eine Analyselösung angeschafft, die den Anforderungen von Big Data gewachsen ist.

ZUSTÄNDIGKEIT

1.1.5

Die Ergebnisse der Studien von TDWI und BARC stimmen im Hinblick auf die Zuständigkeit für Big Data in Unternehmen weitgehend überein. Während 41 % der Teilnehmer an der TDWI-Studie die Zuständigkeit beim Data-Warehouse bzw. BI-Team sehen (Russom 2011), sehen das 47 % der BARC-Studienteilnehmer so (Bange et al. 2013). Neben dem Data-Warehouse- bzw. BI-Team spielen beiden Studien zufolge auch die IT und die einzelnen Geschäftsbereiche im Kontext der Analyse großer Datenmengen eine wichtige Rolle. Während allerdings bei der BARC-Studie mehr Teilnehmer die Zuständigkeit bei der IT sahen (23 zu 16 %), war es bei den Teilnehmern der TDWI-Studie genau umgekehrt (12 zu 21 %). Dieser Unterschied lässt sich vermutlich zum Teil auf regionale Unterschiede zurückführen. Sowohl die TDWI-Studie als auch die Studie von TCS lassen erkennen, dass sich das Verständnis der Zuständigkeiten für die Datenanalyse regional recht stark unterscheiden kann (Russom 2011, S.15; TCS 2013, S.31). Die Ergebnisse der TCS-Studie zeigen darüber hinaus auch, dass die Bedeutung der IT bzw. der einzelnen Geschäftsbereiche stark variiert, je nachdem ob die Zuständigkeit für die Datenanalyse oder die Datenverarbeitung betrachtet wird (TCS 2013). Auch das könnte aufgrund der unterschiedlichen Frageformulierungen zu der deutlichen Abweichung zwischen den Studienergebnissen geführt haben. Im Hinblick auf die Bedeutung des Datenarchitekturteams stimmen die Studien von TDWI und BARC wieder überein (Bange et al. 2013; Russom 2011).



ZUFRIEDENHEIT**1.1.6**

Die Studie von TCS (2013) zeigt, dass Unternehmen, die sich auf Big Data eingelassen haben und große Datenmengen analysieren, überwiegend mit den Ergebnissen zufrieden sind. Über alle untersuchten Regionen hinweg waren 80 % der Studienteilnehmer der Meinung, dass die durchgeführten Big-Data-Analysen zu besseren Entscheidungen geführt haben. Auch im Hinblick auf die Zufriedenheit mit den Ergebnissen von Analysen großer Datenmengen gibt es regionale Unterschiede. Während beispielsweise im Vereinigten Königreich nur 73 % der im Rahmen der TCS-Studie Befragten zufrieden sind, waren es in Deutschland 90 % und in den Niederlanden sogar 91 % (TCS 2013).

ÖFFENTLICHE VERWALTUNG**1.2**

Einerseits führt die öffentliche Verwaltung selbst Big-Data-Analysen durch, andererseits fördert sie die Weiterentwicklung und Nutzung von Big-Data-Analysen durch die Wissenschaft und Unternehmen.

EINSATZBEREICHE**1.2.1**

Bisher wurden noch keine umfassenden Studien über die Nutzung von Big Data in der öffentlichen Verwaltung veröffentlicht. Einzelne Anwendungsbeispiele zeigen allerdings eindrucksvoll, dass die Durchführung von Big-Data-Analysen auch in der öffentlichen Verwaltung Vorteile bringen kann.

Beispielsweise setzte das Königliche Technologie Institut in Stockholm bei einem Versuch zur Verbesserung des Verkehrsmanagements auf Big-Data-Analysen. Das Ziel war die Prognose und Vermeidung von Problemen auf Grundlage von Verkehrs- und Wetterdaten. Neben Unfall- und Staumeldungen sollten unter anderem auch GPD-Daten, Sensordaten und Videos in die Analyse einbezogen und in Echtzeit ausgewertet werden. Die Lösung, die auf IBM InfoSphere Streams aufbaut, wurde in das »intelligente Verkehrssystem« der Stadt Stockholm eingebunden und sorgt nicht nur für kürzere Fahrzeiten, sondern auch für weniger Verkehr und weniger Emissionen (BITKOM 2012, S. 69).

Laut BITKOM (2012, S. 81) arbeitet die öffentliche Verwaltung in Indien in Zusammenarbeit mit Unternehmen wie Wipro Technologies an einer zentralen Datenbank, in der für jeden indischen Staatsbürger neben einer Identifikations-ID unter anderem auch biometrische Daten gespeichert werden sollen. Die Daten sollen de facto in Echtzeit abfragbar sein, um die Identifizierung von Personen im Rahmen von öffentlichen und kommerziellen Geschäftsprozessen nutzen zu können. Die Lösung für die öffentliche Verwaltung in Indien basiert zu einem großen Teil auf Open-Source-Lösungen wie Hadoop.



Das deutsche Unternehmen Empolis Information Management entwickelte in enger Zusammenarbeit mit dem Europäischen Patentamt mit EPOQUE-Net eine Lösung zur weltweiten Patentrecherche. Neben Patentinformationen werden zunehmend weitere Quellen wie wissenschaftliche Publikationen einbezogen, um bereits veröffentlichte Technologien zu identifizieren. Die Datenbank, auf die rund 6.500 Patentprüfer Zugriff haben, hat derzeit 450 Mio. Einträge. Die Lösung ist mittlerweile in 45 Staaten im Einsatz. Dazu zählen neben den Mitgliedstaaten der EU auch Brasilien, Kanada und Australien (BITKOM 2012, S. 89).

Das Schweizer Staatssekretariat SECO entschied sich im Hinblick auf die Durchführung von komplexen Analysen für die Arbeitsmarktstatistik für eine Datenbanklösung, die den Anforderungen großer Datenmengen genügt. Das vorhandene Datenbanksystem wurde durch die EMC Greenplum Database ersetzt, die eine massivparallele Verarbeitung von Daten auf preisgünstigen Standardservern erlaubt. Diese können von den Bundeslieferanten unkompliziert beschafft werden. Im Frontend zeigte sich durch den Einsatz der neuen Lösung eine deutlich bessere Leistung (BITKOM 2012, S. 90).

Auch die Geheimdienste der USA und des Vereinigten Königreichs sorgten kürzlich für viel Aufmerksamkeit, als Details über ihre Aktivitäten im Bereich Datensammlung und -analyse bekannt wurden. Es wurden nicht nur sehr große Datenmengen analysiert, sondern auch unterschiedlich strukturierte Daten aus mehreren Quellen. Darüber hinaus spielte eine zeitnahe Analyse eine wichtige Rolle, um entsprechend auf Entwicklungen reagieren zu können. Big-Data-Analysen werden auch im Gesundheitswesen sowie zur Verbrechensbekämpfung eingesetzt.

Interessante Anwendungsszenarien für Big-Data-Analysen ergeben sich auch durch die zunehmende Verfügbarkeit öffentlicher Verwaltungsdaten (Open Data).

FÖRDERUNG

1.2.2

Mit der »Big Data Research and Development Initiative« beschloss die US-Regierung 2012 ein umfassendes Forschungsprogramm. Insgesamt sollen mehr als 200 Mio. US-Dollar in Forschungsprojekte im Kontext von Big Data fließen. In der EU gibt es derzeit kein direkt vergleichbares Forschungsprogramm. Aktuell setzt sich das im Rahmen des 7. Forschungsrahmenprogramms (FRP 7) der EU geförderte Projekt »BIG – Big Data Public Private Forum«⁵ vor allem mit der Entwicklung einer Strategie für die Big-Data-Forschung und der Förderung des Austausches zwischen Vertretern aus Forschung, Wirtschaft und Politik auseinander. Im Rahmen von Horizont 2020, dessen erste Ausschreibungen im Dezember 2013 veröffentlicht wurden, spielt Big Data eine zentrale Rolle mit rela-

5 www.big-project.eu



tiv großen Förderbudgets. Diese Bedeutung wird ebenfalls durch den Beschluss des Europäischen Rates zur Stärkung der europäischen IT-Industrie unterstrichen, der Big Data neben Cloud Computing als zentrales Gebiet benennt (ER 2013). In Deutschland fördert das Bundesministerium für Bildung und Forschung (BMBF) derzeit im Rahmen des Förderprogramms »IKT 2020 – Forschung für Innovationen« den Aufbau von drei Kompetenzzentren sowie Forschungsvorhaben im Bereich der Verwaltung und Analyse großer Datenmengen. Das Bundesministerium für Wirtschaft und Technologie (BMWi) richtete Ende 2013 die Veranstaltung »Smart Data. Kongress – Workshop – Labor. Intelligente IKT und effiziente Wertschöpfung für Big Data« aus, um Vertreter aus Forschung, Wirtschaft und Politik zusammenzubringen, um eine Big-Data-Strategie zu erarbeiten, die den spezifischen Bedürfnissen und Prioritäten Deutschlands und Europas gerecht wird. Außerdem fördert das BMWi im Rahmen des Technologiewettbewerbs »Smart Data« die Entwicklung von innovativen Big-Data-Lösungen.

WISSENSCHAFT

1.3

Auch wenn Big Data heutzutage vor allem als Herausforderung für Unternehmen sowie staatliche Stellen und Behörden gesehen wird, gehen die Ursprünge von Big Data eigentlich auf den wissenschaftlichen Anwendungsbereich zurück. Neben eher wissenschaftstheoretischen Diskussionen über die Frage, wie viel Informationen überhaupt existieren oder ob sie noch vom Menschen verarbeitet werden können, traten ab den 1990er Jahren zunehmend praktische Fragen der Speicherung, Bearbeitung und Analyse großer Datenmengen in den Vordergrund. Auslöser dafür war die Digitalisierung einzelner Wissenschaftsbereiche. So sammelte der 2000 gestartete Sloan Digital Sky Survey pro Nacht über 200 Gigabyte an digitalen astronomischen Daten (The Economist 2010c). Doch schon vorher tauchten Begriffe wie Large Data Sets oder gar Big Data auf, und wissenschaftliche Artikel beschäftigten sich mit deren Verarbeitung (z. B. Bryson et al. 1999; Cox/Ellsworth 1997). Deutlich wird dabei, dass der Begriff der Größe immer flexibel war und sich im Lauf der Zeit änderte. Ebenso ist zu bedenken, dass diese Formen großer Datenmengen nicht mit der heutigen Definition von Big Data übereinstimmen. Es handelte sich zwar auch oft um Bilddaten oder andere Daten, die sich schlecht in den gängigen Datenhaltungssystemen (insbesondere SQL-Datenbanken) speichern ließen, aber dennoch waren es selten polystrukturierte Daten aus mehreren Quellen. So gesehen überrascht es kaum, dass viele theoretische Grundlagen zu Datenhaltung oder Analyse und insbesondere zur visuellen Analyse im wissenschaftlichen Umfeld entstanden. Die Kommerzialisierung erfolgte dann durch die großen Internetunternehmen, die auch auf einen großen Erfahrungsschatz in der Sicherheitsforschung zurückgreifen konnten.



Dennoch hat auch in der Wissenschaft die Digitalisierung weiter Einzug gehalten, und neue Analysemethoden und -modelle führten zu immer neuem Bedarf an fortschrittlicheren Lösungen. Diese heute auch als »eSciences« bezeichnete Entwicklung löste daher viele andere Entwicklungen aus, wie beispielsweise das Grid Computing, welches zu den direkten Vorläufern von Cloud Computing gehört. Deshalb ist es auch nicht weiter überraschend, dass der Wert von wissenschaftlichen Daten zunehmend in den Fokus geriet, insbesondere als es möglich wurde, diese Daten miteinander oder mit Datensätzen aus anderen Quellen zu kombinieren. Dementsprechend überrascht es nicht, dass bereits im September 2008 die Fachzeitschrift »Nature« eine Sonderausgabe zum Thema Big Data in der Wissenschaft veröffentlichte.

Ein Beispiel für diese neuen Möglichkeiten ist die Untersuchung eines Konsortiums rund um die Harvard School of Public Health über den Zusammenhang zwischen der Mobilität von kenianischen Staatsbürgern und der Verbreitung von Malaria mithilfe der Analyse großer Datenmengen. Dazu werteten die Forscher über mehrere Jahre gesammelte Mobilfunkdaten von knapp 15 Mio. Kenianern sowie detaillierte Informationen über das regionale Auftreten von Krankheitsfällen aus. Die Datenanalysen legten den Schluss nahe, dass die Krankheit ihren Ursprung in der Region rund um den Victoriasee hat und sich primär ostwärts in Richtung Nairobi ausbreitet. Die Analysen erlauben Rückschlüsse, die für die Vermeidung von Malariainfektionen von großer Bedeutung sind (Harvard School of Public Health 2012). Gleichzeitig warf das Projekt ethische Fragen auf, da beispielsweise die Auswertung der Mobilfunkdaten ohne Zustimmung der Nutzer erfolgte. Ein besonders häufig genanntes Beispiel für die Durchführung von Analysen extrem großer Datenmengen im Dienste der Wissenschaft ist der bereits erwähnte LHC am CERN (The Economist 2010b). Neben der Epidemiologie und der Kernphysik spielt Big Data aber auch in anderen Disziplinen wie Astronomie, Geologie, Genetik oder Klimaforschung eine zunehmend wichtige Rolle (z. B. Andersen 2012; Sellars et al. 2013). Hierbei treten aber teilweise auch zunehmend wissenschaftstheoretische und ethische Fragen auf, die noch adressiert werden müssen (The Royal Society 2012).

Die zunehmende Bedeutung von Daten in der Wissenschaft wurde auch auf politischer Ebene erkannt, wie der Bericht der Expertenkommission der EU zu wissenschaftlichen Daten (EK 2010) sowie die darauf basierende Erklärungen der Europäischen Kommission zeigen (EK 2009). Dementsprechend wurden auf europäischer Ebene mit dem EUDAT-Projekt erste Schritte in die Wege geleitet, eine multidisziplinäre Plattform zur Datenspeicherung zu schaffen, an der auch deutsche Forschungseinrichtungen beteiligt sind (EUDAT). Darüber hinaus werden im Rahmen des Programms »Infrastructure« sowie des Programms »Science in and with Society« von Horizont 2020 sowohl weitere technische Projekte als auch Projekte zu ethischen und forschungspolitischen Fragestellungen im Kon-



text von Responsible Research and Innovation (RRI) adressiert (EK 2013b u. 2013c). Auch in Deutschland hat das Förderprogramm des BMBF zu Forschungsvorhaben in Big Data sowohl wissenschaftliche Forschungsprojekte als auch Begleitforschungen adressiert.

PRIVATPERSONEN

1.4

Privatpersonen führen in der Regel selbst keine Big-Data-Analysen durch. Sollte es doch einmal der Fall sein, bieten Big-Data-Lösungen auf Basis von Cloud Computing ein großes Potenzial, da keine Infrastruktur aufgebaut werden muss, sich die erforderlichen Kenntnisse auf die Datenauswahl, -analyse und -interpretation beschränken und die Kosten überschaubar bleiben. Einige Anbieter von cloudbasierten Big-Data-Lösungen erlauben Privatpersonen die kostenlose Nutzung ihrer Infrastruktur in begrenztem Umfang. Google verfolgt beispielsweise ein entsprechendes Modell im Zusammenhang mit BigQuery.

Ein Anwendungsgebiet, das sich zunehmender Beliebtheit erfreut, ist im Umfeld der Quantified-Self-Bewegung zu finden. Immer mehr Menschen finden daran Gefallen, unterschiedlichste persönliche Daten auszuwerten, um ihre Persönlichkeitsentwicklung und Gesundheit zu beobachten und in der Folge fördern zu können. Vor allem mithilfe von Funkschnittstellen und Smartphone-Apps lassen sich Daten von zahlreichen Geräten aus dem medizinischen Bereich aufzeichnen. Aufgezeichnet werden nicht nur Daten zu Gewicht, Blutdruck, Herzratenvariabilität oder Atemfrequenz, sondern immer mehr auch Daten zu Stress, Entspannung oder psychischem Wohlbefinden (Liebram 2013). Besonders große Datenmengen kommen in diesem Zusammenhang nicht nur durch die kontinuierliche Erfassung zustande, sondern auch dadurch, dass Anhänger der Quantified-Self-Bewegung ihre eigenen Werte mit den Werten von Menschen derselben Altersklasse oder desselben Fitnesslevels vergleichen möchten. Aktuell werden im Kontext der Bewegung fast ausschließlich strukturierte Daten ausgewertet.

Auch wenn Privatpersonen nur in Ausnahmefällen Analysen durchführen, profitieren sie doch in vielerlei Hinsicht von Big-Data-Analysen, die von Unternehmen, der öffentlichen Verwaltung oder der Wissenschaft durchgeführt werden. Vor allem im Hinblick auf die Wissenschaft ist das Verhältnis allerdings keineswegs einseitiges. Im Rahmen der Citizen Science sind Privatpersonen aufgerufen, sich in wissenschaftliche Prozesse einzubringen (Haarmann 2013). In vielen Projekten handelt es sich vor allem um eine Unterstützung von Wissenschaftlern bei der Erhebung von Daten. Dieses Vorgehen ermöglicht das rasche Erstellen umfangreicher Datensätze mit überschaubaren Kosten. Privatpersonen können sich aber auch als Mitglieder von Beratungsgremien an Forschungsprojekten beteiligen oder private Rechenleistung für wissenschaftliche Projekte zur Verfügung stellen. Daten, die von Stellen und Behörden der öffentlichen Verwal-

tung erhoben wurden, stehen mittlerweile in vielen Fällen auch Privatpersonen in Form von Open Data zur Verfügung.

CLOUD COMPUTING

2.

Es werden die Diffusion und Nutzung von Clouddiensten in Unternehmen, in der öffentlichen Verwaltung, in der Wissenschaft und von Privatpersonen diskutiert. Diffusion und Nutzung unterscheiden sich über die Anwendergruppen zum Teil erheblich. Außerdem sind innerhalb von bestimmten Gruppen auch regionale Unterschiede erkennbar. Anwender erwarten sich von Clouddiensten unter anderem höhere Flexibilität und sinkende Kosten.

UNTERNEHMEN

2.1

Im Rahmen einer IDC-Studie wurden von Cattaneo et al. (2012) Vertreter von Unternehmen und Einrichtungen der öffentlichen Verwaltung im Hinblick auf die Nutzung von Cloudangeboten befragt. Von den 1.056 Teilnehmern aus neun EU-Mitgliedstaaten gaben 45 % an, dass sie zumindest in einem Bereich Services aus der Cloud nutzen. 32 % der Unternehmen nutzen Cloudangebote in mehr als einem Bereich (Cattaneo et al. 2012, S. 16). Einer Studie des Cloud Industry Forum (2012b, S. 5) mit 400 Teilnehmern zufolge nutzen bereits 76 % der Unternehmen und Einrichtungen der öffentlichen Verwaltung in den USA zumindest in einem Bereich Dienste aus der Cloud. Befragt wurden in beiden Studien leitende Entscheidungsträger. Auch wenn die Ergebnisse der beiden Studien aufgrund unterschiedlicher methodischer Ansätze nicht ohne Weiteres vergleichbar sind, deuten sie doch auf eine stärkere Nutzung von Cloudangeboten in den USA hin.

Aus den Ergebnissen der Studie von Cattaneo et al. (2012) geht klar hervor, dass größere Unternehmen in Europa eher Cloudangebote nutzen als kleinere. In den USA greifen laut Cloud Industry Forum (2012b) im Gegensatz dazu neben großen Organisationen auch sehr kleine Organisationen auf Cloudangebote zurück. Den Ergebnissen der Studie zufolge greifen Organisationen mit 100 bis 999 Mitarbeitern in den USA seltener auf Cloudangebote zurück als andere. Für Europa konnten keine Auffälligkeiten im Hinblick auf die Nutzung von Cloudangeboten in unterschiedlichen Branchen festgestellt werden (Cattaneo et al. 2012). Ein Vergleich mit der Situation in den USA ist aufgrund fehlender Daten nicht möglich.

Cattaneo et al. (2012, S. 14) zufolge werden in Europa vor allem E-Mail- und Security-Dienste aus der Cloud genutzt. Überraschend ist, dass Back-Office-Dienste bereits an dritter Stelle liegen. Back-Office-Dienste umfassen eine breite Palette an Diensten von Beschaffungslösungen über Lösungen für das Rechnungs-



wesen bis hin zu umfassenden ERP-Lösungen und haben normalerweise einen hohen Individualisierungsgrad. Der Grund für die Platzierung liegt wahrscheinlich in der sehr breit angelegten Kategorie Back-Office-Dienste. Datenbank- und Speicherdienste werden ebenfalls häufig aus der Cloud bezogen. Überraschend ist auch der hohe Wert für BI-Dienste. Es fällt auf, dass vor allem einfache Dienste vermehrt im Rahmen von Cloudangeboten genutzt werden. Im Hinblick auf die verwendeten Dienste ist die Situation in den USA ähnlich wie in Europa (Cloud Industry Forum 2012b, S. 9). Spiceworks (2013, S. 9) befragte im Rahmen einer Studie weltweit 1.000 IT-Spezialisten und kam zu ähnlichen Ergebnissen.

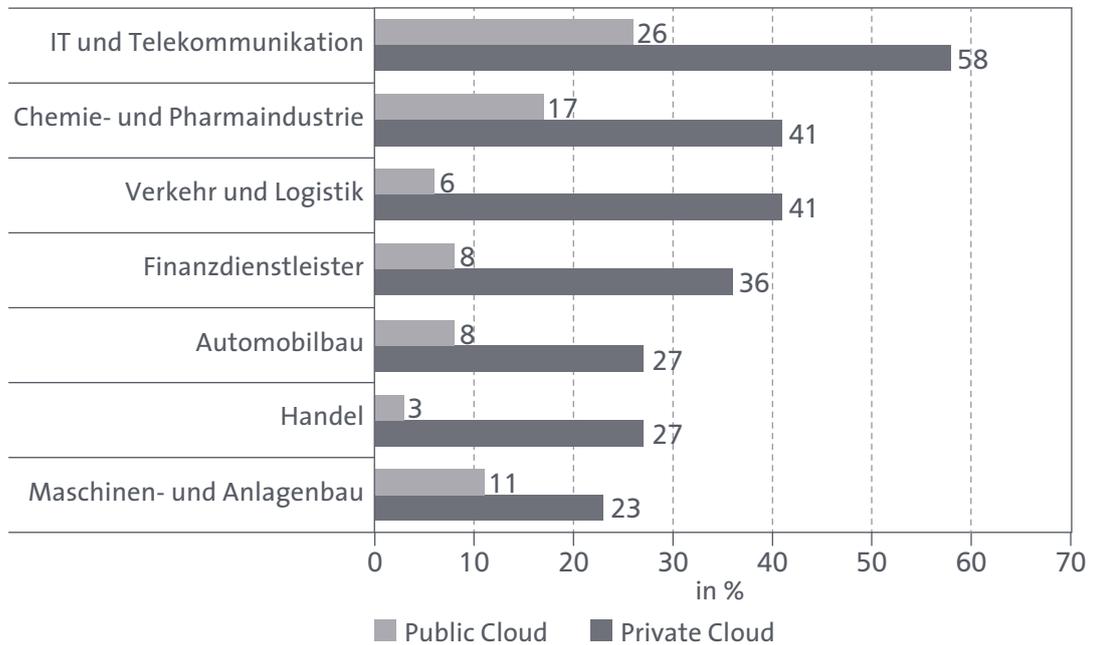
Es fällt auf, dass vor allem KMU in den USA bei der Nutzung von Cloudangeboten deutlich weniger zögerlich sind als entsprechende Unternehmen in Europa. Da KMU einen wesentlichen Teil der europäischen Unternehmenslandschaft ausmachen, kann der kleinere Markt für Cloud Computing in Europa teilweise auf dieses Zögern zurückgeführt werden. Gartner (2012a) geht davon aus, dass der Rückstand hinter den USA zumindest zwei Jahre ausmacht. Andere – wie z. B. Borja (2012) – gehen davon aus, dass der Abstand noch größer ist.

Nach den Ergebnissen einer Umfrage, die von KPMG und BITKOM durchgeführt wurde, nutzten Ende des Jahres 2012 rund 37 % aller Unternehmen in Deutschland Cloud Computing, während der entsprechende Wert im Vorjahr nur bei 28 % lag (KPMG/BITKOM 2013). Bei Großunternehmen ab 2.000 Mitarbeitern liegt der Wert mit 65 % deutlich höher als bei kleineren Unternehmen. Während 34 % der deutschen Unternehmen auf Private-Cloud-Angebote setzten, entscheiden sich nur 10 % für die Public Cloud. Der Anteil der Unternehmen, für die Cloud Computing nicht in Betracht kommt, ist 2012 gegenüber dem Vorjahr in allen Größenklassen deutlich gesunken. Vor allem sehr große Unternehmen kommen dem »Cloud-Monitor 2013« zufolge nur sehr schwer um das Thema herum. Während sich 2011 noch 50 % der deutschen Unternehmen nicht mit Cloud Computing beschäftigten, waren es 2012 nur noch 34 %. IDC (2012a) weist im Rahmen des Berichts zur Studie »Cloud Computing in Deutschland 2012« darauf hin, dass in 56 % der befragten Unternehmen Cloud Computing formell Bestandteil der IT-Strategie ist. Dieses Ergebnis zeigt deutlich, welchen Stellenwert Cloud Computing mittlerweile in deutschen Unternehmen einnimmt.

Im Branchenvergleich liegt in Deutschland sowohl bei der Nutzung von Private-Cloud-Angeboten als auch bei der Nutzung von Public-Cloud-Angeboten der Bereich IT und Telekommunikation mit 58 % bzw. 26 % an der Spitze. Mit je 41 % werden Private-Cloud-Angebote auch in der Chemie- und Pharmaindustrie und im Bereich Verkehr und Logistik häufig genutzt. Mit 36 % setzt auch gut ein Drittel der Finanzdienstleister auf Private-Cloud-Angebote. In der Chemie- und Pharmaindustrie nutzen immerhin 17 % der Unternehmen in Deutschland Public-Cloud-Angebote (Abb. IV.5).

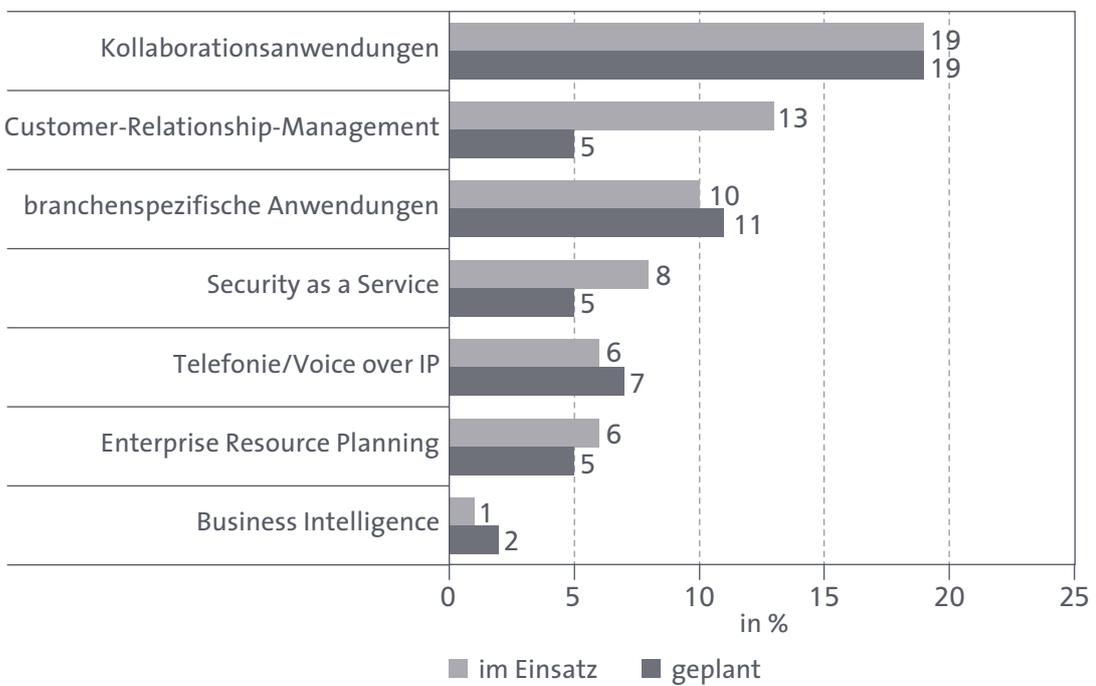


ABB. IV.5 NUTZUNG VON CLOUDDIENSTEN NACH BRANCHE



Quelle: nach KPMG/BITKOM 2013

ABB. IV.6 NUTZUNG VON SAAS-DIENSTEN IM RAHMEN VON PUBLIC-CLOUD-ANGEBOTEN



Quelle: nach KPMG/BITKOM 2013



Kollaborations-/Kommunikations- und CRM-Lösungen sind derzeit laut »Cloud-Monitor 2013« die wichtigsten Anwendungsfelder für SaaS-Services aus der Public Cloud. Die Vorteile von Cloud Computing wie verteilter Zugriff von verschiedenen Standorten oder die Nutzung unterschiedlicher Endgeräte kommen hier besonders stark zum Tragen. Von den Unternehmen, die Public-Cloud-Angebote nutzen, haben 19 % Kollaborations-/Kommunikationslösungen und 13 % CRM-Lösungen im Einsatz. Laut KPMG/BITKOM (2013, S.28) gehört der Zugriff auf Unternehmensdaten und -anwendungen zu den wichtigsten Motiven für die Nutzung von Cloudangeboten. ERP-Lösungen aus der Public Cloud werden nur von 6 % eingesetzt. Mit 1 % spielen BI-Lösungen in Deutschland noch eine untergeordnete Rolle (Abb. IV.6).

ÖFFENTLICHE VERWALTUNG

2.2

Um die Diffusion und Nutzung von Cloud Computing insgesamt zu fördern, schlägt KPMG (2012) vor, den Einsatz von Cloud Computing in der öffentlichen Verwaltung zu forcieren.

Einerseits nutzt die öffentliche Verwaltung selbst Clouddienste, andererseits fördert sie die Weiterentwicklung und Nutzung von Cloud Computing durch die Wissenschaft bzw. Unternehmen.

EINSATZBEREICHE

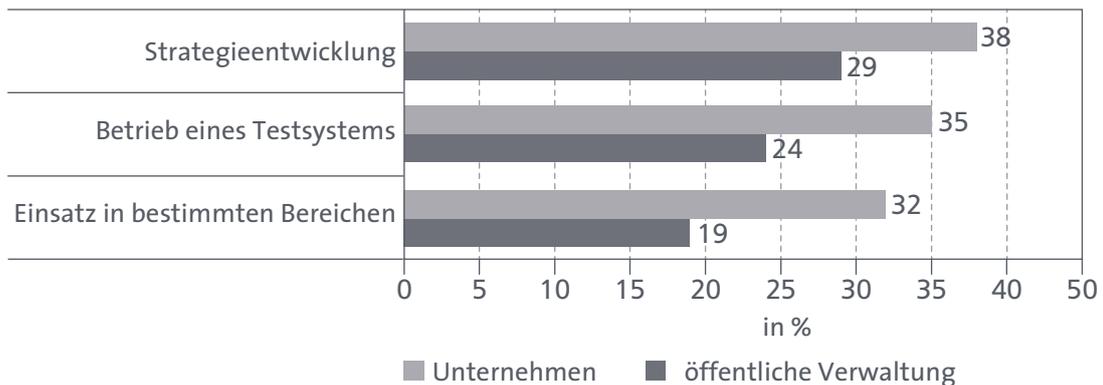
2.2.1

Die beiden bereits erwähnten Studien von Cattaneo et al. (2012) und vom Cloud Industry Forum (2012b) berücksichtigten neben Unternehmen auch Einrichtungen der öffentlichen Verwaltung. In Europa nutzen rund 43 % der Studienteilnehmer aus dem Bereich Regierung (Government) Services, die im Rahmen von Cloudangeboten bereitgestellt werden. Aus den Bereichen Gesundheitswesen und Bildung (Healthcare/Education) gaben 41 % der Teilnehmer an, zumindest in einem Bereich entsprechende Services zu nutzen (Cattaneo et al. 2012, S. 19). In den USA nutzen 63 % der Studienteilnehmer aus dem öffentlichen Sektor zumindest einen Dienst aus der Cloud (Cloud Industry Forum 2012b, S. 5).

Eine Studie von KPMG (2012) untersuchte die Nutzung von Cloud Computing in zehn Staaten und stellte zahlreiche Unterschiede zwischen den verfolgten Ansätzen fest. Die Studie konzentrierte sich vor allem auf die verwendeten Kategorien von Services und die Bereiche, in denen Clouddienste zum Einsatz kommen. An der 2011 durchgeführten Studie nahmen 429 Regierungsvertreter und 808 Führungskräfte von Unternehmen teil. Zusätzlich wurden mehrere Regierungschefs im Rahmen von Interviews befragt. Den Ergebnissen der Studie zufolge liegen staatliche Stellen und Behörden im Hinblick auf den Einsatz von Cloud Computing deutlich hinter Unternehmen zurück (Abb. IV.7). Die Ergeb-

nisse lassen allerdings ein zunehmendes Interesse der öffentlichen Verwaltung an Clouddiensten erkennen.

ABB. IV.7 NUTZUNG VON CLOUD COMPUTING IN UNTERNEHMEN UND DER ÖFFENTLICHEN VERWALTUNG



Quelle: nach KPMG/BITKOM 2013

Mit 32 % gab knapp ein Drittel der Befragten an, eine Private Cloud zu nutzen bzw. dies zu planen. 26 % der Regierungsvertreter gaben an, dass die staatliche Stelle oder Behörde, die sie vertreten, eine Hybrid-Cloud-Umgebung nutzt oder zu nutzen plant. Weniger als ein Viertel der Befragten (22 %) gab an, dass die Nutzung von Diensten aus einer Public Cloud favorisiert werde. Genutzt werden in der öffentlichen Verwaltung überwiegend SaaS-Dienste. Während 38 % der befragten Regierungsvertreter bereit sind, in SaaS-Dienste zu investieren, können sich das nur 28 % bei PaaS- und 26 % bei IaaS-Diensten vorstellen.

Auf Ebene der Mitgliedstaaten unterscheiden sich die Ansätze zur Nutzung von Cloud Computing von Staat zu Staat deutlich. Von den zehn im Rahmen der KPMG-Studie untersuchten Staaten sind fünf Mitgliedstaaten der EU (Dänemark, die Niederlande, das Vereinigte Königreich, Italien und Spanien). Während vor allem in Dänemark und Italien die Möglichkeiten des Cloud Computing umfassend genutzt werden, ist das in anderen Staaten weniger der Fall. Mögliche Einflussfaktoren könnten laut KPMG (2012, S. 21 ff.) unter anderem die Größe des Staates, der Zentralisierungsgrad (zentrale vs. föderale Strukturen) oder das Interesse an Kosteneinsparungen sein.

In Dänemark werden die Daten der öffentlichen Verwaltung bereits seit geraumer Zeit zentral gespeichert (CPR 2009; Willumsen 1999). Dokumente der dänischen Agentur für Digitalisierung zeigen, dass sich zahlreiche Initiativen auf verschiedenen Ebenen mit Cloud Computing beschäftigen. Beispielsweise soll die zentrale E-Invoicing-Plattform »NemHandel« in Zukunft durch einen Cloudservice realisiert werden. Im Hinblick auf Dänemark sind keine Pläne zum Aufbau einer nationalen Cloudinfrastruktur bekannt.



Im Gegensatz dazu realisierten die Regierungen von Frankreich und dem Vereinigten Königreich mit »Andromède« bzw. »G-Cloud« nationale Cloudumgebungen, die als Plattform für den gesamten öffentlichen Sektor dienen sollen. Das primäre Ziel der Initiativen war die Reduktion von Kosten. Im Gegensatz zur Cloudumgebung des Vereinigten Königreichs soll die französische Cloudumgebung nicht nur von Einrichtungen der öffentlichen Verwaltung, sondern auch von Unternehmen genutzt werden.

Im Hinblick auf die Verwendung von Cloud Computing im öffentlichen Sektor in Europa fallen im Vergleich zu den USA drei wesentliche Unterschiede auf. Die Nutzung von Cloud Computing wird in Europa auf nationaler Ebene vorangetrieben und unterscheidet sich stark von einem Staat zum anderen. Die meisten Initiativen in Europa sind noch in einem sehr frühen Stadium. Die nationalen Strategien zielen häufig auf große nationale Cloudumgebungen ab, die nicht nur vom öffentlichen Sektor, sondern auch von Geschäftskunden genutzt werden sollen. Diese Cloudumgebungen werden häufig als kritisch für die Wettbewerbsfähigkeit des nationalen IT-Sektors angesehen. Obwohl dieser Aspekt nicht offiziell Teil der Strategie der USA ist, wird es wahrscheinlich zumindest als eine Nebenerscheinung angesehen (Higgins 2012).

Auch Japan baute mit der »Kasumigaseki Cloud« eine große nationale Cloudumgebung auf (Wyld 2010). In den USA haben einige Initiativen bereits einen sehr hohen Reifegrad erreicht. Während in den USA auf bereits existierende Cloudinfrastrukturen zurückgegriffen wird, bevorzugten einige Staaten Europas den Aufbau eigener Infrastrukturen, überwiegend in Zusammenarbeit mit nationalen IT-Dienstleistern. Aufgrund der Tatsache, dass der Aufbau einer neuen Cloudinfrastruktur viel Zeit in Anspruch nimmt und um von niedrigeren Kosten zu profitieren, haben mittlerweile die meisten europäischen Staaten ihre bestehenden Rechenzentren durch Virtualisierung in Private Clouds umgewandelt.

In den letzten Jahren wurde auch in Deutschland über den Aufbau einer nationalen Cloudumgebung für die öffentliche Verwaltung und Unternehmen nach dem Vorbild Frankreichs diskutiert (Kalenda/Pößneck 2011). Während entsprechende Ideen zunächst wieder verworfen wurden, tauchen sie im neuen Koalitionsvertrag wieder auf, wobei hier allgemeiner von einer »einheitlichen Plattform ›Netze des Bundes« gesprochen wird (CDU/CSU/SPD 2013). In Deutschland gibt es auch auf Landesebene zahlreiche Initiativen zur Nutzung von Cloud Computing. Allerdings sind viele, wie z. B. die im Rechenzentrum Region Stuttgart, auf die Reduktion von Kosten und die Verbesserung der Dienstgüte durch die Nutzung einer Private Cloud beschränkt (Kommune21 2012).

FÖRDERUNG**2.2.1**

Die Fördersituation in Deutschland ist vielfältig ausgeprägt. Auf Bundesebene betreibt das BMWi das Technologieprogramm »Trusted Cloud«. Das Programm zielt auf die Entwicklung innovativer, sicherer und rechtskonformer Cloudlösungen ab. Das BMWi legte den Schwerpunkt des Programms auf Lösungen, die sich für den Einsatz in KMU eignen und fördert Projekte zur Entwicklung von Basistechnologien genauso wie Projekte, die sich mit der Anwendung von Cloud Computing in Industrie und Handwerk, im Gesundheitssektor und in der öffentlichen Verwaltung beschäftigen. Im Rahmen der Hightech-Strategie 2020 der Bundesregierung wurde das Zukunftsprojekt »Internetbasierte Dienste für die Wirtschaft« initiiert, das auf den Ergebnissen des vom BMWi geförderten Projekts »THESEUS – Neue Technologien für das Internet der Dienste« aufbauen soll. Sowohl das Technologieprogramm »Trusted Cloud« als auch das Zukunftsprojekt »Internetbasierte Dienste für die Wirtschaft« sind Teile des Aktionsprogramms »Cloud Computing«. Auf EU-Ebene spielt Cloud Computing ähnlich wie Big Data eine wichtige Rolle im Rahmen von Horizont 2020. Beide stehen auch im Zentrum der vom Europäischen Rat im Herbst 2013 verabschiedeten Strategie zur Stärkung der europäischen IT-Industrie. Darüber hinaus wurde Mitte 2013 das von der EU mit 10 Mio. Euro mitfinanzierte FRP 7-Projekt »Cloud for Europe«⁶ mit Teilnehmern aus 22 Mitgliedstaaten gestartet. Im Rahmen des Projekts sollen innovative Lösungen für die Nutzung von Cloud Computing in der öffentlichen Verwaltung durch die Nutzung des Instruments der innovativen, vorkommerziellen Beschaffung entwickelt und umgesetzt werden. Darüber hinaus soll somit auch sichergestellt werden, dass Lösungen in verschiedenen Mitgliedstaaten miteinander kompatibel sind.

WISSENSCHAFT**2.3**

Cloud Computing spielt in der Wissenschaft vor allem dann eine Rolle, wenn hohe Rechenleistung gefordert ist oder bestimmte Dienste oder Daten einem verteilten Nutzerkreis zur Verfügung stehen sollen. Grundsätzlich neu ist das Thema in der Wissenschaft nicht, denn gerade wissenschaftliche Einrichtungen waren die Treiber der Entwicklung im Distributed Computing und Grid Computing der späten 1990er und 2000er Jahre, die, wie bereits erwähnt, eng mit Cloud Computing verwandt sind. Auslöser dafür war der große Bedarf nach Rechenleistung, der in einzelnen Disziplinen mit dem verstärkten Einsatz von computerbasierten Modellen und Simulationen oder anderen computerbasierten Verfahren entstand. Diese zunehmende Digitalisierung der wissenschaftlichen Arbeit, die mit dem Begriff »eSciences« umschrieben wird, hat auch letztlich da-

6 www.cloudforeurope.eu



zu geführt, dass sich die Anforderungen und der Bedarf an Rechenleistung gewandelt haben (z. B. EK 2009). Neben den eher einfachen Services, die über die jeweiligen Hochschulrechenzentren angeboten wurden, entstand der Bedarf nach flexiblerem Zugang zu Rechenleistung jenseits der lokalen Infrastrukturen.

Als Antwort darauf wurden die Gridaktivitäten massiv ausgebaut. Flaggschiff dabei war die sich seit Mitte der 2000er Jahre formierende nationale Gridinitiative für Deutschland (NGI-DE), die vom BMBF finanziert wird. Ziel war der Aufbau des D-Grid für die Wissenschaft. Gleichzeitig ist NGI-DE, das 2010 institutionalisiert und bei der Gauß-Allianz angesiedelt wurde, deutscher Vertreter in der European Grid Initiative (EGI). Diese zielt darauf ab, europaweit einheitliche und zugängliche Plattformen für alle Wissenschaftsdisziplinen anzubieten und wird nicht nur durch nationale Beiträge, sondern auch von der Europäischen Kommission gefördert. Neben Gridstrukturen werden im Rahmen der EGI auch Cloudinfrastrukturen angeboten, wobei sich hier aber vieles noch in der Entwicklung befindet, wie gerade laufende Forschungsprojekte im Rahmen von FRP 7 zeigen. Weitere Aktivitäten in diesem Rahmen zielen darauf ab, auch Hochleistungsrechnerinfrastrukturen, die im Rahmen der PRACE (Partnership for Advanced Computing in Europe) zusammengeführt wurden, ebenfalls einzubinden. Insgesamt sind an EGI zurzeit über 300 Einrichtungen aus über 50 Ländern beteiligt, die über 300.000 Rechnerkerne und 400 Petabyte Speicherplatz anbieten, auf denen monatlich über 40 Mio. Jobs durchgeführt werden (EGI 2013). Das Anwendungsspektrum für solche Grid-/Clouddienste ist dabei sehr breit. Neben dem LHC am CERN benutzen verschiedene astronomische Einrichtungen diese Infrastrukturen zur Analyse ihrer Ergebnisse. Ebenso kommt es in Genom-Entschlüsselungsprogrammen als auch bei der Rekonstruktion in der Archäologie zum Einsatz (Gater 2011).

Neben diesen übergeordneten Initiativen führen in Deutschland Wissenschaftsorganisationen und Universitäten teilweise eigene Cloudinfrastrukturen ein. So bietet die Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen der Max-Planck-Gesellschaft Cloudlösungen für die angeschlossenen Einrichtungen an, die der flexiblen Bereitstellung von virtuellen Servern und damit Rechenleistung sowie dem Datenaustausch und der Datensynchronisation dienen (Yahya-pour/Wieder 2013). Ein anderes Beispiel ist das Projekt GeneCloud im Rahmen des BMWi-Programms »Trusted Cloud«, in dem an der TU Dresden eine Infrastruktur entwickelt wird, die mittelständischen Unternehmen aus der Pharmabranche Zugang zu cloudbasierten Hochleistungsrechnern für die Analyse von Testergebnissen ermöglicht. Darüber hinaus gibt es nationale und internationale Veranstaltungen zum Einsatz von Cloud Computing in der Wissenschaft, wie beispielsweise den »Workshop on Scientific Cloud Computing«, der 2013 zum vierten Mal in Berlin stattfand.

PRIVATPERSONEN**2.4**

Der Vergleich von Studien zur Verwendung von Cloudangeboten durch Privatpersonen zeigt, dass die Abgrenzung des Begriffs »Cloud Computing« in diesem Bereich besonders schwierig ist. Es fällt auf, dass Cloud Computing in Europa breiter ausgelegt wird als in den USA. Es stellt sich die Frage, inwieweit alle im Rahmen der Studien untersuchten Dienste auch tatsächlich Clouddienste sind. Aus diesem Grund und aufgrund unterschiedlicher methodischer Ansätze ist der Vergleich der Studien nicht ohne Weiteres möglich.

Eine Studie, an der 947 Privatpersonen aus neun EU-Mitgliedstaaten teilnahmen, zeigt, dass sich die Nutzung von Cloudangeboten zwischen diesen Ländern deutlich unterscheidet (Cattaneo et al. 2012, S. 50). Während in Ungarn mehr als 30 % der Studienteilnehmer angeben, Cloudangebote zu nutzen, sind es in Deutschland weniger als 10 %. Die großen Unterschiede lassen sich auf Grundlage der Studie nicht erklären. Andere Studien bestätigen allerdings im Wesentlichen die geringe Nutzung in Deutschland. Im Hinblick auf die verwendeten Dienste sind die Unterschiede zwischen den einzelnen Staaten vernachlässigbar. Besonders häufig werden laut Cattaneo et al. (2012, S. 51) Suchdienste, Blogging- und Streamingdienste genutzt. Deutlich geringer ist das Interesse an Diensten, bei denen Daten in der Cloud gespeichert werden. Eine Ausnahme stellen allerdings Social-Network-Dienste dar.

Es ist nicht überraschend, dass Privatpersonen in der Regel nicht bereit sind, für dieselben Dienste, die kostenlos zur Verfügung gestellt werden, zu bezahlen. Am ehesten sind Nutzer bereit, für Streamingdienste zu bezahlen, bei denen Inhalte bezogen werden können (Cattaneo et al. 2012, S. 51). Studien in den USA kommen zu ähnlichen Ergebnissen (PwC 2012).

Einer Studie von Forrester (2012) mit mehr als 2.136 Teilnehmern zufolge nutzen 65 % der Privatverbraucher in den USA zumindest einen Cloudservice. Kalender- und Fotospeicherdienste werden der Studie zufolge am häufigsten verwendet. 44 bzw. 40 % der Studienteilnehmer gaben an, entsprechende Dienste zu nutzen. 19 % der Teilnehmer nutzen Kollaborationsdienste aus der Cloud (Forrester 2012, S. 7).

Insgesamt zeigen die Studienergebnisse, dass Cloudangebote nicht nur von Geschäftskunden, sondern auch von Privatpersonen gerne genutzt werden. Einige Autoren gehen sogar davon aus, dass insbesondere in den USA die Zahl der Early Adopter im Umfeld der Privatpersonen größer war als im Bereich der Geschäftskunden (Layo 2012; Schofield 2012).

FAZIT – BIG DATA ALS TREIBER FÜR CLOUD COMPUTING 3.

Im Rahmen einer Studie von Gartner (2012b) gab knapp ein Drittel der Teilnehmer (27 %) an, dass ihr jeweiliges Unternehmen bereits cloudbasierte Angebote zur Unterstützung seiner Fähigkeiten zur Datenanalyse in ausgewählten Geschäftsbereichen nutzt oder plant, dies innerhalb eines Jahres zu tun. Im Vergleich dazu haben nur 17 % der Befragten ihre vorhandene Infrastruktur teilweise oder komplett durch cloudbasierte Angebote ersetzt oder planen entsprechende Schritte. An der Umfrage, die 2011 durchgeführt wurde, nahmen 1.364 Fachbereichs- und IT-Verantwortliche teil. Datenanalyzelösungen auf Basis von Cloud Computing werden als kostengünstig und relativ leicht einführbar wahrgenommen. Gartner identifizierte mit kurzen Time-to-Value-Zyklen, Kostenvorteilen und fehlender Datenanalyseexpertise drei wesentliche Treiber für die Nutzung von cloudbasierten Datenanalyzelösungen. Dieses Ergebnis unterstreicht, dass Big-Data-Anwendungen ein zentraler Treiber für Cloud Computing werden können. Wesentlich dafür ist das Zusammenspiel zweier Faktoren. Einerseits besteht die Notwendigkeit von Investitionen in teure Infrastrukturen für diese Art von Datenanalysen, die vielen KMU in der Anfangszeit als zu riskant erscheinen könnten. Andererseits gibt es oftmals nur wenige schon existierende Infrastrukturen, die obsolet würden. Dementsprechend bietet es sich an, diese Leistung über cloudbasierte Lösungen zu beziehen. Im Gegenzug steigt dadurch aber auch die Erfahrung im Umgang mit Cloudangeboten. Fehlende Erfahrung mit Cloud Computing sowie die Expertise für größere Eingriffe in die Infrastruktur, die wahrscheinlich wesentliche Gründe für eine Zurückhaltung sind, könnten dadurch verringert werden und dazu führen, dass Cloud-Computing-Angebote auch für andere Anwendungsbereiche bei den regelmäßig notwendigen Entscheidungen im Rahmen des Investitionszyklus berücksichtigt werden. Diese eher evolutionäre Hybridisierung der IT-Infrastruktur würde auch mit neueren Marktforschungsergebnissen übereinstimmen, die einen Trend in diese Richtung voraussehen (Böckle 2014). Im Bereich der öffentlichen Verwaltungen dürfte der Wunsch nach dem Einsatz solcher Analysemöglichkeiten aus den gleichen Gründen den Bedarf an Cloud Computing steigern, aber es ist davon auszugehen, dass hier eventuell andere Lösungen im Rahmen gemeinsamer öffentlicher Infrastrukturen angestrebt werden.



POTENZIALE

V.

SOZIOÖKONOMISCHE POTENZIALE

1.

Sozioökonomische Potenziale werden auf der Mikroebene im Hinblick auf einzelne Organisationen betrachtet. Der Fokus liegt zunächst auf Unternehmen, Stellen und Behörden der öffentlichen Verwaltung sowie sonstige Organisationen (z. B. Nichtregierungsorganisationen) und dann auf wissenschaftlichen Einrichtungen. Anschließend werden Potenziale für Privatpersonen adressiert. Auf der Makroebene werden die Potenziale für die Gesellschaft und die Wirtschaft als Ganzes betrachtet (Kap. V.1.1–V.1.4). Sowohl auf der Mikro- als auch auf der Makroebene lassen sich sozioökonomische Potenziale nur zum Teil quantifizieren.

POTENZIALE FÜR ORGANISATIONEN

1.1

Vor allem im Hinblick auf Unternehmen können auf Grundlage aktueller Studien zahlreiche Potenziale von Big Data und Cloud Computing identifiziert werden. Die Studienteilnehmer wurden in der Regel nach Treibern für die Durchführung von Big-Data-Analysen bzw. die Nutzung von Clouddiensten sowie nach den damit verbundenen Zielen oder Vorteilen gefragt. Teilnehmer von Unternehmen, die Big-Data-Analysen bereits durchführen bzw. Clouddienste nutzen, wurden in ausgewählten Studien auch nach konkret beobachteten Auswirkungen von Big-Data-Analysen bzw. der Nutzung von Clouddiensten gefragt. Vor allem daraus sind zuverlässig Potenziale für Unternehmen ableitbar, die bisher keine oder nur sehr begrenzt Erfahrungen mit Big Data bzw. Cloud Computing gesammelt haben.

Die im Rahmen der Analyse von Studien identifizierten sozioökonomischen Potenziale wurden den Kategorien Entwicklung und Innovation, Prozesse und Organisation, Kunden und Finanzen (Kap. V.1.1.1–V.1.1.4) zugeordnet. Im Hinblick auf die Auswahl der Potenzialkategorien erscheint eine Orientierung an den klassischen Dimensionen der Balanced Scorecard sinnvoll, da von ihnen die wesentlichen Komponenten betrieblichen Erfolgs, aber auch gesellschaftlich relevante Komponenten abgedeckt werden. Die Balanced Scorecard wurde ursprünglich als Konzept zur Messung und Steuerung der Aktivitäten einer Organisation im Hinblick auf seine Vision und Strategie entwickelt (Kaplan/Norton 2009). Die Bedeutung der Potenzialkategorien sowie die konkrete Ausprägung der Potenziale unterscheiden sich teils deutlich zwischen Big Data und Cloud Computing. Dass die Zuordnung der identifizierten Potenziale nicht immer ein-



deutig möglich ist, hängt damit zusammen, dass die Dimensionen im Sinne einer Ursache-Wirkungs-Kette zusammen hängen (Kaplan/Norton 1997). Die im Rahmen der Entwicklungsdimension betrachteten Aspekte, wie das vorhandene Expertenwissen, beeinflussen Faktoren wie Qualität und Durchlaufzeit der internen Prozesse. Die Aspekte der Prozessdimension wiederum beeinflussen Aspekte wie Kundentreue, die der Kundendimension zugeordnet werden. Am Ende der Kette steht die Finanzdimension, für die der Return on Investment (ROI) ein zentrales Erfolgskriterium ist.

Im Hinblick auf Stellen und Behörden der öffentlichen Verwaltung, die Wissenschaft und Privatpersonen gibt es nur sehr wenige Studien, die sich mit sozio-ökonomischen Potenzialen von Big Data und Cloud Computing beschäftigen. Während anzunehmen ist und von einigen Studien bestätigt wird (z. B. KPMG 2012), dass die Potenziale für Stellen und Behörden der öffentlichen Verwaltung und wissenschaftliche Einrichtungen denen von Unternehmen recht ähnlich sind, muss davon ausgegangen werden, dass sich die Potenziale für Privatpersonen aufgrund anderer Zielsetzungen und Rahmenbedingungen zum Teil erheblich davon unterscheiden.

Sieht man von Umsatzsteigerungen und dem Aufbau strategischer Wettbewerbsvorteile ab, die bei staatlichen Stellen und Behörden sowie bei wissenschaftlichen Einrichtungen eine untergeordnete Rolle spielen, ergeben sich in diesen Bereichen im Prinzip dieselben sozioökonomischen Potenziale durch Big Data und Cloud Computing wie bei Unternehmen. Neben Kostensenkungen sind Verbesserungen bei Kundenzufriedenheit, Produktivität und Innovationsfähigkeit auf Grundlage von besseren Entscheidungen und höherer Flexibilität in allen Bereichen gefragt.

Im Hinblick auf Cloud Computing beziehen sich die besprochenen Potenziale primär auf die Nutzung von Diensten aus Public Clouds.

ENTWICKLUNG UND INNOVATION

1.1.1

Sowohl die Durchführung von Big-Data-Analysen als auch die Nutzung von Cloud Computing lassen den Studienteilnehmern zufolge Auswirkungen auf die Entwicklung von Organisationen erwarten. Im Hinblick auf Big Data ist in der Interxion-Studie von höherer Innovationsfähigkeit die Rede (Interxion 2013, S.11). 46 % der Studienteilnehmer gaben der Hoffnung Ausdruck, dass die Durchführung von Big-Data-Analysen zu Verbesserungen im Hinblick auf die Innovationsfähigkeit führt. Laut Schäfer et al. (2012, S.47) stellt die Entwicklung neuer Produkte und Dienstleistungen ein primäres Ziel der Durchführung von Big-Data-Analysen dar. Beim Aufbau strategischer Wettbewerbsvorteile handelt es sich um ein weiteres Ziel, das Schäfer et al. im Rahmen dieser Studie identifizierten. Keines der von Schäfer et al. identifizierten Ziele bekam eine brei-



tere Zustimmung als der Aufbau strategischer Wettbewerbsvorteile. Außerdem werden von der Durchführung von Big-Data-Analysen Vorteile bei der Erkennung von Absatzmöglichkeiten (Russom 2011, S. 11), eine schnellere Platzierung neuer Produkte und Dienstleistungen am Markt (Interxion 2013, S. 11) und die Möglichkeit zur Einführung neuer Geschäftsmodelle (Schroeck et al. 2012) erwartet.

Auch im Hinblick auf Cloud Computing ist von einer höheren Innovationsfähigkeit die Rede. Eine immer wichtigere Rolle spielt dabei mit Cloud Collaboration die Möglichkeit, durch cloudbasierte Lösungen sowohl innerhalb einer Organisation als auch über die Grenzen einer Organisation hinaus unabhängig vom tatsächlichen Aufenthaltsort der Mitarbeiter zusammenzuarbeiten. Umfragen zufolge fördert Cloud Collaboration in besonderer Weise das Innovationspotenzial von Organisationen (Forbes Insights/Cisco 2013). Die Zusammenarbeit kann dabei sowohl über spezialisierte Lösungen innerhalb existierender Anwendungen realisiert werden (z. B. Groupwareanwendungen) als auch über die Nutzung von speziellen, auf Cloud Collaboration ausgerichteten Anwendungen. Letztere werden auch als Social-Enterprise-Lösungen bezeichnet, da ihrer Funktionalität teilweise an klassische Social Networks erinnert. Bekannteste Beispiele sind Yammer oder blueKiwi, die bis zur Übernahme durch Microsoft bzw. Atos rein cloudbasierte Lösungen waren.

Bei einer von KPMG/BITKOM (2013, S. 22) durchgeführten Studie gaben mit 49 % knapp die Hälfte der Befragten an, dass sie die Verbesserung der eigenen Innovationsfähigkeit als sehr wichtiges oder eher wichtiges Ziel der Nutzung von Public-Cloud-Diensten sehen. 26 % der Vertreter von Unternehmen, die entsprechende Dienste nutzen, gaben an, dass die Innovationsfähigkeit des von ihnen vertretenen Unternehmens zugenommen oder deutlich zugenommen hat. Nur 2 % berichteten, dass es im Hinblick auf die eigene Innovationsfähigkeit zu einer Abnahme kam. KPMG/BITKOM (2013) kam außerdem zum Ergebnis, dass kürzere Implementierungszeiten für neue Anwendungen und Lösungen ein zentrales Ziel der Nutzung von Clouddiensten ist. Bei Betrachtung der tatsächlichen Auswirkungen fällt allerdings auf, dass sie äußerst uneinheitlich sind. Während 40 % der Befragten angaben, dass die Implementierungszeiten abgenommen oder deutlich abgenommen haben, gaben 41 % an, dass die Implementierungszeiten zugenommen haben. Dieses Ergebnis legt den Schluss nahe, dass die Auswirkungen der Nutzung von Clouddiensten auf die Implementierungszeiten für neue Anwendungen und Lösungen maßgeblich von weiteren Faktoren abhängt.

Ein weiteres von KPMG/BITKOM (2013) identifiziertes Ziel ist die Abnahme der Notwendigkeit, eigene IT-Fachkräfte zu beschäftigen. 26 % der Vertreter von Unternehmen, die Public-Cloud-Dienste nutzen, konnten bestätigen, dass es tatsächlich zu einer Abnahme kam. Nur 4 % der Befragten nahmen eine Zunahme bei der Notwendigkeit, eigene IT-Fachkräfte zu beschäftigen, wahr. Auch



das Cloud Industry Forum konnte im Rahmen von Studien belegen, dass die Kompensation von fehlendem Expertenwissen durchaus ein Grund für die Nutzung von Clouddiensten ist (Cloud Industry Forum 2012a u. 2012b). In der Möglichkeit, neue Dienste anzubieten, sieht das Cloud Industry Forum noch einen weiteren Grund für die Nutzung von Clouddiensten, der mit der Entwicklung von Organisationen zusammenhängt. Vor allem in den USA fand dieser Grund mit 22 % relativ breite Zustimmung (Cloud Industry Forum 2012b).

PROZESSE UND ORGANISATION

1.1.2

Sowohl die Durchführung von Big-Data-Analysen als auch die Nutzung von Cloud Computing lassen verschiedenen Studien zufolge Verbesserungen im Hinblick auf die internen Prozesse von Unternehmen erwarten. Im Zusammenhang mit Big Data werden vor allem bessere Entscheidungen erwartet (Bange et al. 2013; Interxion 2013; Schäfer et al. 2012; TCS 2013). Bange et al. (2013, S.25) unterscheiden zwischen strategischen Entscheidungen und der Steuerung operativer Prozesse. Während 59 % der Teilnehmer an der BARC-Studie bessere strategische Entscheidungen erwarteten, ging mit 51 % noch gut jeder zweite Befragte von einer besseren Steuerbarkeit operativer Prozesse als Folge der Durchführung von Big-Data-Analysen aus. Auch Schroeck et al. (2012, S.7) stellten fest, dass mithilfe von Big-Data-Analysen die Optimierung betrieblicher Abläufe erwartet wird.

Die Studienteilnehmer gaben darüber hinaus an, dass sie als Folge der Durchführung von Big-Data-Analysen sowohl eine höhere Arbeitsproduktivität als auch eine höhere Qualität von Produkten und Dienstleistungen erwarten (Bange et al. 2013; Interxion 2013; Schäfer et al. 2012). In der Interxion-Studie gehen 26 % der Teilnehmer davon aus, dass die Durchführung von Big-Data-Analysen allgemein zu höherer Arbeitsproduktivität führt. Eine empirische Untersuchung mit Daten börsennotierter Unternehmen zeigt beispielsweise, dass Unternehmen, die ihre Entscheidungen auf der Grundlage einer systematischen Datenanalyse treffen, eine um 5 bis 6 % höhere Produktivität erzielen (Brynjolfsson et al. 2011). Mit 46 % erwartet knapp die Hälfte der Studienteilnehmer, dass mithilfe von Big-Data-Analysen vor allem neue Produkte und Dienstleistungen schneller auf den Markt gebracht werden können (Interxion 2013, S.11). Auch Schäfer et al. (2012, S.47) kamen zu dem Ergebnis, dass die Erhöhung der Produktivität zu den wichtigsten Zielen von Big Data in Unternehmen gehört. Jeder vierte Teilnehmer der BARC-Studie gab an, dass eine höhere Qualität von Produkten und Dienstleistungen zu den wichtigsten Vorteilen der Durchführung von Big-Data-Analysen gehört (Bange et al. 2013, S.25). Während mit 24 % knapp ein Viertel der Teilnehmer der Interxion-Studie zusätzlich auf die Bedeutung von Big-Data-Analysen für den Bereich Compliance und Governance hinweist, heben Schroeck et al. (2012, S.7) die Bedeutung für den Bereich Risiko- und Finanzmanagement



hervor. Dem Bereich Risiko- und Finanzmanagement wurde von 15 % der Befragten hohe Bedeutung beigemessen.

Im Hinblick auf Cloud Computing wird vor allem größere organisatorische Flexibilität erwartet (Cloud Industry Forum 2012a u. 2012b; KPMG/BITKOM 2013). KPMG/BITKOM (2013, S.22) haben im Rahmen der Studie »Cloud-Monitor 2013« nicht nur nach den Zielen der Nutzung von Public-Cloud-Diensten gefragt, sondern auch nach den bisherigen Auswirkungen der Nutzung entsprechender Dienste. Höhere organisatorische Flexibilität stellt für 55 % der Befragten ein sehr wichtiges oder ein eher wichtiges Ziel der Nutzung dar. Von den Vertretern von Unternehmen, die Public-Cloud-Dienste nutzen, gaben 36 % an, dass die organisatorische Flexibilität zugenommen hat oder sogar deutlich zugenommen hat. Während im Rahmen der Studie von KPMG/BITKOM ausschließlich Vertreter deutscher Unternehmen befragt wurden, konzentrierte sich das Cloud Industry Forum auf die USA bzw. das Vereinigte Königreich. In den USA gaben 31 % der Befragten an, dass sie die größere organisatorische Flexibilität noch vor niedrigeren Kosten als wichtigsten Grund für die Nutzung von Clouddiensten halten (Cloud Industry Forum 2012b, S.7). Im Vereinigten Königreich wird den Kosten etwas mehr Bedeutung beigemessen als der organisatorischen Flexibilität. Das Cloud Industry Forum (2012a, S.7) weist allerdings darauf hin, dass in früheren Untersuchungen die organisatorische Flexibilität vorne lag.

Bradshaw et al. (2012, S.23) beschreiben im Hinblick auf interne Prozesse weitere Potenziale wie z.B. höhere Produktivität, einheitliche Abläufe, einfachere Eröffnung von Zweigstellen und einfachere Einführung neuer Produkte oder Produktlinien. Wie bei der Studie von KPMG/BITKOM wurde auch bei der Studie von Bradshaw et al. nicht nur der erwartete, sondern auch der tatsächlich beobachtete Nutzen erhoben. Während beispielsweise die einfachere Eröffnung von Zweigstellen und Einführung neuer Produkte oder Produktlinien eng mit einer höheren organisatorischen Flexibilität zusammenhängen, ist das bei der höheren Produktivität und den allgemein einfacheren Abläufen nicht der Fall.

Auch im Bereich der IT werden Potenziale in der Nutzung von Diensten aus der Cloud gesehen. KPMG/BITKOM (2013, S.22) kommen zum Ergebnis, dass durch die Nutzung von Clouddiensten nicht nur der IT-Administrationsaufwand zurückgehen, sondern auch die Verfügbarkeit und Performance von IT-Leistungen steigen kann. In beiden Fällen handelt es sich nicht nur um Ziele, sondern um tatsächlich beobachtete Auswirkungen der Nutzung von Clouddiensten. Bei 56 % der Befragten hat der IT-Administrationsaufwand durch die Nutzung von Clouddiensten abgenommen oder stark abgenommen. 14 % der Befragten gaben an, dass der Administrationsaufwand zugenommen hat. Ähnlich gaben 47 % der Studienteilnehmer an, dass die Verfügbarkeit und Performance von IT-Leistungen zugenommen oder deutlich zugenommen hat. Bei nur 6 % kam es den Ergebnissen der Studie zufolge zu einer Abnahme im Hinblick auf Verfügbarkeit



und Performance. Bei der Möglichkeit zur schnelleren Skalierbarkeit von IT-Leistungen handelt es sich um eine weitere zentrale Auswirkung, die von KPMG/BITKOM (2013, S.22) untersucht wurde. 57 % der Studienteilnehmer gaben an, dass die Skalierbarkeit zugenommen oder deutlich zugenommen hat. Nur 6 % berichteten, dass die Skalierbarkeit abgenommen oder deutlich abgenommen hat. Im Rahmen einer Studie des Cloud Industry Forum wurde nach Kosteneinsparungen und Flexibilitätssteigerungen die Skalierbarkeit von IT-Leistungen als häufigster Grund genannt, warum Clouddienste genutzt werden. Ein weiterer Vorteil ist in umfassenden, meist uneingeschränkten Servicezeiten zu sehen (Cloud Industry Forum 2012b).

Es fällt auf, dass sowohl die Nutzung von Clouddiensten als auch die Durchführung von Big-Data-Analysen das Potenzial haben, die Produktivität von Organisationen zu steigern.

KUNDEN

1.1.3

Während angenommen wird, dass die Durchführung von Big-Data-Analysen durchaus direkte Auswirkungen auf die Kunden einer Organisation haben kann, ist das bei Cloud Computing eher nicht der Fall. Im Rahmen mehrerer Studien wurde festgestellt, dass durch Big-Data-Analysen eine bessere Kundenorientierung erreicht werden kann (Schroeck et al. 2012). Den Studien zufolge wird erwartet, dass zielgerichtetes Marketing (Bange et al. 2013; Russom 2011) und besserer Kundenservice (Bange et al. 2013; Schäfer et al. 2012) zu höherer Kundenzufriedenheit (Interxion 2013) und besserer Kundenbindung (Bange et al. 2013) führen können. Laut Schroeck et al. (2012, S.7) handelt es sich bei Kundenorientierung um das mit Abstand am häufigsten genannte Ziel von Big-Data-Analysen. Mit 49 % gab fast die Hälfte der Befragten an, dass es bei der Analyse großer Datenmengen vorrangig um Kundenorientierung geht. Im Rahmen der TDWI-Studie gaben 61 % der Befragten an, dass sie sich durch Big-Data-Analysen eine bessere Ausrichtung der Marketingmaßnahmen erhoffen (Russom 2011, S.11). Zielgerichtete Maßnahmen im Marketing nehmen 31 % der im Rahmen der BARC-Studie Befragten als einen der wichtigsten Vorteile von Big-Data-Analysen wahr (Bange et al. 2013, S.25). Ähnlich viele Teilnehmer dieser Studie (32 %) halten einen verbesserten Kundenservice für einen wichtigen Vorteil. Schäfer et al. (2012, S.47) kommen ebenfalls zum Schluss, dass die Verbesserung des Kundenservice ein zentrales Ziel von Big-Data-Anwendungen ist. Aus Sicht der Interxion-Studie handelt es sich bei der Verbesserung der Kundenzufriedenheit um den zweitwichtigsten Nutzen aus der Durchführung von Big-Data-Analysen nach allgemein besseren Entscheidungen (Interxion 2013). Mit 54 % gaben mehr als die Hälfte der Befragten an, dass aus ihrer Sicht im Bereich der Kundenzufriedenheit der zentrale Nutzen liegt. Eine stärkere Kundenorientierung und damit eine höhere Kundenzufriedenheit werden vor allem durch die

Berücksichtigung der Ergebnisse umfassenderer Analysen erzielt. Kundenorientierung ist keinesfalls ein Thema das nur für Unternehmen relevant ist. Auch für Stellen und Behörden der öffentlichen Verwaltung ist Kundenorientierung von großer Bedeutung. Gleichzeitig stellen aber Studien der öffentlichen Verwaltung in Deutschland in dieser Hinsicht ein schlechtes Zeugnis aus (z.B. Stellermann et al. 2011).

FINANZEN

1.1.4

Sowohl die Durchführung von Big-Data-Analysen als auch die Nutzung von Cloud Computing lassen den Teilnehmern verschiedener Studien zufolge Auswirkungen auf die finanzielle Situation von Organisationen erwarten. Im Hinblick auf Big Data werden vor allem Umsatzsteigerungen erwartet (Bange et al. 2013; Schäfer et al. 2012), aber auch eine Reduktion der Kosten erscheint nicht unwahrscheinlich (Interxion 2013; Schäfer et al. 2012). Laut Schäfer et al. (2012, S.47) gehören sowohl die Steigerung des Umsatzes als auch die Einsparung von Kosten, abgesehen vom Aufbau strategischer Wettbewerbsvorteile, zu den wichtigsten Zielen, die mit Big-Data-Analysen verbunden werden. Im Rahmen der Interxion-Studie gab mit 47 % fast jeder zweite Teilnehmer an, dass sich insbesondere durch Cross- und Up-Selling auf Grundlage von Big-Data-Analysen Umsatzsteigerungen erzielen lassen (Interxion 2013, S.11). Während beim Cross-Selling dem Kunden Zusatzprodukte anderer Produktfamilien angeboten werden, geht es beim Up-Selling um den Verkauf eines höherwertigen Produkts derselben Familie. Empfehlungssysteme für Aktivitäten wie Cross- und Up-Selling basieren in der Regel auf der Analyse großer Datenmengen. Im Rahmen der von TCS (2013, S. 33) durchgeführten Studie gaben 43 % der Teilnehmer an, dass sie im Hinblick auf Big Data von einem ROI von mehr als 25 % ausgehen. Ein Drittel der Befragten erwartete einen positiven ROI von bis zu 25 %. Immerhin 28 % der Teilnehmer an der BARC-Studie gaben an, dass sie in geringeren Kosten einen wichtigen Vorteil der Nutzung von Big-Data-Analysen sehen (Bange et al. 2013, S.25). Kostensenkungen können durch die Berücksichtigung der Ergebnisse umfassenderer Analysen bei Entscheidungen erzielt werden.

Auch Cloud Computing soll sowohl eine Kostenreduktion als auch Umsatzsteigerungen möglich machen (Bradshaw et al. 2012; Cloud Industry Forum 2012a u. 2012b; KPMG/BITKOM 2013). Im Hinblick auf die Reduktion von Kosten ist einerseits von niedrigeren Anschaffungskosten (Cloud Industry Forum 2012a u. 2012b) und niedrigeren Betriebskosten (Cloud Industry Forum 2012a) für die IT-Infrastruktur die Rede und andererseits von einer Umwandlung von einmaligen Investitionskosten aufseiten der IT in laufende Kosten aufseiten der Fachbereiche (Bradshaw et al. 2012). Diese Möglichkeit macht Cloud Computing vor allem für KMU interessant. Studien zeigen, dass viele Unternehmen durch die Nutzung von Clouddiensten ihre IT-Kosten um 10 bis 20 % reduzieren können



(EK 2012b). Im Rahmen einer vom Cloud Industry Forum (2012b, S.7) durchgeführten Studie gaben 28 % der Teilnehmer an, dass eine Reduktion der Kosten der wichtigste Grund für die Nutzung von Clouddiensten ist. Nur 3 % der Befragten gaben an, dass der erwartete ROI eine wichtige Rolle spielt. Im Rahmen der von KPMG/BITKOM durchgeführten Studie wurden neben den Zielen der Nutzung von Public-Cloud-Diensten auch die bisherigen Auswirkungen der Nutzung untersucht. Von den Studienteilnehmern bezeichneten 55 % die Senkung der IT-Ausgaben als sehr wichtiges oder eher wichtiges Ziel. Bei Betrachtung der tatsächlichen Auswirkungen fällt allerdings auf, dass sie uneinheitlich sind. Während 25 % der Vertreter von Unternehmen, die Public-Cloud-Dienste nutzen, bestätigten, dass die Höhe der IT-Ausgaben abgenommen oder deutlich abgenommen hat, gaben ebenso 25 % an, dass die IT-Ausgaben sogar zugenommen haben (KPMG/BITKOM 2013, S.22).

Dies legt den Schluss nahe, dass die Auswirkungen der Nutzung von Clouddiensten auf die Kosten maßgeblich von weiteren Faktoren abhängen. Ein Grund für die uneinheitlichen Ergebnisse könnte darin liegen, dass sich eine vollständige Erhebung der mit Cloud Computing verbundenen Kosten im Vorfeld der Nutzung oft schwierig gestaltet. Violino (2011) weist beispielsweise darauf hin, dass mitunter unerwartet hohe Kosten für den Transfer von Daten in die Cloud oder für die langfristige Speicherung von Daten in der Cloud anfallen. Dieser Faktor hat im Kontext von Big Data eine besondere Relevanz. Auch die Kosten für die Integration von Clouddiensten können unerwartet hoch sein und Kostenvorteile durch die Nutzung von Cloud Computing rasch zunichtemachen (Violino 2011). Bradshaw et al. (2012, S.23) fanden heraus, dass eine Steigerung der Umsätze durch die Nutzung von Clouddiensten nicht nur erwartet wird, sondern auch tatsächlich beobachtet werden kann. Umsatzsteigerungen durch Clouddienste sind unter anderem auf eine bessere Bewältigung von Anforderungsspitzen und auf die einfachere Einführung neuer Produkte oder Produktlinien zurückzuführen.

POTENZIALE FÜR DIE WISSENSCHAFT

1.2

Wie bereits im Kapitel IV erwähnt, bieten Big Data und Cloud Computing auch Potenziale für die Wissenschaft. Vor allem die Kombination von Big Data und Cloud Computing eröffnet neue Möglichkeiten.

Bei Big Data gibt es grundsätzlich zwei wesentliche Aspekte, die ein großes Potenzial für wissenschaftliches Arbeiten beinhalten. Der erste Aspekt betrifft die Datenverfügbarkeit. Parallel zur Open-Data-Debatte für Informationen, gibt es auch in der Wissenschaft eine Diskussion über den Zugang zu Forschungsdaten. Dies geht über das wesentlich ältere Open Access, das auf einen freien Zugang zu wissenschaftlichen Publikationen abzielt, deutlich hinaus. Zentraler Punkt ist hier nicht nur der Zugang zu den Publikationen, sondern auch zu den dahinter



stehenden Forschungsmaterialien und Datensätzen. Da diese zunehmend digital produziert oder digitalisiert vorliegen, entstehen immer mehr Möglichkeiten, dass andere Forscher diese ebenfalls nutzen können. Hinzu kommt, dass sie bei Big Data noch mit öffentlich verfügbaren Datensätzen der Verwaltungen, Datenbanken oder weiteren Internetquellen verbunden werden könnten. Dadurch würden sich vollkommen neue Möglichkeiten für die Bearbeitung wissenschaftlicher Fragestellungen ergeben. Dies führt in direkter Linie zu dem zweiten Aspekt, den Potenzialen für die wissenschaftlichen Vorgehens- und Arbeitsweisen. Grundsätzlich erlaubt Big Data zunehmend eine empirisch-deduktive Vorgehensweise, bei der aus Datenauswertungen Aussagen abgeleitet werden, im Gegensatz zu einer eher theorie- und modellgetriebenen Vorgehensweise. Manche Vertreter dieser Entwicklung gingen in der ersten Euphorie soweit, vom Ende der Theorie zu sprechen (Anderson 2008), was sich aber heute schon wieder etwas relativiert hat. Im Laufe der Zeit haben sich dabei verschiedene Richtungen und Tendenzen entwickelt, meist getrennt nach den klassischen Bereichen Naturwissenschaften und Geistes- und Sozialwissenschaften.

Eine Strömung sieht Big Data als neues, viertes Paradigma der Wissenschaft (Hey et al. 2009). Darunter wird verstanden, dass sich die Technologien und die Techniken, die bei Forschung mit Big Data (Data-intensive Science) eingesetzt werden, von den bisherigen derart unterscheiden, dass hierunter ein neues Paradigma ausgerufen werden kann (Lynch 2009). Aus ihrer Sicht ist dies die logische Konsequenz aus dem Einsatz von IT in der Forschung. Diese neue Art der Forschung wird dabei als »capture, curation and analysis« (Hey et al. 2009) zusammengefasst. Dementsprechend beziehen sich deren Vertreter auf Wissenschaftsgebiete, in denen durch Experimente und Verfahren prinzipiell viele Daten anfallen. Vorreiter sind daher vor allem Naturwissenschaften, sei es in der Astronomie, in der Meteorologie (z. B. Mangelsdorf 2012), in der Physik mit dem LHC oder der Biologie durch die Sequenzierung von Genen (z. B. Anderson 2008). Dies sind nur einige wenige Momentaufnahmen, aber ohne den Einsatz von Big Data wären viele Forschungen in diesen Bereichen nur deutlich langsamer oder überhaupt nicht möglich. Gleichzeitig kann die Verwendung dieser Techniken zu völlig neuen Blickwinkeln und Denkweisen führen, die neue Erkenntnisse, neue Forschungsfragen und -felder und Innovationen hervorbringen (Harris 2012).

Eine etwas gemäßigtere Strömung innerhalb der Naturwissenschaften sieht Big Data mehr als ein drittes wissenschaftliches Werkzeug neben Theorie und Experiment (The Royal Society 2012). Dies zeigt auch der Trend, der nach der ersten Euphorie Big Data nicht als grundsätzlichen Bruch mit bisherigen wissenschaftlichen Arbeitsweisen sieht, sondern vielmehr als eine, wenn auch teilweise grundlegend neue, Ergänzung der wissenschaftlichen Methoden. Unabhängig davon, ob Big Data nun einen Paradigmenwechsel darstellt oder eine wichtige Erweiterung, sind sich Vertreter beider Richtungen darüber einig, dass nicht nur inner-



halb der einzelnen Disziplinen, sondern auch über die Disziplinen hinweg zusammengearbeitet werden muss (Hey et al. 2009; The Royal Society 2012). Dies ist beispielsweise auch Ziel der entsprechenden Programme auf EU-Ebene (eInfrastructures), die vor allem auf generische, disziplinenübergreifende Plattformen ausgerichtet sind. Ein Beispiel dafür ist die EUDAT-Plattform, die zu einer Plattform für wissenschaftliche Daten aus allen Disziplinen nach einheitlichen Standards weiterentwickelt werden soll. Hierbei spielen aber vor allem die Naturwissenschaften eine wichtige Rolle.

Im Gegensatz zu diesen Strömungen, wo die Nutzung von IT in der Forschung schon verbreitet ist, sieht die Lage in den Geistes- und Sozialwissenschaften wesentlich differenzierter aus. So verwenden einige Disziplinen wie beispielsweise die Archäologie, die empirische Sozialforschung oder Teile der Wirtschaftswissenschaften schon länger IT zu Forschungszwecken. Beispiele sind Berechnungen und Simulationen von Fundstellen oder Datenanalysen in größeren Datenmengen. Jedoch ist in vielen anderen Disziplinen IT bis heute kein Werkzeug der Forschung selbst. Dementsprechend zielen verschiedene Maßnahmen wie die BMBF-Förderlinie eHumanities darauf ab, eine Entwicklung in diese Richtung anzustoßen. Dennoch ist diese Entwicklung kein Selbstläufer, und viele Fachvertreter stehen den Zielen solcher Vorhaben ambivalent bis ablehnend gegenüber. Ein Grund dafür sind wahrscheinlich sehr ausgeprägte Strömungen wie Social Physics (Pentland 2012b) oder auch Computational Social Science (Lazer et al. 2009), die durch den Einsatz von IT im Allgemeinen und Big Data im Speziellen darauf abzielen, die Geistes- und Sozialwissenschaften zu »quantifizieren«. Ziel ist es dabei, beispielsweise gesellschaftliche Entwicklungen oder soziale Phänomene in Gegenwart oder auch Geschichte durch Einbeziehung von vielen Daten zu quantifizieren und zu dynamisieren, also Entwicklungen im Zeitverlauf darzustellen (Lazer et al. 2009; Manovich 2012). Pentland (2014) geht sogar von einer völligen Umkehrung der Ansätze aus. Aus seiner Sicht sind insbesondere soziale Verbindungen (Social Ties), d.h. mit wem und wie kommuniziert wird (Metadaten), Grundlage für das Verständnis sozialer Systeme, weniger die Analyse beispielsweise von einzelnen Institutionen oder der Handlungen Einzelner. Ausgehend von diesen Kenntnissen über die sozialen Verbindungen kann dann seiner Meinung nach ein Verständnis für das Gesamtsystem gewonnen werden. Dies setzt natürlich Zugang zu großen Datensätzen voraus, die gerade in einigen Teilbereichen kaum möglich oder wünschenswert sind. Zugleich lösen solche Ansprüche im Hinblick auf Erklärungsansätze den Widerspruch der etablierten Strömungen aus und sind vielleicht sogar kontraproduktiv. Denn letztlich behindern solche Antagonismen die fortschreitende Nutzung neuer Methoden und Werkzeuge, auch jenseits von Big Data, von denen auch die Geistes- und Sozialwissenschaften in Kombination mit ihren tradierten Vorgehensweisen auf viele Weisen, wie beispielsweise höhere Effizienz durch bessere Kommunikation und offenere Diskurse, profitieren könnten.



Unabhängig von der Entwicklung dieser teilweise sehr unterschiedlichen Strömungen wird Big Data ein Teil der wissenschaftlichen Arbeit werden. Dies wird auch zur Entwicklung neuer Techniken und Ansätze führen, die wiederum auch neue Erkenntnisse hervorbringen (Harris 2012; Vande Moere et al. 2012). In welchem Umfang wird sich aber erst in den nächsten Jahren zeigen, sodass es zu früh ist, bereits heute darüber zu urteilen. Jedoch verdeutlicht die Analyse auch die (künftige) Bedeutung von Cloud Computing im Wissenschaftsbereich, denn gerade jenseits der »Big Sciences«, also jenseits von Projekten wie dem Teilchenbeschleuniger LHC am CERN, haben viele Wissenschaftler keinen ausreichenden Zugang zu leistungsfähigen IT-Ressourcen. Daher werden Projekte, die cloudbasierte Lösungen für Big Data in der Wissenschaft bereitstellen, eine entscheidende Rolle bei der Realisierung der Potenziale spielen.

POTENZIALE FÜR PRIVATPERSONEN

1.3

Grundsätzlich muss festgehalten werden, dass Privatpersonen, wie bereits angedeutet (Kap. IV.1.4 u. IV.2.4), entweder als Konsument und Nutzer oder als Bürger, sehr unterschiedlich von Cloud Computing und Big Data profitieren können. Einerseits gibt es direkte Potenziale durch die eigene Nutzung von Cloud Computing und Big Data und andererseits indirekte Potenziale, die durch die Nutzung durch andere entstehen. Bei Letzterem ist aber die Grenze zu den gesamtgesellschaftlichen Effekten sehr ungenau.

Die direkten Potenziale sind am deutlichsten bei der Nutzung von Cloud Computing erkennbar. Im Hinblick auf Cloud Computing ist vor allem der Komfort beim Zugriff oder Teilen von Daten bemerkenswert. Die Möglichkeiten reichen von einfachen Datenspeicherlösungen bis hin zur Nutzung von Streamingangeboten für Filme oder Musik, die teilweise, aber nicht nur cloudbasiert sind. Hinzu kommen die indirekten Potenziale durch die Nutzung vieler Apps, die wiederum auf Cloud-Computing-Lösungen aufgebaut sind. Neben Dropbox, das auf Amazon Web Services aufbaut, basieren viele andere Apps ebenfalls auf den Angeboten großer IaaS-Anbieter. Hierzu zählen insbesondere solche Apps, die von Privatpersonen eingesetzt werden, um Arbeitsprozesse zu erleichtern. Dies reicht von einfachen E-Mail- und Office-Anwendungen zur mobilen Arbeit bis hin zu Kollaborationsanwendungen. Offen ist jedoch, ob Trends in diesem Bereich wie IT-Consumerization oder »bring your own device« (BYOD) und die damit einhergehende Verschmelzung von Privat- und Berufsleben ausschließlich positiv zu beurteilen sind. Weitere Potenziale liegen darin, dass manche Unternehmen ihre Kollaborationslösungen auch für ihre Kunden in bestimmten Bereichen öffnen und somit die Möglichkeit besteht, Entwicklungen mitzugestalten. Noch weiter geht diese Mischung der Rolle von Produzent und Konsument (Prosument) im Bereich des sogenannten Maker Movement, wo Privatleute und Kleinunternehmer (z.B. Künstler, Freiberufler) beispielsweise die Möglichkeiten der immer



günstiger werdenden 3-D-Drucker ausnutzen und individuelle Gegenstände herstellen. Auch hier hat sich in den letzten Jahren eine zunehmende Anzahl von Cloudangeboten entwickelt, die einen einfacheren Zugang ermöglichen sollen. Beispiele sind das Erstellen von dreidimensionalen Druckvorlagen bis hin zu Druck und Lieferung. Aktiv sind dabei Start-ups wie Sculpteo oder Netfabb, aber auch etablierte Softwareanbieter wie Autodesk. Manche Visionen in diesem Umfeld gehen gar so weit, hier von einer neuen industriellen Revolution zu sprechen (Anderson 2012), wobei es auch hier berechtigte Kritik gibt (Rotman 2013). Dennoch sollte abschließend angemerkt werden, dass Privatpersonen bei der Nutzung von Cloud Computing vor allem durch ein Streben nach Komfort getrieben werden, während die Bereitschaft, Kosten dafür in Kauf zu nehmen, vorläufig noch gering ist (Bradshaw et al. 2012, S.25 f.). Jedoch bezahlen viele bei den großen, stark genutzten Diensten für E-Mail, Datenaustausch oder Ähnliches mit ihren Daten.

Auf diesem Weg tragen Privatpersonen als Kunden zur Sammlung großer Datenmengen bei. Auch in der Rolle als Bürger tragen sie zur Generierung von Daten bei. In beiden Fällen führen sie in der Regel aber selbst keine Analysen durch. Der Nutzen und das Potenzial für sie liegen vielmehr darin, dass es Unternehmen oder der öffentlichen Verwaltung möglich wird, angepasste Angebote und Dienstleistungen, z.B. in Form präziserer Kaufempfehlungen oder optimierter Bereitstellung von Infrastrukturen, anzubieten. Da den betroffenen Konsumenten oder Bürgern dabei oftmals nicht bewusst ist, dass diese Dienstleistungen und Angebote auf solchen Vorarbeiten beruhen, sind sie nicht ganz unumstritten. Einen anderen Ansatz verfolgen hier sogenannte Datenmarktplätze, wo Privatpersonen die Möglichkeit haben, ihre Daten einzustellen und darüber zu entscheiden, wer in welcher Form Zugang dazu erhält. Inwieweit sich solche Konzepte durchsetzen können, ist aber noch offen. Während in diesen Lösungen die Rolle von Privatpersonen eher passiv ist, gibt es in zwei Bereichen Möglichkeiten wenigstens teilweise aktiv zu partizipieren. Die erste Möglichkeit liegt hierbei in den Bereichen Open Data und Open Government. So bieten staatliche Einrichtungen wie Kommunen immer öfter die Möglichkeit, als Bürger an Entwicklungen zu partizipieren, indem Daten beispielsweise zur Haushaltsgestaltung oder Ähnlichem offengelegt werden. Diese können dann für Auswertungen genutzt werden (z. B. Klessmann et al. 2014). Eine zweite Möglichkeit ergibt sich bei der Nutzung öffentlich verfügbarer Forschungsdaten. Hierdurch würde die Vision einer Wissenschaft durch Bürger für Bürger (Citizen Science), die schon seit den 1970er Jahren immer wieder propagiert wird, möglich werden (Hand 2010). Denn gerade cloudbasierte Big-Data-Lösungen (Kap. IV.1.4) eröffnen hier ein großes Potenzial, da keine Infrastruktur aufgebaut werden muss, sich die erforderlichen Kenntnisse in Grenzen halten und die Kosten überschaubar bleiben. Dennoch gilt es dabei noch viele Herausforderungen zu bewältigen, sodass offen bleibt, ob und wie sich diese Möglichkeiten in Zukunft entwickeln werden.

GESAMTGESELLSCHAFTLICHE POTENZIALE**1.4**

Ähnlich wie im Hinblick auf die sozioökonomischen Wirkungen für Unternehmen, die öffentliche Verwaltung und Privatpersonen gibt es auch im Zusammenhang mit den gesamtwirtschaftlichen und gesamtgesellschaftlichen Potenzialen von Big Data und Cloud Computing eine Vielzahl von teilweise kontroversen Debatten. Grundsätzlich gibt es neben vielen Veröffentlichungen, die vor allem auf pauschalen Annahmen oder vereinfachenden Schätzungen beruhen, nur eine überschaubare Zahl an Studien, die mit wissenschaftlichen Methoden entweder qualitativ oder quantitativ eine Erfassung oder gar Bewertung der Potenziale versuchen. Letztere zielen in der Regel auf eine Quantifizierung der ökonomischen Potenziale – also Produktivität, Wachstum und Beschäftigung – ab. Während für Cloud Computing mehrere Studien vorliegen (Bradshaw et al. 2012; DIW econ 2010; Etro 2009, 2011; Hogan et al. 2010, 2011), ist die Auswahl an Studien für Big Data deutlich kleiner (Manyika et al. 2011; OECD 2013a, 2013b). Daneben existiert eine Reihe von Artikeln und Büchern, die sich abstrakt mit diesen Potenzialen von Big Data beschäftigen (z. B. Mayer-Schönberger/Cukier 2013). Eine gemeinsame Bewertung beider Technologien existiert nicht, wobei jedoch in einzelnen Studien zu Big Data die besondere Rolle von Cloud Computing als sogenannter Enabler thematisiert wird (z. B. OECD 2013b, S. 324). Darüber hinaus finden sich noch Studien, die sich eher generell mit den gegenwärtigen und zukünftigen Potenzialen internetbasierter Dienste beschäftigen (Dean et al. 2012; Hoorens et al. 2012; Pélassié du Rausas et al. 2011). Diese sollen hier vor allem zur Ergänzung, Erweiterung und Verbindung der vorliegenden Ergebnisse genutzt werden. Darüber hinaus werden an einzelnen Punkten auch weniger quantifizierbare Aspekte wie Nachhaltigkeit genannt, aber meistens nicht ausführlich beleuchtet. Diese Potenziale sind oftmals in einzelnen, eher fachbezogenen Studien zu finden, die sich eher gezielt mit Themen wie Nachhaltigkeit (z. B. Hintemann/Fichter 2012) oder Transparenz (z. B. Janssen et al. 2012) als mit Technologien beschäftigen. Jedoch handelt es sich hierbei vor allem um qualitative Einschätzungen möglicher zukünftiger Auswirkungen, die sich nur teilweise mit Big Data oder Cloud Computing beschäftigen. Dementsprechend sollen hier nur wesentliche Ergebnisse und Aussagen ausgewählt dargestellt werden.

Grundsätzlich ist bei dieser Art von Studien anzumerken, dass es eine Reihe von Herausforderungen für die Abschätzung solcher Effekte gibt, die man bei der Bewertung der Ergebnisse berücksichtigen muss. Dies gilt insbesondere für die Studien, die den Versuch einer Quantifizierung von Potenzialen unternehmen. Ein erster genereller Aspekt dabei ist, dass natürlich alle auf Prognosen zukünftiger Entwicklungen beruhen. So werden in der Regel generelle Wachstumstrends, wie z. B. die Wachstumsprognosen des Internationalen Währungsfonds (IWF) zugrunde gelegt. Ein weiterer Faktor ist, dass die Berechnungen und Einschät-



zungen sehr stark von historischen Werten und/oder Erwartungen und weniger von aktuellen empirischen Beobachtungen gestützt werden. Abschließend hängen die Ergebnisse zudem stark von den verwendeten Bewertungsverfahren und Modellen ab. Dies soll bei der Analyse der Potenziale reflektiert werden.

PRODUKTIVITÄT UND FIRMENGRÜNDUNGEN**1.4.1**

Fast alle Schätzungen und Berechnungen insbesondere der gesamtwirtschaftlichen Potenziale beruhen auf der Annahme, dass Cloud Computing und Big Data zu Produktivitätssteigerungen und Unternehmensgründungen führen. Grundlegende Annahme ist, dass diese wie für die IKT als Ganzes schon mehrfach in Studien aufgezeigt zu wirtschaftlichem Wachstum und einer Steigerung des Bruttoinlandsprodukts (BIP) führen (z. B. Eicher/Strobel 2009; OECD 2004). Dementsprechend soll analysiert werden, welche Potenziale hier vermutet werden und welche Annahmen dem zugrunde liegen.

Grundsätzlich werden die Produktivitätssteigerungen auf zwei Effekte zurückgeführt: Kosteneinsparungen sowie die Geschäftsentwicklung, d. h. der Ausbau oder die Weiterentwicklung existierender Unternehmen. Zusammen mit den Neugründungen machen sie den wesentlichen Teil der wirtschaftlichen Effekte für Wachstum und Beschäftigung aus. Alle drei Effekte können sowohl bei Big Data als auch bei Cloud Computing beobachtet werden, wobei jedoch ihre Ausprägungen durchaus variieren können. Zu beachten ist, dass abhängig vom verwendeten Modell einige Studien daraus Produktivitätssteigerungen berechnen, während andere Wirkungen auf Beschäftigung oder Wachstum ableiten. Auffallend ist die unterschiedliche Datenlage: Während für Cloud Computing mehrere Studien vorliegen, sind für Big Data nur einzelne Daten vorhanden. Auch die Qualität der Daten unterscheidet sich deutlich, wobei dies unabhängig von der Zahl der Studien ist.

Für Big Data gibt es bisher nur wenige wesentliche Arbeiten. Die bekannteste von allen ist die von McKinsey aus dem Jahre 2011 (Manyika et al. 2011). Hier werden verschiedene Werte zu Potenzialen berechnet. Einer davon ist der Wert von 250 Mrd. Euro an jährlichen Einsparungen in den öffentlichen Verwaltungen innerhalb der EU-Mitgliedstaaten aufgrund von Produktivitätssteigerungen. Betrachtet man die zugrundeliegende Methodik, so wird deutlich, dass die Berechnungsgrundlage nicht vollkommen transparent oder etabliert ist und auf einem sehr holzschnittartigen Ansatz beruht (Manyika et al. 2011, S. 129 ff.). Im Gegensatz dazu setzt das von der OECD durchgeführte Projekt »New sources of growth«, das sich mit Knowledge-based Capital (KBC) beschäftigt, auf einen anderen Ansatz (OECD 2013a u. 2013b). Hier hat man durch die Auswertung verschiedener Datenbanken und Umfragen nachzuweisen versucht, welchen Einfluss Wissenskapital auf Produktivität hat. Dabei hat sich beispielsweise gezeigt,



dass in den USA Firmen, die ihre Entscheidungen wesentlich auf Datenanalyse stützen, eine 5 bis 6 % höhere Produktivität aufweisen (OECD 2013a, S. 18). Ausgehend von diesen und anderen Ergebnissen wird die Notwendigkeit abgeleitet, die entsprechenden Rahmenbedingungen für Big Data zu schaffen, um die Nutzung von Daten und Wissen als neue Quelle des ökonomischen Wachstums zu ermöglichen (OECD 2013a u. 2013b). So werden zwar auch Maßnahmen zur Ausbildung von qualifizierten Fachkräften gefordert, doch es wird nicht wirklich deutlich gemacht, dass letztlich kein automatischer Zusammenhang zwischen datengestützter Entscheidung und mehr Produktivität besteht. Denn nur wenn man Big Data zu nutzen weiß, führt dies letztlich zu mehr Produktivität, und dabei stellen sich neben Fachkräften noch weitere Herausforderungen auf betrieblicher Ebene, die mindestens ebenso gewichtig sind wie die regulatorischen Herausforderungen. Neben diesen generellen Arbeiten gibt es noch zwei weitere Arbeiten, die sich mit den ökonomischen Potenzialen von Big Data für das Vereinigte Königreich und Irland beschäftigen (Mohamed et al. 2012, 2013).

Für Cloud Computing existieren im Vergleich zu Big Data deutlich mehr Studien, wobei insbesondere die Arbeiten von Etro (2009 u. 2011) als Basis für einige weitere Studien dienen (z. B. Hogan et al. 2010). Etro berechnete dabei aber keine expliziten Produktivitätssteigerungen, sondern Firmengründungen, zusätzliches BIP-Wachstum sowie Beschäftigung, auf die hier später eingegangen wird. Er verwendete dazu ein dynamisch-stochastisches Gleichgewichtsmodell, für das er zwei unterschiedliche Szenarien (langsame und schnelle Diffusion) mit unterschiedlichen Ausprägungen in unterschiedlichen Industrien unterstellte. Grundlage der Berechnungen sind Kosteneinsparungen, die Etro basierend auf einer Analyse von IDC aus dem Jahr 2009 mit bis zu 50 % ansetzt (Etro 2011). Dies ist insofern kritisch, da in einer anderen Studie von IDC (Bradshaw et al. 2012), die im Auftrag der Europäischen Kommission durchgeführt wurde, die Mehrheit der befragten rund 1.000 Experten nur noch Kosteneinsparungen von zwischen 10 und 30 % erwarten. Dies zeigt deutlich, dass sich die Erwartungen im Laufe der Zeit deutlich verändert haben, was nicht untypisch ist in einer solchen Entwicklung. Darüber hinaus verdeutlicht es auch ein zweites Problem, nämlich dass die Werte vor allem auf Erwartungen und weniger auf empirischen Zahlen basieren. Diese liegen noch nicht in großer Zahl und vor allem nicht über längere Zeiträume hinweg vor. Auch Bradshaw et al. (2012, S. 47) berechnen ebenso wie Hogan et al. (2010 u. 2011) den kumulierten Wert der durch Cloud Computing zusätzlich entstandenen Wachstums- und Beschäftigungseffekte, ohne genauere Details zu ihren Produktivitätsannahmen offenzulegen. Die Ergebnisse werden im nächsten Abschnitt behandelt.

Vergleichbar zu den Produktivitätssteigerungen spielen die möglichen Kosteneinsparungen auch bei den Firmengründungen eine Rolle. Jedoch kommen hier je nach Technologie noch andere Aspekte wie Flexibilisierung oder Innovationsak-



tivitäten hinzu. Diese ermöglichen in der Regel, dass sich neue Firmen einfacher etablieren können, was sich wiederum positiv auf Wachstum und vor allem Beschäftigung auswirken soll. Zurzeit liegen entsprechende Berechnungen nur für Cloud Computing vor. Maßgeblich sind dabei die Arbeiten von Etro (2009 u. 2011) zur Zahl der Firmengründungen, auf denen alle weiteren Ansätze zur Bestimmung des ökonomischen Wertes beruhen (Hogan et al. 2010 u. 2011). Etro (2009 u. 2011) geht also davon aus, dass durch Cloud Computing und den verbundenen Kostensenkungen nicht nur die Betriebskosten, sondern auch die Markteintrittsbarrieren für Neugründungen über alle Branchen hinweg sinken. Unterschiede hängen hierbei insbesondere von der Intensität des verfügbaren Kapitals ab. Dies wiederum wird aus Sicht von Etro (2009) zu einer Welle von Neugründungen führen, die maßgeblich als Treiber für Beschäftigung und Wachstum wirken. Seinen Berechnungen zufolge entstehen in Europa durch Cloud Computing in einer mittelfristigen Perspektive (fünf Jahre) zwischen rund 83.000 (langsame Diffusion) und 431.000 (schnelle Diffusion) neue KMU zusätzlich zu den normalen Gründungen. Etro gibt zwar an, dass seine Berechnungen auf Daten von Eurostat basieren, nicht aber auf welches Jahr sie sich genau beziehen. Deutschland liegt dabei mit 7.600 bzw. 39.000 eher im Mittelfeld vergleichbar mit dem Vereinigten Königreich (6.700 bzw. 35.000). Deutlich größer sind die Zahlen in Italien (15.000 bzw. 81.000), Spanien (10.500 bzw. 55.000) und Frankreich (9.300 bzw. 48.000). Eine Erklärung dieser großen Unterschiede wird nicht von ihm diskutiert. Etro betont allerdings, dass die Auswirkungen auf die Beschäftigung im Vereinigten Königreich und in Deutschland trotzdem größer sind als in den drei anderen Ländern (Etro 2009).

Auf den Zahlen von Etro aufbauend versuchen Hogan et al. (2010) den kumulierten ökonomischen Wert dieser Entwicklungen für das BIP in den fünf größten Volkswirtschaften der EU zu berechnen. Danach liegt dieser in Deutschland trotz der geringeren Anzahl von Unternehmensgründungen höher (22,3 Mrd. Euro) als in Frankreich (17,7 Mrd. Euro), Italien (12,9 Mrd. Euro), Spanien (7,4 Mrd. Euro) oder dem Vereinigten Königreich (6,8 Mrd. Euro). Die Unterschiede werden mit unterschiedlichen Faktoren wie Skaleneffekten oder Adoptionsverhalten begründet, die aber nur qualitativ bewertet wurden. Zudem basieren die Berechnungen nur auf dem positiven Szenario von Etro (Hogan et al. 2010 u. 2011). Im Hinblick auf die einzelnen Branchen versucht Etro (2009) ebenfalls die Veränderungen zu berechnen. Dazu fasst er die wesentlichen Teile der Wirtschaft in fünf aggregierten Sektoren zusammen (Angaben in Klammern geben die Zahl der zusätzlichen Firmengründungen bei langsamer bzw. schneller Adoption in mittelfristiger Perspektive in Deutschland wieder): produzierendes Gewerbe (955 bzw. 4.931), Einzel- und Großhandel (2.301 bzw. 11.880), Hotel- und Gaststättengewerbe (873 bzw. 4.505), Transport und Kommunikation (462 bzw. 2.384) sowie Immobilien und unternehmensnahe Dienstleistungen (3.010 bzw. 15.542). Von der Verteilung zwischen den Sektoren entspricht dies



ungefähr der Entwicklung in den anderen Ländern. Die Unterschiede zwischen den Sektoren wirken im ersten Moment deutlich, sie sind aber, wenn man sie auf die Grundgesamtheit der einzelnen Sektoren bezieht, nicht so unterschiedlich wie vermutet. Denn das produzierende Gewerbe ist z. B. mit seiner eher industriellen Struktur stärker konzentriert als die unternehmensnahen Dienstleistungen, wovon beispielsweise Beratung, Steuerprüfung, Ingenieurbüros oder IT-Dienstleistungen fallen. Dementsprechend ist die Gesamtzahl an Unternehmen hier deutlich größer. Darauf aufbauend versuchen Hogan et al. wiederum den kumulierten Wert der Neugründungen in den Sektoren zu berechnen. Da sie die Zahlen jedoch teilweise neu berechnen und die Sektoren neu aggregieren, ohne dies genau zu erklären, ergeben sich daraus jedoch keine neuen Erkenntnisse.

Abschließend kann konstatiert werden, dass es durchaus konkrete Hinweise auf die positiven Wirkungen insbesondere von Cloud Computing, aber auch von Big Data auf die Produktivität sowie auf Unternehmensgründungen gibt. Wie groß jedoch diese Effekte tatsächlich sind, ist noch offen. Die Ergebnisse der Schätzungen hängen sehr stark von den verwendeten Daten, Modellen und Vorgehensweisen ab. Deutlich wird dabei, dass die wenigen Schätzungen sich sehr stark auf positive Grundannahmen beziehen, beispielsweise bezüglich der Kostensenkungen. Insgesamt beruhen diese vor allem auf Einschätzungen von Experten oder Marktteilnehmern und weniger auf empirischen Werten. Dementsprechend sind sie auch sehr stark von den positiven Erwartungen an die jeweilige Technologie geleitet. Ein anderer Punkt ist, dass zwar einige wenige Studien zumindest verschiedene Szenarien aufzeigen (Etro 2009), andere sich aber in ihrer Nachfolge dann auf die Nutzung der positiven Entwicklungen fokussieren (Hogan et al. 2010).

WACHSTUM UND BESCHÄFTIGUNG

1.4.2

Im Hinblick auf Big Data gibt es keine umfassenden Studien zu möglichen Wirkungen auf Wachstum und Beschäftigung, die Aussagen für Deutschland oder Europa ermöglichen. Neben den Aussagen von Manyika et al. (2011) bleiben damit nur die Studien des CEBR, die sich mit der Situation in Großbritannien und Irland beschäftigen (Mohamed et al. 2012 u. 2013). Die Berechnungen ergeben für Großbritannien einen kumulierten gesamtwirtschaftlichen Nutzen von rund 240 Mrd. britischen Pfund sowie 58.000 neue Stellen in der Zeit zwischen 2011 und 2017. Wesentliche Aspekte sind dabei Effizienz- und Innovationszunahme in existierenden Unternehmen sowie Unternehmensneugründungen. Interessant ist dabei, dass der berechnete Nutzen von Big Data höher angesetzt wird als der in einer ähnlichen CEBR-Studie für Cloud Computing (Hogan et al. 2010 u. 2011). Dies kann natürlich teilweise an den Besonderheiten im Vereinigten Königreich liegen, doch bestätigt die Einschätzung die Tendenz, das gesamtwirtschaftliche Potenzial von Big Data als besonders hoch einzuschätzen. Neben dem



Nutzen aus Unternehmensneugründungen wird vor allem großes Potenzial in Bereichen wie dem produzierenden Gewerbe, dem Handel oder unternehmensnahen Dienstleistungen, aber auch im öffentlichen Sektor wie z.B. in der Verwaltung oder im Gesundheitssystem gesehen. Diese Einschätzung stimmt mit den Einschätzungen von McKinsey (Manyika et al. 2011) überein. Gleichzeitig wird aber auch eine weiterführende methodische Herausforderung aufgezeigt. Während es in Ansätzen möglich ist, den Wert von Big Data für Unternehmen und mit Abstrichen für die öffentliche Verwaltung zu schätzen, war die Suche nach Ansätzen, welche die Potenziale für die Wissenschaft und wissenschaftlichen Daten erfassen können, bisher noch nicht erfolgreich (EK 2010).

Mit den Studien von Etro (2009 u. 2011) liegen Berechnungen vor, wie stark sich Cloud Computing auf das jährliche Wachstum des BIP auswirken könnte. Er benutzt dazu ein dynamisch-stochastisches Gleichgewichtsmodell (Dynamic Stochastic General Equilibrium Model, DSGE) mit einer endogenen Marktstruktur, um die Auswirkungen neuer Technologien zu ermitteln. Dies ist ein recht neuer Ansatz, auch wenn die DSGE-Modelle selbst heute beispielsweise in Zentralbanken als Ergänzung zu klassischen Vorhersagemodellen verwendet werden. Spätestens seit der letzten Finanzkrise ist ihr Nutzen aber umstritten (The Economist 2010a). Etro (2009) kommt zum Ergebnis, dass das zusätzliche Wachstum mittelfristig (fünf Jahre) bei langsamer Diffusion bei 0,1 % und bei schneller Diffusion bei 0,3 % liegen wird. Auch wenn dieser Wert auf den ersten Blick sehr klein wirkt, ist das angesichts der Größenordnung des BIP ein starker zusätzlicher Impuls. Jedoch muss er, wie Etro selbst anmerkt, mit Vorsicht gesehen werden.

Im Gegensatz zu Etro bemessen andere Studien den Wert nicht in einem zusätzlichen Wachstum des BIP, sondern berechnen, wie bereits erwähnt, den kumulierten Nutzen. Hogan et al. (2010 u. 2011) berechnen den Nutzen auf der Basis mehrerer Komponenten. Dazu zählen neben den schon beschriebenen Neugründungen die Nettokosteneinsparungen durch Verwandlung von Investitionskosten (CapEx) in operative Ausgaben (OpEx) und weitere Einsparungen abzüglich der Ausgaben für Clouddienste sowie Innovationen innerhalb bestehender Unternehmen. Für Deutschland ergibt sich für fünf Jahre ein kumulierter Wert von 221 Mrd. Euro für den Zeitraum von 2010 bis 2015. Im Vergleich zu den fünf anderen betrachteten Ländern Frankreich (162 Mrd. Euro), Vereinigtes Königreich (117 Mrd. Euro), Italien (150 Mrd. Euro) sowie Spanien (11 Mrd. Euro) erzielt Deutschland den Berechnungen zufolge den größten Nutzen. Dies liegt einerseits natürlich an der Größe der Volkswirtschaft, andererseits aber auch an anderen in die Berechnung eingeflossenen Faktoren wie z.B. der qualitativen Bewertung von Diffusionsraten in Ländern und Branchen (Hogan et al. 2010). Als wesentliche Treiberbranchen werden die Bereiche Finanzen, unternehmensnahe Dienstleistungen sowie Logistik und Handel gesehen (Hogan et al. 2011). Dies lässt sich teilweise nur bedingt mit den Ergebnissen von Etro vergleichen,

der eine andere Einteilung der Sektoren gewählt hat, obwohl Hogan et al. (2011) sich teilweise auf dessen Arbeiten stützen.

Auch Bradshaw et al. (2012) schätzen einen kumulierten Nutzenwert von Cloud Computing. Jedoch beziehen sich ihre Schätzungen auf den Zeitraum 2015 bis 2020 und unterscheiden zwischen einem No-Policy-Intervention- und einem Policy-driven-Szenario. Ersteres steht für eine eher langsame Diffusion. Letzteres für eine schnelle Diffusion. Auf Grundlage eines selbst entwickelten Wirkungsmodells kommen Bradshaw et al. zum Ergebnis, dass im No-Policy-Intervention-Szenario der kumulierte Wert für die gesamte EU bei 357 Mrd. Euro und beim Policy-driven-Szenario bei 920 Mrd. Euro. liegen würde. Letztlich geben sie aber auch zu bedenken, dass diese Berechnungen sehr vorsichtig zu bewerten sind, da das Modell noch Lücken aufweist (Bradshaw et al. 2012, S. 58 ff.).

Ähnlich wie bei den Wachstumswirkungen ist die Lage auch bei den Beschäftigungswirkungen schwer zu beurteilen. Grundsätzlich stellt sich die Frage, ob hier nur die zusätzlich geschaffenen Stellen berechnet oder ob auch mögliche Stellenverluste durch Cloud Computing berücksichtigt werden. Bei diesen würde es sich beispielsweise um Verluste in IT-Unternehmen und IT-Abteilungen handeln, die teilweise befürchtet werden. Bisher konnten diese jedoch weder beobachtet werden, noch ist es wahrscheinlich, dass dies angesichts des Mangels an Fachkräften in diesem Bereich im größeren Umfang geschieht. Auf lange Sicht stellt sich die Frage, inwieweit Effizienz- und Produktivitätssteigerungen sowie neue Innovationen und Geschäftsmodelle in den Anwenderbranchen zum Verlust von Jobs führen. Die Frage nach den Nettobeschäftigungseffekten wird von den vorliegenden Studien sehr unterschiedlich beantwortet.

Bradshaw et al. (2012) errechnen für ihre beide Szenarien für den Zeitraum 2015 bis 2020 mögliche Beschäftigungseffekte von 1,3 Mio. (no policy intervention) bis zu 3,8 Mio. (policy driven) neuen Arbeitsplätzen durch Cloud Computing in der EU. Die Frage nach möglichen Jobverlusten können sie mit ihrem Modell nicht beantworten (Bradshaw et al. 2012, S. 61). In den Studien des CEBR zu Big Data werden für den Zeitraum 2011 bis 2017 für das Vereinigte Königreich 58.000 neue Arbeitsplätze durch Unternehmensgründungen und die höhere Nachfrage nach Spezialisten berechnet (Mohamed et al. 2012, S. 6, 14 ff.). Die Frage nach möglichen Verlusten wird nicht thematisiert. In den Studien zu Cloud Computing wird das Thema jedoch adressiert. Während langfristig keine genauen Aussagen möglich sind, werden kurzfristig neugeschaffene Positionen als positive Nettobeschäftigungseffekte betrachtet (Hogan et al. 2010, S. 29 f.). Insgesamt berechnen sie in etwa spiegelbildlich zum ökonomischen Nutzen für den Zeitraum 2010 bis 2015 789.000 neue Stellen in Deutschland und somit die meisten neuen Stellen in Europa (Frankreich: 489.000; Italien: 456.000; Spanien: 392.000; Vereinigtes Königreich: 292.000). Anders sieht dies bei Etro (2011) aus, der bewusst das Thema der Nettobeschäftigungseffekte adressiert und sein



DSGE-Modell nutzt, um diese zu berechnen. Zwar bleibt die Reihenfolge der Staaten fast identisch, doch trotz der sehr positiven Grundannahmen liegen die Arbeitsplatzzahlen deutlich niedriger. So errechnet er für Deutschland für den Fünfjahreszeitraum im Fall einer schnellen Diffusion circa 121.000 neue Nettoarbeitsplätze und im Fall einer langsamen Diffusion nur rund 11.000 neue Nettoarbeitsplätze. Ähnlich niedriger fallen die Zahlen für die anderen Staaten aus (z. B. Frankreich circa 54.000 bzw. circa 5.000) (Etro 2011).

Letztlich weist dies auf eine Diskussion zu den Beschäftigungseffekten des Internets hin, die schon länger geführt wird. So legen die Ergebnisse einer McKinsey-Studie, die auf Befragungen von mehreren Tausend Unternehmen beruht, nahe, dass das Internet in den letzten 15 Jahren für jeden verlorenen Job 2,6 neue Jobs geschaffen hat (Pélessié du Rausas et al. 2011, S. 3). Zu ähnlichen Schlüssen gelangen auch andere Studien, die sich mit der Wirkung des Internets beschäftigen (z. B. Dean et al. 2012). Jedoch merken Hoorens et al. (2012, S. 65 ff.) in ihrer Studie, die sich mit den möglichen sozioökonomischen Effekten des sogenannten Future Internet beschäftigt, also dem Bündel aus Cloud Computing, Social Media, Mobile Computing und vielem mehr, kritisch an, dass dies wenig über die Qualität und Sicherheit dieser Jobs aussagt. Zwar haben Langzeitstudien gezeigt, dass die Einführung von Computern in Unternehmen in den 1970er und 1980er Jahren tendenziell zu einem Ersatz weniger qualifizierter Bürojobs durch höher qualifizierte Jobs geführt hat (z. B. Appelquist 2005), ob dies aber auch für die letzten zehn Jahre zutrifft, gilt es noch zu klären.

Insgesamt lässt sich festhalten, dass parallel zu den Produktivitäts- und Gründungseffekten auch die Wachstums- und Beschäftigungseffekte sowohl von Big Data als auch von Cloud Computing positiv erscheinen. Allerdings zeigt sich, dass die Prognosemethoden sehr unterschiedliche Qualität besitzen und in ihren Möglichkeiten begrenzt sind. Dies gilt insbesondere im Hinblick auf die Frage der Beschäftigungseffekte. Gerade hier wird aber die Notwendigkeit deutlich, die Folgen solcher und ähnlicher Entwicklungen auf Sicherheit und Qualität der Beschäftigung weiter zu erforschen.

INNOVATIONSFÄHIGKEIT

1.4.3

Produktivitätssteigerungen und Unternehmensgründungen als Basis gesamtwirtschaftlicher Potenziale sind eng mit Innovationen verbunden. Oftmals ermöglichen erst Innovationen Kostensenkungen oder die Gründung neuer Unternehmen, wobei es sich nicht nur um technische, sondern auch um Prozess- oder Marketinginnovationen handeln kann.

Sowohl Big Data als auch Cloud Computing für sich genommen sowie in Kombination bieten vielfältige Möglichkeiten, Innovationen zu ermöglichen oder zumindest zu erleichtern. Dies geht so weit, dass heute schon Verfahren getestet



werden, Kundenwünsche zu antizipieren, bevor diese selbst die Wünsche explizit formulieren können (Greif 2014). So ermöglicht beispielsweise Big Data eine deutlich höhere Kundenorientierung, die Möglichkeiten zur Entstehung spezialisierter Angebote und Anbieter beinhaltet. Gleichzeitig sollen Big-Data-Analysen in Verbindung mit neuen Produktionstechnologien (Industrie 4.0) dazu beitragen, die Massenproduktion zu flexibilisieren. Ziel ist es, Massenprodukte möglichst individualisiert anzubieten. Dies kann beispielsweise durch additive Produktion (Additive Manufacturing) geschehen (LaMonica 2013; Rotman 2013). Dabei werden 3-D-Drucker im industriellen Kontext genutzt, um individualisierte Produkte zu erzeugen. Additive Produktion kommt bisher nur im kleineren Maßstab, z. B. bei der Fertigung von Prothesen zum Einsatz, es gibt allerdings Pläne, den Ansatz auch in größerem Maßstab zu nutzen. Dadurch könnten entsprechend den Wünschen der jeweiligen Kunden Produkte als Ganzes oder durch Ergänzungen gezielt individualisiert werden. Damit würde aus dem Maker Movement tatsächlich eine industrielle Entwicklung, wenn auch anders als vielleicht von einigen erwartet, die die Entwicklung eher im Bereich von Nischenproduzenten sahen (Anderson 2012).

Cloud Computing ermöglicht gleichzeitig den Zugang zu Computerressourcen, die entweder früher nicht vorhanden waren oder die sich KMU oder Neugründer bis dahin nicht leisten konnten. Dies reicht von Kollaborationswerkzeugen, die es ermöglichen, Inhalte zu teilen oder Daten gemeinsam zu bearbeiten, über Werkzeuge zur Einbindung vieler Akteure (z. B. Kunden, Lieferanten), über die Entwicklung, Planung und Durchführung von Projekten und die Steuerung von Wertschöpfungsketten bis hin zu Anwendungen wie cloudbasierten Werkzeugen zur Herstellung von Produkten, beispielsweise mit 3-D-Druckern. Im Kapitel V.2 werden die vielfältigen Innovationspotenziale im Zusammenhang mit Big Data und Cloud Computing ausführlicher behandelt.

Sowohl Big Data als auch Cloud Computing bieten schon für sich allein genommen ein großes Potenzial, die Innovationsfähigkeit der Gesellschaft und Wirtschaft zu erhöhen. Diese Wirkungen wurden auch für andere Anwendungen im Bereich der Informations- und Kommunikationstechnik nachgewiesen (DIW 2008). Doch gerade die Möglichkeit, Big-Data-Analysen als cloudbasierte Lösungen zu nutzen, würde diesen Effekt weiter verstärken. Wie bereits gezeigt, beinhaltet das Konzept eine Vielzahl neuer Möglichkeiten für Innovationen in vielen Anwendungsfeldern. Cloud Computing bietet auch KMU sowie Neugründern die Möglichkeit, Big-Data-Analysen flexibel und kostengünstig durchzuführen. Zu bedenken ist allerdings, dass die Nutzung von cloudbasierten Big-Data-Lösungen mit einigen Herausforderungen verbunden ist. So stellen der Transfer großer Datenmengen und der Zugang zu Daten gerade für diese Zielgruppe eine nicht leicht zu überwindende Hürde dar. Auch generell ist daher anzumerken, dass das Potenzial zur Steigerung der Innovationsfähigkeit zwar gegeben ist, dass



es aber noch eine Reihe von technischen, betrieblichen und strukturellen Herausforderungen gibt, die adressiert werden müssen, um dieses Potenzial auch zu realisieren.

NACHHALTIGKEIT**1.4.4**

In vielen Studien zum Thema Green IT wird insbesondere Cloud Computing ein großes Potenzial im Hinblick auf Energie- und Ressourceneffizienz zugesprochen. Grundlage dafür ist, dass Cloud Computing und zugrundeliegende Technologien wie Virtualisierung eine Reihe von Möglichkeiten zur Effizienzsteigerung mit sich bringen. So führt Cloud Computing in der Regel zu einer besseren Auslastung der verfügbaren Infrastrukturen. Vor allem Rechenzentren können durch Virtualisierung ihre Energieeffizienz verbessern. Zudem, so die Annahme, führt die bessere Auslastung zu einem geringeren Bedarf an neuen Servern und Rechenzentren. Dies wiederum würde sowohl den Ressourcenverbrauch, z. B. im Hinblick auf seltene Erden, als auch die Elektroschrottmenge reduzieren. Darüber hinaus bietet die Verwendung von Cloud Computing auch Potenziale zu Einsparungen auf der Anwenderseite. Hier kann es durch die Nutzung alternativer Endgeräte wie Thin Clients oder Tablets anstelle von Desktopcomputern ebenfalls zu Einsparungen beim Energie- und Ressourcenverbrauch kommen, da solche Endgeräte in der Regel energieeffizienter sind und weniger Materialien verbaut werden (u. a. EK 2012a; Stobbe et al. 2009).

Sowohl auf deutscher als auch auf europäischer Ebene wird das Thema Green IT mit Förderprogrammen (z. B. IT2Green) adressiert. Berechnungen zeigen, dass in Deutschland in den kommenden Jahren im Bereich der Rechenzentren circa 30 bis 50 % des Energiebedarfs eingespart werden könnte (Hintemann/Fichter 2012, S. 2 ff.). Zudem belegen Studien, dass einige erwartete Entwicklungen wie z. B. der Rückgang kleinerer Serverinstallationen und kleinerer Rechenzentren tatsächlich eingetreten sind. Jedoch zeigen sie auch, dass es neben dem Trend zu Cloud Computing weitere Faktoren gibt, die zu mehr Energie- und Ressourceneffizienz führen. Einerseits sind das z. B. Fortschritte im Bereich der Server- oder der Versorgungstechnik, andererseits aber auch externe Faktoren wie die Finanzkrise. Inwieweit sich die positive Entwicklung bei Energie- und Ressourceneffizienz fortschreiben lässt, bleibt daher offen, vor allem da auch die Gefahr besteht, dass langfristig durch Reboundeffekte die Einsparungen wieder aufgezehrt werden. Ein Grund dafür kann beispielsweise der Gebrauch mehrerer neuer Endgeräte anstelle eines herkömmlichen Desktopcomputers sein, aber auch neue Anwendungen wie beispielsweise Big Data, die vor allem durch ihre Datenintensität den Energie- und Ressourcenverbrauch steigern.

Es muss davon ausgegangen werden, dass der generelle Energie- und Ressourcenverbrauch im Bereich der IKT aufgrund der zunehmenden Digitalisierung fast

aller Lebensbereiche weiter zunehmen wird, sodass die Verbreitung von Cloud Computing bestenfalls zu einer Abflachung dieser Wachstumskurve führen wird. Dies liegt vor allem daran, dass Rechenzentren nur für einen relativ kleinen Teil am gesamten Stromverbrauch verantwortlich sind (Stobbe et al. 2009, S. 68).

TRANSPARENZ UND PARTIZIPATION

1.4.5

Grundsätzlich werden zwei Aspekte – in Bezug vor allem auf Big Data – mit Potenzialen für Transparenz und Partizipation gesehen. Beim ersten Aspekt geht es um die Idee, dass die Nutzer selbst entscheiden (können) sollen, ob und wem und wie viele Informationen sie von sich preisgeben. Viele Konzepte gehen sogar so weit, dass die Nutzer selbst für die Preisgabe von Informationen bezahlt werden sollen. Grundlegende Idee ist, dass Onlinemarktplätze es Nutzern erlauben sollen, ihre Daten zu hinterlegen. Dort könnten Unternehmen ihre Interessen und Datenwünsche bekunden, und der Nutzer entscheidet dann individuell, zu welchem Preis er welche Daten weitergeben will. Eine weitere Idee ist, dass sich beispielsweise wissenschaftliche Projekte bei diesen Plattformen bewerben und Nutzer entscheiden können, ob sie diesen ihre Informationen kostenfrei überlassen. Damit sollen Nutzer verfolgen können, wer welche Daten benutzt, und bestimmen, wer ihre persönlichen Daten wie nutzen darf. Die Ideen dazu existieren schon seit mehreren Jahren, wobei in den letzten beiden Jahren das Interesse merklich zugenommen hat. Sichtbarste Beispiele dafür sind die Gründung von Datacoup, welches explizit eine solche Idee verfolgt, sowie die Ankündigung von reputation.com, ähnliche Services anbieten zu wollen (Simonite 2013, 2014).

In Deutschland wird im Rahmen des Trusted-Cloud-Programms im Projekt MIA (Marktplatz für Informationen und Analysen) auch an solchen Modellen geforscht. Dabei zeigt sich eine Reihe von Problemen. Dazu zählen die hohe Volatilität des Marktes, bei dem der Preis in verschiedenen Kontexten sehr stark schwanken kann, die Überforderung der Nutzer, die oftmals den Wert ihrer Daten und die Konsequenzen von Entscheidungen nicht einschätzen können oder beispielsweise auch weiter führende rechtliche Aspekte wie die Frage der Kontrolle und der Durchsetzbarkeit von Weitergabeverboten und Ähnlichem (Acquisti 2010; Acquisti et al. 2013). Zudem erschweren die schon existierenden Marktstrukturen einen solchen Systemwechsel, da insbesondere Anbieter angesichts funktionierender Modelle wohl nicht geneigt sind, diese zu verändern. Abschließend stellt sich auch die übergeordnete Frage, ob eine solche Ökonomisierung von Privatheit gesellschaftlich gewünscht bzw. wünschenswert ist.

Beim zweiten Aspekt geht es um mit Open Data verbundene Ideen, die es Privatpersonen ermöglichen sollen, Einblick in Daten und Prozesse der öffentlichen Verwaltung zu nehmen. Big Data würde dabei vor allem zur Analyse von Daten dienen und die Entwicklung von neuen Lösungsansätzen oder die Identifikation



von Handlungs- und Verbesserungsmöglichkeiten unterstützen. Letzten Endes wäre das Ziel dieser als Open Government bezeichneten Idee, Bereiche der Verwaltung und Regierung sowohl Privatpersonen als auch der Wirtschaft zu öffnen. Einerseits sollen dadurch Prozesse vereinfacht werden, andererseits aber auch eine größere Möglichkeit zur Partizipation an politischen Prozessen geschaffen werden (Janssen et al. 2012; Klessmann et al. 2014). Die Potenziale dieser Entwicklungen werden dabei in verschiedenen Studien diskutiert (z. B. Golliez et al. 2012; Huber et al. 2013). Da dies aber sowohl im Hinblick auf die verwendeten Technologien und Verfahren wie Partizipationsportale, Bürgerhaushalte oder Ähnliches als auch im Hinblick auf die damit verbundenen Potenziale über Big Data hinausgeht, soll hier nur auf den genannten Aspekt der Nutzung von Big-Data-Tools durch Bürger hingewiesen werden.

ANWENDUNGSPOTENZIALE

2.

Die Vielfalt der Anwendungspotenziale von Big Data lässt sich unter Bezugnahme auf die Bedarfswelder der Hightech-Strategie der Bundesregierung veranschaulichen. Entsprechend werden aktuelle Anwendungen von Big Data sowie Überlegungen zur zukünftigen Anwendung in den Feldern Klima/Energie (Kap. V.2.1), Gesundheit/Ernährung (Kap. V.2.2), Mobilität (Kap. V.2.3), Sicherheit (Kap. V.2.4) und Kommunikation (Kap. V.2.5) diskutiert. Besonderes Augenmerk wird dabei auf die Kombination von Big Data und Cloud Computing gelegt. Für jedes Bedarfsweld wird zunächst ein Überblick über mögliche Anwendungspotenziale gegeben, und anschließend werden mögliche zukünftige Entwicklungen für ein bis zwei ausgewählte Anwendungspotenziale mithilfe von Miniszenarien antizipiert.

KLIMA/ENERGIE

2.1

Im Hinblick auf das Bedarfsweld Klima/Energie hat sich die Bundesregierung eine Reihe von Zielen gesetzt. Einerseits soll die Wissensbasis im Hinblick auf den Klimawandel erweitert und die Anwendung von Klimaschutztechnologien und Anpassungslösungen vorangetrieben werden. Andererseits soll die Energieversorgung auf eine nachhaltigere Basis gestellt werden. Für Big Data gibt es in diesem Zusammenhang zahlreiche Anwendungspotenziale. Vor allem in den Bereichen Klimaforschung, Smart Cities, Smarte Energienetze und Umweltschutz spielt Big Data bereits heute eine wichtige Rolle.

Im Bereich der Klimaforschung lassen sich durch die Analyse großer Datenmengen aus unterschiedlichen Quellen Vorhersagen zur Entwicklung des Klimas verbessern bzw. werden solche Vorhersagen überhaupt erst möglich. Wissen-



schaftler der Aristoteles Universität von Thessaloniki setzten sich beispielsweise das Ziel, die Entwicklung der Konzentration des bodennahen Ozons in Europa über die nächsten 100 Jahre vorherzusagen (Katragkou et al. 2011). Um die rechenintensive Simulation durchführen zu können, wurde auf die IT-Infrastruktur des Scientific Computer Centers der Aristoteles Universität von Thessaloniki zurückgegriffen. Auch die US-amerikanische Bundesbehörde für zivile Luft- und Raumfahrt NASA greift auf eine äußerst leistungsfähige IT-Infrastruktur zurück, um Simulationen im Hinblick auf die Entwicklung des Klimas durchführen zu können (Mangelsdorf 2012). Kürzlich versuchte die NASA, das Klima auf der Erde mit einer relativ hohen Auflösung für zwei Jahre vorzuberechnen. Der verwendete Hochleistungsrechner »Discover« hat eine Speicherkapazität von 37 Petabyte. Am Beispiel von Hurrikan Sandy lässt sich der Nutzen von Wettervorhersagen auf Grundlage von Big-Data-Analysen verdeutlichen (Wakeman 2012). Dank der Analyse großer Datenmengen ließ sich der Weg des Hurrikans, bereits drei Tage bevor der Hurrikan in den Vereinigten Staaten auf Land traf, relativ genau vorhersagen und damit ein noch größeres Unglück vermeiden. Der Einsatz von Big-Data-Analysen im Zusammenhang mit dem Hurrikan Sandy wird im Rahmen eines Miniszenarios detaillierter behandelt.

Auch im Bereich Smart Cities gibt es im Hinblick auf Big Data vielversprechende Anwendungspotenziale, vor allem im Hinblick auf die Bedarfsfelder Klima/Energie und Mobilität. Laut IBM (2012) zeichnen sich intelligente Städte vor allem durch einen hohen Grad an Nachhaltigkeit und Wirtschaftlichkeit aus, der durch die kontinuierliche Erfassung und Analyse von umfangreichen Datenmengen erzielt wird. Ansätze im Hinblick auf die Optimierung der Verkehrsführung in Smart Cities werden im Kapitel V.2.3 adressiert. An Ansätzen zur Verbesserung der Energieversorgung wird beispielsweise im Rahmen eines Smart-City-Projekts unter der Leitung von Siemens geforscht (Pease 2013). Eine Rolle spielen dabei neben Themen wie lokale Energieversorgung und intelligente Stromnetze (Smart Grid) auch verschiedene Automatisierungstechniken. Laut Fitze (2013) wird allein das Volumen des Marktes für Smart Grids im Jahr 2020 bei fast 400 Mrd. Euro liegen.

Aus der zunehmenden Nutzung regenerativer Energiequellen, die ein zusätzliches volatiles Element in das Stromnetz einführen, ergeben sich für die Erzeugung und Verteilung von Strom neue Herausforderungen. Einerseits produzieren immer mehr Verbraucher auch selbst Strom, der zum Teil ins Stromnetz eingespeist werden soll, und andererseits ist das Angebot an Strom mehr und mehr von Wind und Sonnenschein abhängig. Damit das Energienetz stabil bleibt, müssen Produzenten und Verbraucher koordiniert werden. Big-Data-Analysen spielen bei der Koordination in intelligenten Stromnetzen eine immer wichtigere Rolle. Die Daten, die analysiert werden, sind zwar strukturiert, stammen aber von sehr vielen Quellen und müssen, um die Stabilität des Netzes nicht zu gefährden, in



Quasiechtzeit verarbeitet werden (Gomes 2011). Strom muss entweder gespeichert oder umgehend dorthin transportiert werden, wo gerade Bedarf ist (Bleyle 2013, S.27). Die Daten werden in der Regel mittels Smart Meter erfasst. Zum Speichern der Daten bieten sich Cloudlösungen an, da die Daten dann für Analysen zentral zur Verfügung stehen. Das Potenzial entsprechender Ansätze ist beispielsweise im Zusammenhang mit Elektroautos gut erkennbar, da Akkus genau dann aufgeladen werden können, wenn es im Netz einen Stromüberschuss gibt.

Auch im Bereich Umweltschutz eröffnet Big Data neue Möglichkeiten. Auf Grundlage der Ergebnisse von Big-Data-Analysen lassen sich einerseits umweltfreundlichere Herstellungsmethoden und Produkte entwickeln und andererseits gefährdete Ökosysteme schützen. Beispielsweise wird die Herstellung von biologisch abbaubarem Plastik von einem Forscherteam der Ludwig-Maximilians-Universität in München nicht nur auf Grundlage von zahlreichen Versuchsreihen weiterentwickelt, sondern auch unter Berücksichtigung der Ergebnisse von Berechnungen, die durchgeführt wurden, um relevante Reaktionsmechanismen besser zu verstehen (Birkenheuer et al. 2011). Am portugiesischen Labor für Bauwesen wird Big Data eingesetzt, um die gefährdete Aveiro Lagune besser verstehen und schützen zu können (Rodrigues et al. 2009). Beide Projekte greifen auf eine leistungsstarke IT-Infrastruktur zurück, die im Rahmen der EGI bereitgestellt wird.

MINISZENARIO: HURRIKAN SANDY

Hurrikan Sandy hat im Herbst 2012 in der Karibik und in den USA zahlreiche Menschenleben gekostet und verheerende Schäden verursacht. Allein in den USA kamen 113 Menschen ums Leben. Mit einem auf rund 50 Mrd. US-Dollar geschätzten Schaden war Hurrikan Sandy nach Hurrikan Katrina, der im Jahr 2005 an der Golfküste enorme Schäden anrichtete, der zweitteuerste Hurrikan in der US-amerikanischen Geschichte (Konkel 2012). Aufgrund der Verfügbarkeit von relativ genauen Vorhersagen über den Weg und die Stärke des Hurrikans konnte ein noch größeres Unglück vermieden werden (Konkel 2012). Das Hurrikanzentrum der US-amerikanischen Wetter- und Ozeanografiebehörde NOAA konnte bereits drei Tage, bevor Sandy am Abend des 29. Oktobers das Festland der Vereinigten Staaten erreichte, mit einer Ungenauigkeit von nur rund 15 km vorhersagen, wo der Hurrikan an der Küste von New Jersey auf Land treffen würde (Wakeman 2012). Bereits am Morgen des 26. Oktobers gab das Hurrikanzentrum an, dass der Hurrikan in der Nacht vom 29. auf den 30. Oktober das US-amerikanische Festland erreichen werde. Die Genauigkeit war vor allem deshalb beeindruckend, da der Hurrikan im Hinblick auf Größe und Stärke ein nahezu einzigartiges Wetterereignis darstellte.

Die NOAA verwendete für ihre Berechnungen hochkomplexe Modelle und eine leistungsfähige IT-Infrastruktur (Jackson 2012). Laut Jackson (2012) werden für

die Vorhersage von Wetterereignissen im Allgemeinen und Hurrikans im Speziellen Modelle verwendet, die Informationen über Ereignisse in der Vergangenheit, über die Physik der Atmosphäre und über aktuelle und zukünftige Wetterbedingungen kombinieren. Relevant sind unter anderem die sich permanent ändernden Windgeschwindigkeiten, Druckunterschiede sowie Luft- und Wassertemperaturen (Johnson 2012). Die Informationen, die mit geografischen Informationssystemen zusammengeführt werden müssen, stammen beispielsweise von Satelliten, Wetterbojen oder Flugzeugmessungen (Konkel 2012). Die Mitarbeiter des Hurrikanzentrums sind also einerseits mit umfangreichen Datenmengen konfrontiert, deren rasche Auswertung kritisch ist, und andererseits mit der Tatsache, dass die Daten polystrukturiert sind. Damit sind die drei zentralen Charakteristika von Big-Data-Problemen gegeben.

Die Genauigkeit der Ergebnisse hängt nicht nur von der Genauigkeit der verfügbaren Informationen ab, sondern auch von der Tauglichkeit des verwendeten Modells. Aus diesem Grund kommen bei der Vorhersage von Wetterereignissen in der Regel mehrere Modelle zum Einsatz. Die eigentliche Vorhersage entspricht dann dem Durchschnitt der Ergebnisse (Konsensprognose) (Jackson 2012). Die Berechnungen werden in regelmäßigen Abständen mit aktualisierten Daten wiederholt (Johnson 2012; Konkel 2012). Beim Hurrikanzentrum der NOAA werden für Simulationen im Zusammenhang mit Hurrikans aktuell 37 verschiedene Modelle verwendet. Berücksichtigt werden muss allerdings, dass nicht alle Modelle für dieselbe Fragestellung geeignet sind. Während einige Modelle den Weg eines Hurrikans vorhersagen, sind andere beispielsweise auf die Vorhersage der Hurrikanstärke spezialisiert. Außerdem sind einige Modelle für kurzfristige Vorhersagen geeignet, während andere bei längerfristigen Betrachtungen besser abschneiden. Die Anforderungen an die Rechenleistung unterscheiden sich von Modell zu Modell teils erheblich (Jackson 2012). Im Hinblick auf die erforderliche Rechenleistung setzt NOAA auf den Hochleistungsrechner »IBM Bluefire« mit einer Leistung von bis zu 76 Teraflops (Konkel 2012). Der Einsatz von Hochleistungsrechnern ermöglicht zunehmend höhere Auflösungen bei der Vorhersage von Wetterereignissen.

Die zentrale Herausforderung bei der Vorhersage von Hurrikans liegt in der Tauglichkeit der verwendeten Modelle. Das Verständnis von komplexen Wetterereignissen wie Hurrikans reicht noch nicht aus, um auf Grundlage einzelner Modelle zuverlässige Vorhersagen zu machen. Die deutlichen Verbesserungen, die in den letzten Jahren im Hinblick auf die Genauigkeit der Vorhersage und den Vorhersagezeitraum möglich waren, zeigen allerdings deutlich, welches Anwendungspotenzial für die Vorhersage von Wetterereignissen in der Durchführung von Big-Data-Analysen liegt. Konkel (2012) betont beispielsweise, dass die Vorhersagen im Hinblick auf den Ort, an dem Hurrikan Sandy auf Land traf, zehnmal genauer waren als vergleichbare Vorhersagen in den 1990er Jahren. Aufgrund der relativ kurzen Vorhersagezeiträume und der zunehmenden Kom-



plexität der Modelle, stellt auch die Rechenleistung eine Herausforderung dar. Herausfordernd ist darüber hinaus die Interpretation der Berechnungsergebnisse, die nur mit Erfahrung und einschlägigem Fachwissen geleistet werden kann.

GESUNDHEIT/ERNÄHRUNG

2.2

Das Bedarfswelt Gesundheit/Ernährung zielt vor allem auf die Entwicklung besserer Diagnoseverfahren und Therapien ab, aber auch Ansätze zur Prävention und zur Kostensenkung werden in den Mittelpunkt gestellt. Die Bundesregierung hebt im Zusammenhang mit diesem Bedarfswelt die Bedeutung der individualisierten Medizin sowie telemedizinische Therapieformen besonders hervor. Vor allem im Hinblick auf die Erforschung von Krankheitsursachen und Behandlungsmethoden, die Epidemiologie sowie im Bereich der individualisierten Medizin spielt die Analyse von Big Data bereits heute eine zentrale Rolle und bietet zahlreiche Anwendungspotenziale.

Im Bereich der Erforschung von Krankheitsursachen und Behandlungsmethoden lassen sich durch die Analyse großer Datenmengen aus unterschiedlichen Quellen wichtige Erkenntnisse gewinnen. Untersuchungen in diesem Bereich beschäftigen sich mit der Suche nach und der Identifikation von Krankheitserregern (z. B. Vries et al. 2011) sowie der Neu- und Weiterentwicklung von Medikamenten (z. B. Neumann et al. 2010) oder Verfahren wie z. B. im Bereich der Epilepsie.

Im Bereich der Epidemiologie etwa erlaubt die Analyse umfangreicher Datenmengen eine deutlich schnellere und genauere Bestimmung von Infektionsherden für bestimmte Krankheiten, als dies in der Vergangenheit möglich war. Durch die frühzeitige Erkennung von Infektionsherden kann in der Regel die Ausbreitung einer Infektionskrankheit eingedämmt oder verhindert werden. Ein Beispiel, bei dem der Fokus auf der Eindämmung von Malaria in Kenia liegt, ist im Kapitel IV.1.3 beschrieben. Neben Mobilfunkdaten, wie in diesem Beispiel, können Daten aus sozialen Netzwerken oder Daten aus Anwendungen auf mobilen Endgeräten zur Untersuchung der Ausbreitung von Krankheiten verwendet werden (z. B. Kamel Boulos et al. 2010). Die Analyse der Kontaktdaten Einzelner erlaubt unter bestimmten Voraussetzungen die Identifikation sogenannter »Super-Spreader«, die für die Begrenzung der Ausbreitung von Infektionskrankheiten besonders relevant sind (Galvani/May 2005; Sun et al. 2013).

Ein weiterer Bereich, in dem Big Data eine zentrale Rolle spielt, ist die personalisierte Medizin (Murdoch/Detsky 2013). Personalisierte Medizin führt zur Individualisierung der Medizin unter Berücksichtigung der Eigenschaften von einzelnen Menschen. Big Data erlaubt die Verarbeitung der umfangreichen Datenmengen, die bei einzelnen Personen anfallen. Einblicke in die Gene von Individuen erlauben beispielsweise die Optimierung von Behandlungen und Medikamenten, aber auch individuelle Ernährungs- oder Lebensstilempfehlungen. Im Rahmen des



Future-and-Emerging-Technologies-Flagship-Projekts »ITFoM« (IT Future of Medicine) beschäftigt sich ein internationales Konsortium mit Fragen zur Zukunft der Medizin und wie sie durch Informationstechnik unterstützt werden kann. Die personalisierte Medizin spielt dabei eine zentrale Rolle. Ein verwandtes Thema, bei dem sowohl Big Data als auch Cloud Computing eine wichtige Rolle spielen, ist die Überwachung des menschlichen Körpers mithilfe von Sensoren. Beispielsweise arbeitet Google derzeit an einer Kontaktlinse, die den Blutzucker des Trägers kontrollieren und ihn gegebenenfalls warnen oder die Daten an andere Geräte senden kann (Spiegel Online 2014). So oder ähnlich gewonnene Daten können auch für andere Zwecke verwendet werden, beispielsweise zur individuellen Krankheitsvorhersage. Cloud Computing spielt in diesem Zusammenhang nicht nur im Hinblick auf den Zugriff auf die Analysefunktionalität eine wichtige Rolle, sondern auch im Hinblick auf die Zusammenführung der Daten von Individuen.

Big Data hat außerdem das Potenzial, Entscheidungen von Ärzten und anderen Personen im Gesundheitswesen zu unterstützen. Wie im Kapitel I.1 erwähnt, sorgte in der Vergangenheit vor allem der Einsatz von IBMs Watson in diesem Zusammenhang für Aufsehen. Big-Data-Analysen unterstützen dabei nicht nur die Diagnose von Erkrankungen, sondern auch die Auswahl der passenden Therapie (Upbin 2013). Der Einsatz von Watson im Gesundheitswesen wird im Rahmen des folgenden Miniszenarios detaillierter behandelt. In eine ähnliche Richtung geht ein Projekt von Siemens, bei dem eine durchgehende elektronische Krankenakte in der Cloud geführt wird. Der Ansatz erlaubt den direkten Zugriff auf Behandlungsverläufe sowie deren Analyse und soll so Verbesserungen in der Behandlung ermöglichen (Elsener 2013).

MINISZENARIO: WATSON IM GESUNDHEITSWESEN

Watson ist ein System aus dem Bereich der künstlichen Intelligenz, das von IBM im Rahmen des Forschungsprojekts »DeepQA« entwickelt wurde. Das System zielt auf die Beantwortung von Fragen ab, die in natürlicher Sprache eingegeben werden können. Seine Überlegenheit gegenüber Menschen zeigte Watson bereits 2011 in der Fernsehquizshow »Jeopardy!« (Best 2013). Nach dem eindrucksvollen Auftritt im Fernsehen wurde das System weiterentwickelt und stieß bald, vor allem im Gesundheitswesen, auf großes Interesse (Upbin 2013). Bekannte Anwender von IBMs Watson sind der Verwaltungsdienstleister WellPoint (IBM 2013b) und das Memorial Sloan-Kettering Cancer Center (IBM 2013a).

Nur 20 % des Wissens, das Ärzte für Diagnosen und Entscheidungen über die Behandlung von Patienten verwenden, ist evidenzbasiert (Steadman 2013). Eine von fünf Diagnosen ist falsch oder unvollständig, und allein in den USA gibt es 1,5 Mio. Medikationsfehler pro Jahr (IBM o. J.). Die Informationen, die Ärzte für bessere Entscheidungen benötigen, sind bereits verfügbar. Medizinische Fachzeitschriften veröffentlichen jeden Tag neue Behandlungsmethoden und



Forschungsergebnisse. Auch die Krankengeschichte von Patienten liefert Hinweise. Die Menge an medizinischen Daten verdoppelt sich alle fünf Jahre, und ein Großteil der Daten ist in Form von natürlichsprachlichen Texten vorhanden. 81 % der Ärzte verbringen eigenen Angaben zufolge nur maximal fünf Stunden pro Monat mit dem Lesen von Fachzeitschriften (IBM o. J.). Laut IBM ist Watson besser in der Lage, die umfassenden und rasch wachsenden Daten zu berücksichtigen (Steadman 2013). Der Schwerpunkt liegt derzeit im Bereich der Onkologie. Watson kennt bereits mehr als 600.000 Fakten, mehr als 2 Mio. Seiten einschlägiger wissenschaftlicher Zeitschriften. Außerdem kann Watson auf 1,5 Mio. Krankengeschichten zurückgreifen. Laut WellPoint diagnostiziert Watson Lungenkrebs in 90 % der Fälle korrekt, im Vergleich dazu schaffen das Ärzte nur in 50 % der Fälle (Steadman 2013; Upbin 2013).

Watson analysiert umfangreiche Textsammlungen, um Ärzten bei Diagnosen und Entscheidungen über die Behandlung von Patienten zu helfen. Zur Erstellung einer Diagnose muss der Arzt dem System die Symptome und weitere zugehörige Faktoren beschreiben. Watson ist in der Lage, natürliche Sprache zu verarbeiten und kennt medizinische Fachbegriffe. Laut IBM (o. J.) ermittelt Watson zunächst die wichtigsten Informationen der Eingabe und durchsucht dann die Patientendaten nach relevanten Fakten über die Familiengeschichte, die aktuelle Medikation und weitere Bedingungen. Watson kombiniert diese Informationen mit aktuellen Befunden aus Untersuchungen und Diagnosegeräten und analysiert dann alle verfügbaren Daten, um Hypothesen zu generieren und zu überprüfen.

Watson trägt allerdings nicht nur zur Verbesserung in den Bereichen Diagnose und Behandlung bei, sondern bringt auch aus wirtschaftlicher Sicht Vorteile. Laut WellPoint werden in den USA derzeit 30 % des Budgets für das Gesundheitswesen nicht optimal eingesetzt (Steadman 2013). Watson soll hier für Verbesserungen sorgen können. Interessant ist Watson unter anderem aus Kostensichtspunkten auch deshalb, weil das System als Clouddienst genutzt werden kann (Steadman 2013; Upbin 2013).

Da Watson mit umfassenden, polystrukturierten Daten arbeitet, sind zwei der drei zentralen Charakteristika von Big-Data-Problemen gegeben. Während die Verarbeitungsgeschwindigkeit von Watson bei den Auftritten in der Fernsehquizshow »Jeopardy!« kritisch war, ist sie es beim Einsatz im Gesundheitswesen nicht in gleichem Ausmaß. Watson ist in der Lage, spätestens nach wenigen Minuten Antworten auf Fragen zu geben (Steadman 2013). Zumeist antwortet das System jedoch in weniger als drei Sekunden. Durch den Einsatz von UIMA und Hadoop kann eine große Anzahl von Dokumenten parallel verarbeitet werden (Epstein et al. 2012).

Eine der zentralen Herausforderungen ist die Akzeptanz des Systems im Gesundheitswesen. Einerseits muss IBM potenzielle Nutzer überzeugen, dass die

verwendete Datenbasis sowie die generierten Antworten vertrauenswürdig sind. Andererseits stellt die Integration eines neuen Systems in die etablierten Abläufe im Gesundheitswesen eine Herausforderung dar (Roop 2011).

MOBILITÄT

2.3

Im Hinblick auf das Bedarfsweld Mobilität geht es der Bundesregierung einerseits um eine schnelle, sichere und komfortable und andererseits um eine ressourcenschonende Beförderung von Menschen und Gütern. Neben der Entwicklung von neuen Antriebssystemen zielt das Bedarfsweld vor allem darauf ab, Verkehrsinfrastrukturen intelligenter zu machen. Im Hinblick auf das Bedarfsweld Mobilität gibt es Anwendungspotenziale für Big Data. Vor allem in den Bereichen autonome Fahrzeuge und intelligente Transportsysteme spielt Big Data bereits heute eine Rolle.

Die Forschung im Hinblick auf autonome Fahrzeuge wird seit vielen Jahren intensiv vorangetrieben. Während der Fokus zunächst darauf lag, Fahrzeuge in die Lage zu versetzen, mithilfe verschiedener Sensoren ihre Umgebung wahrzunehmen, ein Ziel anzusteuern und Kollisionen auf dem Weg zu vermeiden, spielte in den letzten Jahren die Kommunikation mit anderen Fahrzeugen und der Verkehrsinfrastruktur (Car-to-X-Kommunikation) eine immer bedeutendere Rolle (Mearian 2014). Im Rahmen des sim^{TD}-Projekts wurde unlängst in einem großen Feldversuch im Rhein-Main-Gebiet gezeigt, dass Car-to-X-Kommunikation prinzipiell reif für den Alltagseinsatz ist (sim^{TD}-Konsortium 2013). Die Durchführung von Big-Data-Analysen im Zusammenhang mit dem sim^{TD}-Projekt wird im Rahmen eines Miniszenarios detaillierter behandelt.

Vor allem, wenn Daten in einer Verkehrszentrale aggregiert und analysiert und dann die Ergebnisse Fahrzeughaltern als Dienst bereitgestellt werden, können Big Data und Cloud Computing ihre Stärken ausspielen (Burns 2013). Ziel ist die Optimierung des Verkehrsaufkommens. Dass autonome Fahrzeuge nicht nur auf gut ausgebauten Straßen Potenzial haben, zeigt eine von Lockheed Martin für Militärkonvois entwickelte Lösung (Edwards 2014). Im zivilen Bereich treibt neben den großen Konzernen wie Daimler und Audi vor allem Google die Forschung voran (Byczkowski 2013; n-tv 2013).

Mobilität ist auch im Kontext von Smart Cities ein zentrales Thema. Neben den im Kapitel V.2.1 besprochenen Fragen der Energieversorgung spielt in Smart Cities auch die Optimierung der Verkehrsführung eine wesentliche Rolle. Entsprechende Maßnahmen wirken sich nicht nur positiv auf den Energieverbrauch und die Umwelt aus, sondern erhöhen auch die Sicherheit der Verkehrsteilnehmer und den Komfort der Fahrer (z. B. Mearian 2014; Parent 2013). Beispielsweise lassen sich Staus in einer »intelligenten Stadt« vermeiden oder umfahren.



Im Kapitel V.2.1 wurde das Smart Grid als Beispiel für den Einsatz von Big Data genannt. Elektroautos spielen bei der flexiblen Speicherung von Energie eine zentrale Rolle (Proskawetz 2013). In Zukunft wäre durchaus denkbar, dass jederzeit und überall ein Auto gerufen werden kann und dieses einen dann an jeden gewünschten Ort bringt. Neben dem eigenen Auto wären auch Parkplätze zunehmend obsolet, und man könnte mit rund einem Fünftel der heutigen Autos ohne Weiteres auskommen. Ein erster Schritt in diese Richtung wurde im Bereich des Carsharings bereits gemacht, Auslastung und Koordinierung von Autos konnten optimiert werden (Proskawetz 2013).

MINISZENARIO: sim^{TD}

Zwischen Herbst 2008 und Frühjahr 2013 beschäftigten sich sechs Automobilhersteller, zwei Zulieferer, ein Kommunikationsunternehmen sowie mehrere öffentliche und wissenschaftliche Einrichtungen im Rahmen des sim^{TD}-Projekts mit der Erforschung, Erprobung und Anwendung von Car-to-X-Kommunikation. Das Projekt setzte sich vor allem die Erhöhung der Verkehrssicherheit und die Verbesserung der Verkehrseffizienz zum Ziel, aber auch die Integration von Mehrwertdiensten. Mehrwertdienste, die über die reine Routenführung hinausgehen und im Rahmen des Projekts betrachtet wurden, zielen beispielsweise auf die Lokalisierung von Tankstellen und Raststätten oder das Einholen von Information über Sehenswürdigkeiten und Freizeitangebote am Streckenrand ab. Im Rahmen des Projekts wurden die politischen, wirtschaftlichen und technischen Rahmenbedingungen für die erfolgreiche Einführung der Fahrzeug-zu-Fahrzeug- sowie der Fahrzeug-zu-Infrastruktur-Vernetzung mithilfe einer großflächigen Testfeldinfrastruktur untersucht. Das Projekt wurde von den Bundesministerien für Wirtschaft und Technologie, Bildung und Forschung sowie Verkehr, Bau und Stadtentwicklung gefördert.

Bei dem im Rahmen des Projekts durchgeführten Feldversuch kamen 120 Fahrzeuge sowie 127 Sensoren, die aus Lichtsignalen oder Baken heraus mit den Fahrzeugen kommunizierten, zum Einsatz. Zusammen legten die Fahrzeuge pro Woche 60.000 km zurück und generierten damit einen Datentransfer von rund 30 Terabyte (T-Systems 2012). Im Rahmen eines hybriden Systemansatzes kam sowohl WLAN-basierende Kommunikation als auch Kommunikation über Mobilfunk zum Einsatz (Stockburger 2013). Eine Rolle spielten beim sim^{TD}-Projekt nicht nur Informationen über die Verkehrslage und Informationen von Lichtsignalanlagen oder Wechselverkehrszeichen, sondern auch Umfeldinformationen, die sich beispielsweise auf das Wetter beziehen.

Die drei zentralen Charakteristika von Big-Data-Problemen sind gegeben: einerseits da die umfangreichen Datenmengen in der Verkehrszentrale nahezu in Echtzeit ausgewertet und an die Fahrzeuge sowie die Leitsysteme zurückgegeben werden müssen, um hilfreich zu sein. Andererseits stammen die Daten, die zur

Auswertung herangezogen werden, aus einer großen Anzahl von Quellen. Allerdings ist der Großteil der verarbeiteten Daten stark strukturiert. Da die Anbindung aller Fahrzeuge auf deutschen Straßen pro Tag Milliarden Terabyte an Datenträffic produzieren würde, ist an einem nationalen Rollout von sim^{TD} noch nicht zu denken (T-Systems 2012). Dennoch zeigte der Feldversuch deutlich, welches Anwendungspotenzial für die Verkehrssteuerung in der Durchführung von Big-Data-Analysen liegt. Eine zentrale Herausforderung stellen die enormen Datenmengen dar. Gleichzeitig wird Car-to-X-Kommunikation erst dann richtig interessant, wenn 30 % der Fahrzeuge entsprechend ausgestattet sind (Stockburger 2013). Laut Stockburger (2013) stellen auch die Kosten für den Aufbau der für Car-to-X-Kommunikation notwendigen Infrastruktur eine große Herausforderung dar, dies vor allem angesichts dessen, dass die vorhandenen Mittel für den Erhalt der bestehenden Straßen kaum ausreichen.

SICHERHEIT

2.4

Im Hinblick auf das Bedarfsfeld Sicherheit hat sich die Bundesregierung eine Reihe von Zielen gesetzt. Geschützt werden sollen die Gesellschaft und ihre Infrastrukturen vor Terrorismus, Sabotage, organisierter Kriminalität und den Folgen von Naturkatastrophen oder Unfällen. Da die Infrastrukturen hoch vernetzt und technisiert sind, spielt auch das Thema IT-Sicherheit in diesem Bedarfsfeld eine wichtige Rolle. Im Hinblick auf das Bedarfsfeld Sicherheit gibt es zahlreiche Anwendungspotenziale für Big Data. Vor allem in den Bereichen Kriminalitäts- und Terrorbekämpfung, Schutz vor Naturkatastrophen und IT-Sicherheit kommt Big Data bereits heute zum Einsatz.

Im Bereich der Kriminalitäts- und Terrorbekämpfung lassen sich durch die Analyse großer Datenmengen aus unterschiedlichen Quellen Straftaten unter bestimmten Voraussetzungen vorhersagen. In den USA hat beispielsweise das FBI mit Predictive Policing bereits erste Erfahrungen bei der Vorhersage von Straftaten gesammelt. Während die Software eine bessere Durchsetzung von Rechtsvorschriften verspricht, gibt es Datenschutzbedenken. Befürchtet wird ein Rechtssystem, das von Maschinen statt Menschen durchgesetzt wird. Wie im Kapitel I.1 bereits angedeutet gehen die Meinungen über den Einsatz von Big Data in bestimmten Bereichen auseinander. Während z. B. die Prävention von Verbrechen durch erhöhte Polizeipräsenz an bestimmten Punkten durchaus sinnvoll erscheint, wird die Berechnung der Rückfallwahrscheinlichkeit von einzelnen Straftätern und damit der Möglichkeit auf vorzeitige Entlassung aus dem Vollzug auf Grundlage von Daten über andere Straftäter eher kritisch gesehen. Der Einsatz von PredPol wird im Rahmen eines Miniszenarios detaillierter behandelt.

Auch im Bereich Betrugserkennung spielt Big Data eine wachsende Rolle (Gartner 2014). In der Vergangenheit setzten Unternehmen primär auf speziali-



sierte Einzellösungen, um Betrugsfälle aufzudecken. Entsprechende Lösungen sind beispielsweise auf Fälle im Zusammenhang mit Datenverlust, Finanzbetrug oder Missbrauch von Benutzerrechten ausgerichtet. Es wird angenommen, dass mithilfe von leistungsfähigen Analysemöglichkeiten und durch die Verwendung von Kontextdaten die Zahl von falsch positiven Befunden reduziert werden kann (Gartner 2014). Das ist vor allem deshalb wichtig, weil die Zahl der sicherheitsrelevanten Vorfälle von Jahr zu Jahr steigt. Besondere Aufmerksamkeit in der Öffentlichkeit bekam seit Sommer 2013 allerdings die umfassende Analyse von Kommunikationsdaten durch Geheimdienste wie die NSA. Gerechtfertigt wurden die Maßnahmen mit dem Argument der Terrorbekämpfung, die Erfolge in dieser Hinsicht waren allerdings bisher eher bescheiden (Bergen et al. 2014).

Im Hinblick auf den Katastrophenschutz gibt es eine Reihe von Big-Data-Anwendungen, die ähnlich den zahlreichen Ansätzen aus dem Bedarfsfeld Klima/Energie auf Simulationen beruhen. Gängig sind die Simulation von Erdbeben (z. B. Bergen et al. 2014) oder Unglücken im Zusammenhang mit Dämmen oder Brücken. Big Data kann aber auch bei der Vermeidung von Unglücken hilfreich sein. Wie im Kapitel II.1.4 erwähnt, können Polizeieinheiten bei Großereignissen unter Berücksichtigung von Daten aus dem Mobilfunk und aus sozialen Medien koordiniert werden.

Im Bereich IT-Sicherheit sind die in Big Data gesetzten Erwartungen groß. Arthur Coviello, der Vorstandsvorsitzende von RSA Security, der auf IT-Sicherheit spezialisierten Tochtergesellschaft von EMC, betonte im Rahmen der RSA Conference, wie wichtig es in Zukunft sein wird, große Datenbestände schnell zu analysieren, um angemessene Sicherheitsmaßnahmen treffen zu können. Auch sehr schwache Signale, die von hoch entwickelten Angriffen ausgehen, müssen frühzeitig erkannt werden. Coviello ist der Ansicht, dass der reaktive Ansatz im Bereich IT-Sicherheit zum Scheitern verurteilt ist (Hauck 2013). HP verfolgt mit der Entwicklung von IT-gestützten Sicherheitssystemen, die das Eindringen in IT-Systeme melden, einen entsprechenden Ansatz. Dabei wird ein Grundzustand definiert und entsprechende Abweichungen vom Systemverhalten können Alarme auslösen und zu Maßnahmen führen (HP 2012). In eine ähnliche Richtung gehen die Bestrebungen von IBM hin zum Einsatz von digitalen Wächtern für den Schutz und die Verwaltung von digitalen Identitäten. Dabei werden die einzelnen Identitäten, die z. B. in einer Cloud gespeichert sind, vor Missbrauch oder Fremdnutzung bewahrt, indem durch Big-Data-Analysen Abweichungen vom üblichen Verhalten erkannt werden.

MINISZENARIO: PREDPOL

Systeme zur polizeilichen Vorhersage sind in der Vergangenheit vor allem in den USA sowie im Vereinigten Königreich auf Interesse gestoßen, aber auch das deutsche Bundeskriminalamt beschäftigt sich mit entsprechenden Ansätzen



(Monroy 2012). Während IBM und Microsoft bereits seit einiger Zeit Lösungen für das Predictive Policing anbieten, sorgte die Software PredPol des gleichnamigen US-amerikanischen Start-ups zuletzt für besondere Aufmerksamkeit in der Öffentlichkeit (Bond-Graham/Winston 2013). Das Start-up wurde von Mathematikern und Sozialwissenschaftlern kalifornischer Universitäten gegründet. Die Wissenschaftler verglichen bei der Entwicklung der Modelle, auf denen die Software basiert, Straftaten mit Erdbeben. Sie gingen davon aus, dass so wie Erdbeben in der Nähe des Epizentrums Nachbeben verursachen, auf Straftaten wie Einbrüche meist weitere Einbrüche in räumlicher und zeitlicher Nähe folgen. Die entwickelten Modelle wurden an mehreren Tausend Einbrüchen getestet, die zuvor in einer bestimmten Region verübt wurden.

PredPol greift auf Statistiken früherer Ereignisse zurück. Relevant sind nicht nur verzeichnete Straftaten, sondern auch eingegangene Notrufe. Um aus einer Menge von erfassten Ereignissen potenzielle Hot Spots für zukünftige Ereignisse ableiten zu können, ist die Lokalisierung der erfassten Ereignisse essenziell. Die Wissenschaftler von PredPol gehen davon aus, dass die Wahrscheinlichkeit, dass Straftäter es in derselben Gegend, in der sie erfolgreich waren, noch einmal versuchen, viel höher ist als die Wahrscheinlichkeit, dass sie sich nach einem Erfolg ein neues Zielgebiet suchen.

Primäres Ziel des Einsatzes von Software wie PredPol ist in der Regel der effizientere Einsatz von beschränkten Personalressourcen. Den verfügbaren Polizisten werden Karten zur Verfügung gestellt, auf denen die prognostizierten Hot Spots für bestimmte Straftaten markiert sind. Die Polizisten sind angehalten, dort zu beobachten und Präsenz zu zeigen, wenn sie nicht durch andere Einsätze gebunden sind. Neue Straftaten werden unverzüglich erfasst und in die Berechnungen einbezogen. Zunächst kam PredPol in Santa Cruz zum Einsatz. Über einen Beobachtungszeitraum von sechs Monaten kam es zu einer Reduktion von Einbrüchen um 19 % (Friend 2013). Unklar blieb allerdings, inwiefern der Einsatz der Software dazu beigetragen hat. Aus diesem Grund wurde später in Los Angeles ein kontrolliertes Experiment durchgeführt, bei dem Predictive Policing in einem Stadtteil zur Anwendung kam, und die Entwicklung der vorhersagbaren Straftaten dann über die Stadtteile hinweg verglichen wurde. Während Eigentumsdelikte in Los Angeles im Beobachtungszeitraum insgesamt leicht zunahmen, gingen sie in dem Stadtteil, in dem PredPol eingesetzt wurde, um 12 % zurück. PredPol kommt mittlerweile nicht mehr nur in US-amerikanischen Großstädten wie Los Angeles, Seattle und Atlanta zum Einsatz, sondern z. B. auch in der Grafschaft Kent im Vereinigten Königreich (The Economist 2013a).

Software wie PredPol erfordert die rasche Verarbeitung von polystrukturierten Daten aus zahlreichen Quellen. Damit sind zwei der drei zentralen Charakteristika von Big-Data-Problemen gegeben. Die Datenmenge spielt in diesem Beispiel verglichen mit anderen Big-Data-Anwendungen eine eher kleine Rolle.



Eine der zentralen Herausforderungen ist die Akzeptanz des Systems. Ähnlich wie bei IBMs Watson muss in diesem Fall PredPol potenzielle Nutzer überzeugen, dass die verwendete Datenbasis sowie die generierten Antworten vertrauenswürdig sind. Andererseits stellt auch hier die Integration eines neuen Systems in etablierte Abläufe eine Herausforderung dar.

KOMMUNIKATION

2.5

Das Bedarfswelt Kommunikation zielt auf Information und Kommunikation als Grundlage effizienter Wertschöpfungsprozesse ab. Die Ziele dieses Bedarfswelts wurden von der Bundesregierung sehr weit gefasst und überschneiden sich teils mit Zielen, die auch im Rahmen anderer Bedarfswelder verfolgt werden. Im Kern geht es um die Intensivierung und Ausweitung des Einsatzes von IKT, damit Deutschland seine Position als Hightech-Standort festigen und weiter ausbauen kann. Das Bedarfswelt Kommunikation bietet damit unzählige Anwendungspotenziale für Big Data. Besondere Aufmerksamkeit wurde zuletzt Big-Data-Anwendungen mit Fokus auf den Bereich Kundenorientierung sowie auf die intelligente Fabrik (Smart Factory) und Hochfrequenzhandel zuteil. Aber auch bei der Optimierung von Wahlkampagnen konnte Big Data vor allem in den USA seine Stärken bereits unter Beweis stellen.

Im Handel hat Big Data großes Potenzial, die Kundenorientierung nachhaltig zu verbessern. Die Vorhersage des Verhaltens von Kunden kann in den Bereichen Verkauf und Logistik zu Vorteilen führen (Schäfer et al. 2012, S. 58 ff.).

Auch in der Industrie kommt Big Data zunehmend zum Einsatz. Durch die intelligente Fabrik (Smart Factory), in der die reale und die virtuelle Welt zusammenwachsen, entstehen neue Formen von Wertschöpfung und Beschäftigung. Im Rahmen der Hightech-Strategie der Bundesregierung ist im Hinblick auf die Informatisierung der klassischen Industrien auch von »Industrie 4.0« die Rede. Die durch intelligente Geräte erfassten umfangreichen Daten können als Grundlage für neue Dienstleistungen dienen. Hilfreich ist Big Data z. B. bei der Vorhersage von Ausfällen von bestimmten Maschinenteilen (Schürmann 2013). Predictive Maintenance ist ein relativ neues Feld. Der Verschleiß von Maschinen wird mithilfe von Sensoren festgestellt. Hilfreich ist die Kenntnis von Ausfallwahrscheinlichkeiten vor allem für die Planung von Reparaturen und die Vermeidung von schwerwiegenden Schäden. Siemens setzt Predictive Maintenance beispielsweise bei Turbinen und Tomografen ein. Ein allgemeines wissenschaftliches Modell wurde von Liao et al. (2012) entwickelt. Der Einsatz von Big-Data-Analysen im Zusammenhang mit »Industrie 4.0« wird im Rahmen eines Miniszenarios detaillierter behandelt.

Auch im Finanzbereich gibt es Ansätze auf Grundlage von Big Data, die Vorteile bringen sollen (Schäfer et al. 2012, S. 54 f.). Mit Data Mining wird im Hinblick auf den Hochfrequenzhandel ein konkretes Tool genannt, das anomale Ereignisse auf Twitter in Echtzeit erkennt und angereichert um Daten aus einer eigenen Unternehmens- und Branchendatenbank in Form von Eilmeldungen an Kunden weiterleitet (Schäfer et al. 2012, S. 31). Relevante Ansätze im Finanzbereich zur Betrugserkennung wurden im Kapitel V.2.4 kurz angesprochen.

Das Anwendungspotenzial von Big Data bei der Optimierung von Wahlkampagnen wurde 2012 bei der Präsidentschaftswahl in den USA ersichtlich. Das Wahlkampfteam um Barack Obama wertete über mehr als ein Jahr hinweg große Mengen an unterschiedlichen Daten aus. Der Fokus konnte auf jene Staaten und Personen konzentriert werden, die sich am ehesten mit vertretbarem Aufwand gewinnen ließen (Issenberg 2012; Quack 2013).

In der öffentlichen Verwaltung spielt Big Data eine nicht zu unterschätzende Rolle. Wie im Kapitel IV.1.2.1 erläutert kommt Big Data z.B. in Indien im Zusammenhang mit der Identifizierung von Personen im Rahmen von öffentlichen und kommerziellen Geschäftsprozessen zum Einsatz. Auch weitere Anwendungspotenziale in Bereichen wie der Verkehrsplanung, der öffentlichen Sicherheit oder der Energieversorgung wurden bereits angesprochen (Kommune21 2014; Viola 2013). Spätestens wenn man die Geheimdienstaktivitäten einbezieht, wird klar, dass mit dem Einsatz von Big Data in der öffentlichen Verwaltung neben Potenzialen auch erhebliche Gefahren für demokratische Gesellschaften verbunden sind (Kommune21 2014).

Im Kontext des sogenannten »Connected Home« oder »Smart Home« spielen sowohl Big Data als auch Cloud Computing eine wichtige Rolle. In Haushalten über verschiedenste Sensoren gesammelte Daten im Hinblick auf Themen wie Sicherheit, Energie, Gesundheit oder Unterhaltung können über Cloudlösungen einerseits den Bewohnern komfortabel zugänglich gemacht werden und andererseits in aggregierter Form ausgewertet werden. Auswertungen sind insbesondere im Hinblick auf Verbesserungen in Smart Cities relevant. Privatpersonen können dadurch profitieren, dass sich verschiedene Geräte im Haushalt abstimmen und dass eine Überwachung und Steuerung auch von unterwegs möglich wird.

Big-Data-Analysen kommen aber auch zur Vorhersage von sozioökonomischen Entwicklungen zum Einsatz. Konkrete Anwendungsfälle lassen sich bei der Vorhersage von Armutskrisen in Entwicklungsländern beobachten, wo durch die Analyse von Mobilfunkdaten die wirtschaftliche Entwicklung eines Landes eingeschätzt werden kann und Vorhersagen abgeleitet werden können (Global Pulse 2012). Auch im Rahmen des FRP 7-Projekts »FuturICT«⁷ wird in diesem Bereich geforscht. Verschiedene Zukunftsszenarien sollen getestet und wahr-

7 www.futurict.eu



scheinliche Entwicklungen identifiziert werden können. Teils ist in diesem Zusammenhang von Predictive Modelling sowie von Reality Mining die Rede (Peil 2013). Als zentral wird die Fähigkeit angesehen, schwache Signale, die auf Veränderungen hindeuten, zu erkennen und richtig zu interpretieren.

MINISZENARIO: »INDUSTRIE 4.0«

»Industrie 4.0« ist ein Zukunftsprojekt in der Hightech-Strategie der deutschen Bundesregierung mit dem Ziel, die Verwendung von Informationstechnik in der klassischen Industrie voranzutreiben. Angestrebt wird das Idealbild der intelligenten Fabrik, die sich durch Wandlungsfähigkeit, Ressourceneffizienz und Ergonomie sowie die Integration von Kunden und Geschäftspartnern in Geschäfts- und Wertschöpfungsprozesse auszeichnet (Böhler 2012). Technologische Grundlage sind cyberphysische Systeme und das Internet der Dinge.

In der Praxis sind entsprechende Ansätze noch nicht weit verbreitet. Als Vorreiter gilt beispielsweise der taiwanische Elektronikauftragshersteller Foxconn, der für Apple große Mengen an iPhones herstellt. Da Apple die elektronischen Komponenten bei iPhones häufig ändert, muss Foxconns Fertigung entsprechend anpassungsfähig sein. Um eine hohe Wandlungsfähigkeit bei der Fertigung gewährleisten zu können, setzt Foxconn auf einen weitgehend computergesteuerten Fertigungsprozess (Cassell 2013). Insbesondere Hersteller von hochgradig individualisierten Produkten profitieren von Maßnahmen im Bereich der Automation.

Obwohl in den nächsten Jahren aufgrund des steigenden Bedarfes an kundenindividueller Massenfertigung mit einer zunehmenden Dynamik im Bereich der Industrieautomation zu rechnen ist, sind große Sprünge eher nicht zu erwarten. Hersteller werden nach und nach ihre Fertigungsanlagen weiter vernetzen und mehr und mehr Daten sammeln und in die Steuerung des Fertigungsprozesses einfließen lassen. Wie industrielle Informationstechnik technischen Systemen zu mehr Intelligenz verhelfen kann, ist bereits heute in der Lemgoer Modellfabrik zu sehen. Die vom Institut für industrielle Informationstechnik der Hochschule Ostwestfalen-Lippe zu Forschungszwecken betriebene intelligente Fabrik adressiert die Handlungsfelder Wandlungsfähigkeit, Ressourceneffizienz und Benutzerfreundlichkeit⁸.

Eine zentrale Herausforderung im Kontext von »Industrie 4.0« ist die Analyse der Flut von Daten, die in intelligenten Fabriken erzeugt wird (Cassell 2013). Um Effizienzsteigerungen und einen höheren Grad an Wandlungsfähigkeit in der Fertigung erreichen zu können, ist die Steuerung des Fertigungsprozesses auf Grundlage der vorhandenen Daten unumgänglich. Das Potenzial von computergesteuerten Fertigungsprozessen ist enorm. Laut Cassell (2013) wird im Bereich

8 www.smartfactory-owl.de



der Industrieautomation mit einem Umsatz von 183 Mrd. US-Dollar im Jahr 2013 und 209 Mrd. US-Dollar im Jahr 2016 gerechnet.

Die drei zentralen Charakteristika von Big-Data-Problemen sind gegeben: einerseits da die umfangreichen Datenmengen in intelligenten Fabriken nahezu in Echtzeit ausgewertet und an die Fertigungsanlagen zurückgegeben werden müssen. Andererseits stammen die Daten, die zur Auswertung herangezogen werden, aus einer großen Anzahl von Quellen. Allerdings ist der Großteil der verarbeiteten Daten stark strukturiert.

FAZIT

3.

Im Hinblick auf Unternehmen sowie die öffentliche Verwaltung konnten zahlreiche sozioökonomische Potenziale der Durchführung von Big-Data-Analysen und der Nutzung von Clouddiensten identifiziert werden. Beiden wird Potenzial zur Steigerung des Umsatzes sowie zur Senkung der Kosten zugesprochen. Während im Zusammenhang mit Big Data Umsatzsteigerungen eine größere Bedeutung beigemessen wird, überwiegt im Zusammenhang mit Cloud Computing das Potenzial im Hinblick auf Kostensenkungen. Für die Steigerung der Kundenorientierung, ein zentrales Thema in Wirtschaft und öffentlicher Verwaltung, wird in der Durchführung von Big-Data-Analysen großes Potenzial gesehen. Sowohl Big-Data-Analysen als auch die Nutzung von Cloud Computing können Studien zufolge zur Steigerung der Produktivität beitragen. Während im Zusammenhang mit Big Data vor allem das Potenzial zur Herbeiführung besserer Entscheidungen hervorgehoben wird, steht im Zusammenhang mit Cloud Computing die Steigerung der organisatorischen Flexibilität im Mittelpunkt. Zur Verbesserung der Innovationsfähigkeit wird sowohl in der Durchführung von Big-Data-Analysen als auch in der Nutzung von Clouddiensten Potenzial gesehen. Darüber hinaus fördert Big Data Studien zufolge den Aufbau strategischer Wettbewerbsvorteile, Cloud Computing hingegen kann zur Senkung der Notwendigkeit, eigene IT-Fachkräfte beschäftigen zu müssen, beitragen.

Im Hinblick auf die Wissenschaft wird besonderes Potenzial nicht nur im einfacheren Zugriff auf vorhandene Forschungsdaten gesehen, sondern auch in der Möglichkeit die primär theorie- und modellgetriebene Forschung vermehrt durch datengetriebene Ansätze unterstützen zu können. Die Nutzung von Clouddiensten eröffnet Privatpersonen komfortable Möglichkeiten beim Zugriff auf sowie beim Teilen von Daten. Zunehmend werden Lösungen zur Einbindung von Privatpersonen in die Entwicklung von neuen Produkten oder Dienstleistungen auf Basis von Cloud Computing bereitgestellt, wovon sowohl die jeweiligen Organisationen als auch Privatpersonen profitieren. Ähnlich stellt sich die Situation im Bereich Big Data dar, wo Privatpersonen zur Sammlung großer Datenmengen



beitragen, selbst aber in der Regel nicht in deren Analyse eingebunden sind. Von den Ergebnissen der Datenauswertung profitieren wiederum beide, Organisationen und Privatpersonen.

Grundsätzlich zeigt die Analyse der gesamtgesellschaftlichen Potenziale, dass eine Reihe wesentlicher Potenziale mit Cloud Computing und Big Data verbunden ist. Insbesondere die ökonomischen Potenziale beider Technologien sind dazu geeignet, sich gegenseitig zu verstärken, da Cloud Computing beispielsweise auch kleineren Unternehmen den Zugang zu den Möglichkeiten von Big-Data-Analysen ermöglicht. Nur im Fall der Nachhaltigkeit (Energie- und Ressourcenverbrauch) ist offen, ob Big Data hier nicht eher den Trend zu mehr Verbrauch befördert. Doch aufgrund vieler noch ungeklärter Fragen lässt sich dies nicht abschließend bewerten. Dies weist auch auf ein weiteres Ergebnis der Analyse. Die bisher realisierten Potenziale sind im Vergleich zu den zukünftig erwarteten Potenzialen noch gering. Zudem zeigt sich, dass das Ausmaß der Potenziale stark von verschiedenen Aspekten, wie etwa den zugrundeliegenden Annahmen und Berechnungsverfahren, abhängt. So ist etwa die teilweise kritische Würdigung der verschiedenen Studienergebnisse nicht darauf ausgerichtet, die mit Big Data und Cloud Computing assoziierten Potenziale als Ganzes infrage zu stellen. Vielmehr geht es darum, relevante Problembereiche anzusprechen. So beruhen bis auf wenige Ausnahmen die meisten Studien auf Annahmen und weniger auf tatsächlichen empirischen Werten. Entweder handelt es sich dabei um Daten zu allgemeinen Entwicklungen wie dem erwarteten Wachstum der Weltwirtschaft, das von Institutionen wie dem IWF bezogen wurde, oder um Daten zu Entwicklungen um die speziellen Technologien, denen oftmals verschiedene Arten von Umfragen oder Experteneinschätzungen zugrunde liegen. Gerade Letztere geben vor allem erwartete Annahmen über zukünftige Potenziale wieder. Forschungen haben gezeigt, dass die meisten neuen Technologien einem zweiphasigen Zyklus folgen, dem »Double Boom« (Schmoch 2007), wobei vor allem die erste Phase stark von optimistischen Erwartungen geprägt ist und dementsprechend mögliche Einschätzungen beeinflusst. Grundsätzlich bedeutet dies aber nicht, dass die Potenziale nicht vorhanden sind, sondern nur, dass es in der Regel wesentlich länger dauert, bis sie realisiert werden können. Die Berücksichtigung dieser Aspekte könnte einen Ansatzpunkt für verbesserte Prognosen bieten.

Dass die Realisierung von Potenzialen zumeist eine geraume Zeit benötigt, weist auf einen zweiten Punkt hin, nämlich dass die Voraussetzungen zur Realisierung meistens komplexer sind als ursprünglich erwartet. Dazu zählen unter anderem eine breite Akzeptanz, eine breite Diffusion, die Überwindung von Hemmnissen sowie das Zusammenspiel mit weiteren technologischen Entwicklungen. Anekdotisch wird diese Problematik am sogenannten Solow-Paradoxon deutlich. So löste Robert Solow, Nobelpreisträger für Wirtschaftswissenschaften, mit der im Jahr 1987 getätigten eher beiläufigen Anmerkung »You can see the computer age everywhere these days, except in the productivity statistics« eine Reihe von



Forschungsarbeiten aus, die zwar letzten Endes diesen Einfluss zeigen konnten, zugleich zeigten sie aber auch, dass die erwarteten Wirkungsmechanismen und Zeiträume völlig andere waren, als ursprünglich erwartet (Brynjolfsson 1993; Solow 1987; Wilson 1993).

Ein weiterer Aspekt betrifft die implizit angenommene kausale Beziehung zwischen Produktivitätssteigerungen und der Verbesserung der gesamtgesellschaftlichen Situation. Diese ergibt sich daraus, dass es einen erkennbaren Zusammenhang zwischen der Zunahme der Produktivität und der Zunahme des BIP pro Kopf gibt. Doch wird in den USA in den letzten Jahren verstärkt die Frage diskutiert, warum der Median des realen durchschnittlichen Haushaltseinkommens deutlich langsamer gewachsen ist als die Produktivität und das reale BIP pro Kopf. Dies deutet auf eine zunehmende Ungleichheit in der gesellschaftlichen Entwicklung bzw. an der Wohlfahrtsteilhabe hin. Dieses Phänomen – das auch in etwas schwächerer Ausprägung in anderen Industrienationen auftritt – wird in der wissenschaftlichen Diskussion als »Stagnation« bezeichnet (Cowen 2011; Phelps/Tilman 2010) und könnte letztlich dazu führen, dass gesellschaftliche Inklusion und Kohäsion kontinuierlich schwieriger wird. Während unterschiedliche Ursachen wie zu geringe Produktivitätssteigerungen oder Globalisierungseffekte diskutiert wurden und werden, haben Ökonomen am MIT, Brynjolfsson und McAfee, auf die mögliche Rolle von Technologien, insbesondere der Digitalisierung durch IT, hingewiesen. Ihr Hauptargument zielt, wie schon in der Diskussion um die Beschäftigungswirkungen angedeutet, darauf, dass die Digitalisierung zwar neue Beschäftigung schafft, aber dass die Qualität dieser Beschäftigung sehr unterschiedlich ist. Die immer weiter voranschreitende Digitalisierung und Automatisierung wird ihrer Ansicht nach immer mehr Arbeitsplätze betreffen, an denen Maschinen Menschen ersetzen können. Jedoch lautet ihre Schlussfolgerung nicht, diese Entwicklung aufhalten zu müssen, sondern dass die Gesellschaft möglichst viele der positiven Potenziale realisieren sollte. So weisen sie sehr konkret auf die dafür notwendigen Maßnahmen sowohl im Infrastruktur- und Bildungsbereich als auch zur Verbesserung weiterer Rahmenbedingungen für Wettbewerbsfähigkeit hin (Brynjolfsson/McAfee 2012, 2014). Abschließend soll nochmals unterstrichen werden, dass die Realisierung von Potenzialen kein Selbstläufer ist, sondern im Gegenteil auch die Möglichkeit besteht, dass eine Gesellschaft zu den Verlierern dieser Entwicklungen gehören kann, wenn sie nicht entsprechend reagiert. Dementsprechend gilt es, die zentralen Herausforderungen so zu adressieren.

Big Data bietet zahlreiche Anwendungspotenziale. Dies lässt sich unter Bezugnahme auf die fünf Bedarfsfelder der Hightech-Strategie der Bundesregierung veranschaulichen. Während in einigen Bereichen eine Kombination mit Cloud Computing naheliegend ist, ist das in anderen Bereichen nicht der Fall. Im Hinblick auf das Bedarfsfeld Klima/Energie spielt Big Data vor allem in den Bereichen Klimaforschung, Smart Cities, Smarte Energienetze und Umweltschutz be-



reits heute eine wichtige Rolle. Die Erforschung von Krankheitsursachen und Behandlungsmethoden, die Epidemiologie sowie die individualisierte Medizin sind mögliche Anwendungsbereiche im Zusammenhang mit dem Bedarfsfeld Gesundheit/Ernährung. Auch im Hinblick auf das Bedarfsfeld Mobilität gibt es Anwendungspotenziale für Big Data, vor allem in den Bereichen autonome Fahrzeuge und intelligente Transportsysteme. Darüber hinaus kommt Big Data bereits heute in den Bereichen Kriminalitäts- und Terrorbekämpfung, Schutz vor Naturkatastrophen und IT-Sicherheit zum Einsatz. Diese Bereiche werden dem Bedarfsfeld Sicherheit zugeordnet. Im Hinblick auf das Bedarfsfeld Kommunikation spielt Big Data im Bereich der Kundenorientierung, der intelligenten Fabrik und des Hochfrequenzhandels bereits heute eine große Rolle.

Vor allem anhand der Miniszenerien lässt sich gut erkennen, dass Art und Weise, wie Big-Data-Analysen in den unterschiedlichen Bereichen zur Erreichung von Zielen beitragen, sehr vielfältig sind. Die Rolle von Cloud Computing hängt von den jeweiligen Anforderungen und Rahmenbedingungen ab. Die Beschäftigung mit den Anwendungspotenzialen von Big Data zeigt, dass Cloud Computing vor allem dann nützlich sein kann, wenn sich die auszuwertenden Daten ebenfalls in der Cloud befinden, wenn Ergebnisse von Analysen komfortabel zugänglich gemacht werden sollen und wenn auf Daten aus verschiedenen Quellen zugegriffen werden soll. Gleichzeitig lassen sich allerdings auch zentrale Herausforderungen erkennen, die bewältigt werden müssen, um von den Möglichkeiten, die sich durch Big Data ergeben, profitieren zu können. Es geht dabei nicht nur um technische und betriebliche, sondern auch um rechtliche, gesellschaftliche und strukturelle Herausforderungen. Die Größe der Herausforderungen hängt unter anderem auch davon ab, ob Cloud Computing im Rahmen der Big-Data-Analyse zum Einsatz kommt.

HERAUSFORDERUNGEN

VI.

TECHNISCHE HERAUSFORDERUNGEN

1.

Auf Grundlage aktueller Studien können zahlreiche technische Herausforderungen von Big Data und Cloud Computing identifiziert werden. Die Teilnehmer relevanter Studien wurden in der Regel explizit nach Herausforderungen oder nach Problemen oder Barrieren im Kontext der Durchführung von Big-Data-Analysen bzw. der Nutzung von Clouddiensten gefragt. Im Hinblick auf Cloud Computing beziehen sich die besprochenen Herausforderungen primär auf die Nutzung von Diensten aus Public Clouds.

Neben fehlendem Fachwissen, fehlenden überzeugenden Einsatzszenarien und Kostenaspekten spielen vor allem technische Probleme im Kontext von Big Data eine zentrale Rolle. Im Rahmen der BARC-Studie gab etwa ein Drittel der Befragten an, dass sie mit technischen Problemen zu kämpfen haben (Bange et al. 2013, S.26). Der Anteil derjenigen, die technische Probleme bei der Durchführung von Big-Data-Analysen sehen, ist in den Fachbereichen etwas größer als im Bereich der IT.

ANALYSEARCHITEKTUR

1.1

Technisch anspruchsvoll werden Big-Data-Analysen genau durch jene Charakteristika, über die sich Big Data vorrangig definiert. TCS (2013, S.33) setzte auf Grundlage einer umfassenden Studie den Umgang mit umfangreichen, heterogenen Datenmengen, die rasch verarbeitet werden müssen, auf Rang 2 der Herausforderungen im Big-Data-Umfeld. Auf Rang 1 liegt mit der eingeschränkten Bereitschaft von Fachbereichen, »ihre« Daten zu teilen, eine betriebliche Herausforderung. Vor allem die Verarbeitung von Daten in Echtzeit stellt sehr hohe Anforderungen an die verwendete Analysearchitektur. Da das Standard-Ethernet nicht den Anforderungen der industriellen Echtzeit genügt, die isochrone Zykluszeiten von weniger als eine Millisekunde verlangen, erlauben cloudbasierte Ansätze ohne Anwendung spezieller Verfahren bestenfalls eine Verarbeitung in Quasiechtzeit. Aber auch der Umfang der Datenmengen stellt ein Problem dar. Im Rahmen der Interxion-Studie wurde z. B. festgestellt, dass knapp ein Drittel der befragten Unternehmen keine ausreichenden Speicherkapazitäten hat (Interxion 2013, S.8). Einer IDC-Studie zufolge übersteigt der Umfang der erzeugten Daten den vorhandenen Speicherplatz deutlich (The Economist 2010c).

In der Praxis gibt es vor allem Probleme beim Aufbau einer geeigneten Analysearchitektur, bei der Realisierung des Zugriffs von verschiedenen Endgerätetypen



sowie bei der Erstellung einer für den Endnutzer angemessenen Benutzerschnittstelle. Während im Rahmen der TDWI-Studie ein Drittel der Befragten bestätigte, dass Schwierigkeiten beim Aufbau einer Analysearchitektur zu den zentralen Barrieren im Hinblick auf die Durchführung von Big-Data-Analysen gehören, und 23 % angaben, dass es Probleme vor allem im Hinblick auf die Skalierbarkeit der Architekturen gibt (Russom 2011, S. 12), landete die Auswahl für eine Analysearchitektur geeigneter Big-Data-Technologien bei der TCS-Studie auf Rang 13 der wichtigsten Herausforderungen (TCS 2013, S. 33). Im Rahmen der Interxion-Studie gaben 29 % der Befragten an, dass die Realisierung des Zugriffs von verschiedenen Endgerätetypen eine wichtige Herausforderung darstellt (Interxion 2013, S. 8). TCS (2013, S. 33) setzte die Erstellung einer für den Endnutzer angemessenen Benutzerschnittstelle auf Rang 7 der Herausforderungen im Big-Data-Umfeld. Hervorgehoben wird vor allem die Notwendigkeit einer visuellen Aufbereitung der Ergebnisse in einer Form, die geeignet ist, als Grundlage für Entscheidungen zu dienen. Russom (2011, S. 12) fand heraus, dass 22 % der von ihm befragten Unternehmen Probleme dabei haben, Big Data für die Endnutzer verwendbar zu machen. Im Rahmen der BARC-Studie gaben 15 % der Teilnehmer an, dass Big Data für potenzielle Endnutzer aus den Fachbereichen nicht verwendbar ist (Bange et al. 2013, S. 26).

Auch im Rahmen der Experteninterviews mit Vertretern von Unternehmen, wissenschaftlichen Einrichtungen und Verbänden wurden Herausforderungen im Zusammenhang mit den zentralen Big-Data-Charakteristika und dem Aufbau geeigneter Analysearchitekturen angesprochen. Beispielsweise wurde betont, dass es im Big-Data-Bereich keine Lösungen gibt, die über alle Charakteristika hinweg gut abschneiden. In der Regel ist es daher notwendig, eine für die jeweiligen Anforderungen geeignete Analysearchitektur aufzubauen. In diesem Zusammenhang wurde darauf hingewiesen, dass man im Hinblick auf die Voraussetzungen zum Aufbau der Architektur deutliche Unterschiede zwischen den USA und Europa erkennen kann. Während in den USA bereits viele potenzielle Anwender erste Erfahrungen mit Big-Data-Technologien wie Hadoop gesammelt haben, ist das in Europa nicht der Fall. Einer der befragten Experten gab an, dass aus seiner Sicht die USA im Vergleich zu Europa etwa zwei Jahre voraus sind. Die Experten stellten fest, dass der Aufbau von Analysearchitekturen vor allem deshalb so herausfordernd ist, weil sich die Anwendungsfälle und damit auch die Anforderungen von Anwendern im Laufe der Zeit weiterentwickeln, sich die hoch spezialisierten Big-Data-Lösungen aber nur sehr beschränkt an neue Anforderungen anpassen lassen. Dies kann dazu führen, dass in relativ kurzer Zeit hoch diversifizierte und komplexe Architekturen entstehen. Einer der Experten wies darauf hin, dass es derzeit im Big-Data-Bereich fünf oder sechs verschiedene Architekturkonzepte gibt, die von den Anwendungsgebieten unterschiedlich ausgerichtet sind.



Mögliche Ursachen für die beschriebenen Schwierigkeiten sind neben fehlendem Fachwissen vermutlich einerseits in der Verwendung veralteter Technologien und andererseits in der mangelnden Reife neuer Big-Data-Technologien zu finden. 31 % der im Rahmen der Interxion-Studie Befragten gaben an, dass veraltete Technologie im Kontext von Big Data eine Herausforderung darstellt. Russom (2011, S.12) betrachtete diese Problematik etwas genauer. Im Rahmen der TDWI-Studie wurde festgestellt, dass vorhandene Lösungen einerseits in ihren Analysemöglichkeiten eingeschränkt und andererseits zu langsam sind. Im Hinblick auf fehlende Analysemöglichkeiten gaben Studienteilnehmer beispielsweise an, dass vorhandene Data-Warehouse-Lösungen in der Regel nur für Berichte und OLAP geeignet sind. Außerdem wurde auf Probleme durch langsame Abfrageverarbeitungen und langsames Laden von Daten hingewiesen. Die mangelnde Reife von Big-Data-Technologien wurde vor allem von Schäfer et al. (2012, S.48) thematisiert. 18 % der im Rahmen der IDC-Studie Befragten gaben an, dass langsame Internetverbindungen als Herausforderung betrachtet werden (Bradshaw et al. 2012, S. 34).

Im Hinblick auf die mangelnde Reife neuer Big-Data-Technologien wurde im Rahmen der Interviews vor allem kritisiert, dass weiterführende, fortschrittliche Analysen nur sehr eingeschränkt möglich sind. Die vorhandenen Lösungen sind den befragten Experten zufolge nicht dafür geeignet, komplexes Data Mining durchzuführen, wie es beispielsweise für Cluster- oder Netzwerkanalysen erforderlich ist. Das Kernproblem ist, dass das Konzept der Iteration nicht unterstützt wird. Einer der Experten wies darauf hin, dass Statistiklösungen wie MatLab oder R iterative Algorithmen im Prinzip sehr effizient ausführen können, aber nicht skalieren. Systeme wie Hadoop haben das Problem, dass sie Iterationen nicht nativ unterstützen, es muss also immer ein spezielles Treiberprogramm verwendet werden, das nicht Teil von Hadoop ist. Dafür wird dem Experten zufolge auf Sprachen wie Java oder Scala zurückgegriffen. Besonders hingewiesen wurde von den Experten darauf, dass die Analyse möglichst immer dort durchgeführt werden soll, wo die Daten entstehen, da der Transport von großen Datenmengen ein massives Problem darstellt.

DATENMATERIAL

1.2

Nicht unwesentliche Herausforderungen im Zusammenhang mit Datensicherheit, Datenschutz und auch Datenqualität haben zwar eine technische Komponente, sind aber nicht als ausschließlich technische Herausforderungen zu verstehen. Während Schäfer et al. (2012, S.48) Datensicherheit und Datenschutz als Kernherausforderung im Umfeld von Big Data sehen, landet der Schutz von Daten vor unberechtigtem externen oder internen Zugriff nur auf den Rängen 14 bzw. 16 der TCS-Herausforderungsranliste (TCS 2013, S.33). Immerhin ein Viertel



der im Rahmen der BARC-Studie Befragten bestätigt die Relevanz der Datenschutzproblematik im Kontext von Big Data (Bange et al. 2013, S.26). Bange et al. (2013) führen den relativ niedrigen Wert darauf zurück, dass sich Organisationen der Problematik bewusst sind und entsprechend damit umgehen und gegebenenfalls auf die Verarbeitung von personenbezogenen Daten verzichten. Von Schäfer et al. (2012, S.48) wurden Datensicherheit und Datenschutz nicht für sich alleine, sondern kombiniert betrachtet. Dies könnte ein Grund für die sehr breite Zustimmung sein. Während Schroeck et al. (2012, S.15) allgemein auf das Problem der mangelnden Datenqualität hinweisen, stellen Schäfer et al. (2012, S.48) klar, dass es darüber hinaus bei der Gewinnung, Extraktion und Zusammenführung von Daten zu Problemen kommen kann.

Im Bereich Datensicherheit und Datenschutz sehen Bradshaw et al. (2012, S.34) eine zentrale Herausforderung für die Nutzung von Clouddiensten. Rund 30 % der Teilnehmer der IDC-Studie halten Datensicherheit und Datenschutz für eine massive Barriere im Hinblick auf die Nutzung von Clouddiensten. Wie im Kontext von Big Data, haben Datensicherheit und Datenschutz auch im Zusammenhang mit Cloud Computing zwar eine technische Komponente, sind aber nicht als ausschließlich technische Herausforderungen zu verstehen.

Auch im Rahmen der Interviews mit Experten spielten Herausforderungen für Datensicherheit, Datenschutz und Datenqualität eine große Rolle. Datensicherheit wurde vor allem bei der Nutzung von Big Data in der Cloud als Herausforderung gesehen, wenn auf Public-Cloud-Angebote zurückgegriffen werden sollte. Im Hinblick auf strategisch relevante oder personenbezogene Daten wurde die Nutzung von entsprechenden Angeboten als kritisch angesehen. Im Hinblick auf den Datenschutz wurde beispielsweise im Zusammenhang mit Public-Cloud-Angeboten darauf hingewiesen, dass es schwierig, wenn nicht unmöglich zu steuern ist, welche Daten wo genau landen, wenn sie außer Haus gegeben werden. Als problematisch wird auch die Heterogenität der Datenschutzrichtlinien in Europa angesehen, die verglichen mit den USA, wo es einen sehr großen Binnenmarkt gibt, einen großen Nachteil darstellt. Rechtliche Herausforderungen im Zusammenhang mit dem Thema Datenschutz werden im Kapitel VI.3.1 detaillierter behandelt. Von einem Experten wurde die Datenaufbereitung als besondere Herausforderung hervorgehoben. Berücksichtigt man die Aufbereitung der benötigten Daten, kann sich ein an sich interessanter Anwendungsfall schnell als wirtschaftlich nicht sinnvoll entpuppen. Nicht selten wird für Big-Data-Analysen beispielsweise auf Archive zurückgegriffen, die ursprünglich für eine Auswertung nicht vorgesehen waren – entsprechend niedrig war der Anspruch und ist zugleich das Vertrauen in die Datenqualität. Vor allem dann, wenn der Link zwischen Daten und Ergebnis nicht mehr einfach ersichtlich ist, wird die Verlässlichkeit und die Qualität der Daten zu einem großen Thema.

INTEROPERABILITÄT**1.3**

Im Hinblick auf die Nutzung von Cloud Computing stellt neben Problemen in den Bereichen Datensicherheit und Datenschutz mangelnde Interoperabilität eine zentrale technische Herausforderung dar. Interoperabilität ist nicht nur im Hinblick auf die Integration von Cloudlösungen mit Inhouselösungen ein Thema, sondern auch im Hinblick auf die Integration von Cloudlösungen untereinander. Bei der von KPMG/BITKOM (2013, S.23 f.) durchgeführten Studie »Cloud-Monitor 2013« stimmten 47 % der Vertreter von Unternehmen, die bereits Public-Cloud-Dienste nutzen, der Aussage, dass die Integration von Cloudlösungen mit Inhouselösungen schwierig ist, und 41 %, dass die Interoperabilität verschiedener Cloudlösungen unzureichend ist, voll und ganz oder zumindest eher zu. Zu ähnlichen Ergebnissen kam die Befragung von Vertretern von Unternehmen, bei denen die Nutzung von Clouddiensten zum Zeitpunkt der Befragung kein Thema war. Im Rahmen einer IDC-Studie gab rund ein Viertel der Befragten an, dass Einschränkungen bei der Portabilität von Daten eine massive Barriere im Hinblick auf die Nutzung von Clouddiensten darstellen (Bradshaw et al. 2012, S. 34). Sowohl die wachsende Zahl an Anbietern als auch der Mangel an Standards im Umfeld von Cloud Computing führen zu Problemen bei der Interoperabilität und Portabilität von Diensten. Der Wechsel des Cloudanbieters gestaltet sich in der Regel schwierig und ist häufig mit großem Aufwand und hohen Kosten verbunden. Um die Cloudinfrastrukturen mehrerer Anbieter zu verbinden (Cloud Federation) sind nicht nur entsprechende Standards notwendig, sondern auch Klarheit im Hinblick auf Fragen der Servicequalität und Sicherheit. Cloud Federation trägt maßgeblich zur Interoperabilität und Portabilität von Services bei (Schubert/Jeffery 2012, S.60 f.). Es gibt mehrere Initiativen die sich für die Entwicklung von Standards für Cloudumgebungen einsetzen (z. B. OASIS, DMTF, ETSI).

BETRIEBLICHE HERAUSFORDERUNGEN**2.**

Abgesehen von technischen Herausforderungen sehen sich Organisationen sowohl bei der Durchführung von Big-Data-Analysen als auch bei der Nutzung von Cloud Computing mit betrieblichen Herausforderungen konfrontiert. Im Kontext von Big Data gehören dazu primär fehlende überzeugende Einsatzszenarien, mangelndes Fachwissen und Datenschutzbedenken. Aber auch relativ hohe Kosten und Rivalitäten zwischen Abteilungen spielen eine nicht unbedeutende Rolle.

Eine besondere Herausforderung, die sowohl die Durchführung von Big-Data-Analysen als auch die Nutzung von Cloud Computing betrifft, ist die Umsetzung der jeweils relevanten rechtlichen Anforderungen. In diesem Zusammenhang ist eine Auseinandersetzung unter anderem mit Fragen des Datenschutzes, des Urheberrechts, des Vertragsrechts und des Wettbewerbsrechts notwendig. Eine ausführliche Diskussion rechtlicher Herausforderungen erfolgt im Kapitel VI.3.

WIRTSCHAFTLICHKEIT

2.1

Schroeck et al. (2012, S. 15) betonen beispielsweise, dass ihre Ergebnisse auf eine zentrale Herausforderung bei der Einführung von Big Data hindeuten. Bei dieser Herausforderung handelt es sich um die Erstellung eines belastbaren Business Case. Im gegenwärtigen Wirtschaftsumfeld haben Organisationen nur eine geringe Motivation, in neue Technologien zu investieren, solange sich keine messbaren Vorteile ergeben. Während im Rahmen der BARC-Studie 36 % der Befragten angaben, dass fehlende überzeugende Einsatzszenarien ein wesentliches Problem im Hinblick auf Big Data darstellen (Bange et al. 2013, S. 26), waren es in der TDWI-Studie 28 % (Russom 2011, S. 12). Beide Studien identifizierten auch hohe Kosten als zentrale Herausforderung. Während ein Drittel der Teilnehmer der BARC-Studie in den Kosten ein zentrales Big-Data-Problem sah, nahmen in der TDWI-Studie 43 % der Befragten die Kosten als wesentliche Barriere wahr. Ein Teil der Teilnehmer der TDWI-Studie gab im Rahmen der Befragung an, dass vor allem die Betriebskosten von Hadoop hoch sind (Russom 2011, S. 12).

Der belastbare Business Case stellt eine zentrale Herausforderung dar, die auch im Rahmen der Experteninterviews intensiv diskutiert wurde. Viele potenzielle Anwender stehen einem der befragten Experten zufolge vor dem Problem, dass sie erkannt haben, dass Potenzial in ihren Daten steckt, sie aber nicht wissen, wie sie den vermeintlichen Datenschatz monetarisieren können. Von einem anderen Experten kam die Einschätzung, dass die Anbieter von Big-Data-Lösungen bei potenziellen Anwendern ganz spezielle Szenarien im Kopf erzeugen, die zwar auf den ersten Blick sehr interessant erscheinen, aber nicht auf »normale« Unternehmen übertragbar sind. Diese Diskrepanz ruft laut dem befragten Experten bei den potenziellen Anwendern rasch Ernüchterung hervor. In der Werbung für Big-Data-Lösungen stehen Anwendungsformen wie Onlinemarketing oder Sentimentanalysen häufig im Zentrum der Aufmerksamkeit, wohingegen sich potenzielle Anwender in der Regel eher Verbesserungen bei den operativen Prozessen erhoffen. Aus Sicht der Experten ist es schwierig, Szenarien aufzuzeigen, die zu einem nennenswerten Mehrwert führen.

Auf Grundlage dieser Kenntnisse überrascht es nicht, dass in vielen Organisationen der Investitionswille fehlt (Interxion 2013, S. 8; Russom 2011, S. 12; Schäfer et al. 2012, S. 48; Schroeck et al. 2012, S. 15; TCS 2013, S. 33). Während im Rahmen der Interxion-Studie ein Drittel der Befragten angab, dass die Unternehmensleitung nicht bereit ist, in Big Data zu investieren, waren es bei der TDWI-Studie 38 % der Teilnehmer. TCS (2013, S. 33) setzte die mangelnde Bereitschaft der Unternehmensleitung, in Big Data zu investieren, auf Rang 6 der bedeutsamsten Herausforderungen im Big-Data-Umfeld. Eine weitere wesentliche Barriere im Hinblick auf die Nutzung von Big Data liegt in der Tatsache, dass die IT-Abteilungen dringendere Anforderungen sehen und andere Prioritäten gesetzt haben

(Interxion 2013; Schäfer et al. 2012). Interxion (2013, S.8) identifizierte andere Prioritäten bei der IT-Abteilung als wichtigste Herausforderung bei der Einführung von Big Data. 45 % der Befragten gaben im Rahmen der Studie an, dass sie diese Herausforderung für besonders wichtig halten. Fehlendes Budget oder andere Prioritäten sehen Schäfer et al. (2012, S.48) nach Datenschutz und Datensicherheit als zweitwichtigste Barriere im Hinblick auf Big Data.

FACHWISSEN

2.2

Im Hinblick auf betriebliche Herausforderungen spielen auch Defizite beim erforderlichen Fachwissen eine wichtige Rolle. Sowohl in der TDWI-Studie als auch in der BARC-Studie wurde fehlendes Fachwissen als wichtigste Herausforderung im Kontext von Big Data identifiziert. Im Rahmen der TDWI-Studie gaben 46 % der Befragten an, dass eine unzureichende Personalausstattung und mangelnde Fähigkeiten zu den zentralen Barrieren im Zusammenhang mit der Analyse von Big Data gehören (Russom 2011, S.12). Im Rahmen der BARC-Studie wurden fehlendes technisches Fachwissen (46 %) und fehlende fachliche Expertise (44 %) als zentrale Probleme bei der Nutzung von Big Data gesehen (Bange et al. 2013, S.26). Schäfer et al. (2012, S.48) kamen zu dem Ergebnis, dass fehlende Expertise im Unternehmen nach Datenschutz und Datensicherheit sowie fehlendem Budget oder dringenderen Prioritäten die dritt wichtigste Barriere für Big Data darstellt. Auch bei der Interxion-Studie gab mit 32 % knapp ein Drittel der Befragten an, dass in den jeweiligen Unternehmen das erforderliche Fachwissen fehlt und die Beschaffung zu teuer ist (Interxion 2013, S.8). Der eklatante Mangel an Experten am Arbeitsmarkt wird im Rahmen von Anforderungen an die Rahmenbedingungen (Kap. VI.5.1) genauer diskutiert. Auch Schroeck et al. (2012, S.15) weisen auf die Bedeutung von technischen und analytischen Fähigkeiten bei der Einführung von Big Data hin. TCS (2013, S.33) weist auf das für Big-Data-Analysen erforderliche Fachwissen der IT-Experten hin, bzw. auch auf die notwendige Überzeugung der IT-Mitarbeiter, dass zusätzliches Fachwissen überhaupt erforderlich ist. Die Herausforderungen befinden sich auf den Rängen 11 bzw. 15 der TCS-Liste.

Auch im Rahmen der Experteninterviews wurden Defizite beim erforderlichen Fachwissen thematisiert. Unterstrichen wurde, dass die derzeit entstehenden Lehr- und Studienangebote allein aufgrund der Vielschichtigkeit der Anforderungen an entsprechende Mitarbeiter nicht in der Lage sein werden, das Defizit auszugleichen und für eine Entspannung am Arbeitsmarkt zu sorgen. Von entsprechenden Mitarbeitern werden Kenntnisse der Datenanalyse, Kenntnisse des skalierbaren Datenmanagements sowie Kenntnisse in der jeweiligen Anwendungsdomäne erwartet. Einer der befragten Experten betonte, dass die Defizite beim für die Durchführung von Big-Data-Analysen erforderlichen Fachwissen auch eine technische Herausforderung darstellen. Der Experte forderte Techno-



logien, bei denen man sich um Fragen der Systemprogrammierung als Anwender keine Gedanken machen muss. Eine Abfragesprache sollte es dem Experten zufolge erlauben, deklarativ zu formulieren, was mit den Daten passieren soll, ohne sich um die Infrastruktur im Hintergrund kümmern zu müssen. Die befragten Experten waren sich weitgehend einig, dass es wichtig ist, dass im Prinzip ein normaler Business-Analyst mit den Big-Data-Lösungen umgehen können muss.

PLANUNG UND DURCHFÜHRUNG

2.3

Auch Rivalitäten und mangelndes Vertrauen zwischen Abteilungen spielten eine wichtige Rolle im Hinblick auf die zentralen Herausforderungen bei der Einführung und Nutzung von Big-Data-Lösungen. TCS (2013, S.33) setzte beispielsweise die mangelnde Bereitschaft von Fachbereichen, »ihre« Daten mit anderen zu teilen, auf Rang 1 der auf Grundlage einer umfassenden Studie erstellten Herausforderungsrankliste. Die Herausforderung, Vertrauen zwischen den Data Scientists auf der einen Seite und den Fachbereichsleitern auf der anderen Seite aufzubauen, befindet sich auf Rang 4 der Rangliste. Insgesamt scheinen sich Entscheider nur schwer davon überzeugen zu lassen, auch auf Daten und nicht nur auf ihre Intuition zu vertrauen. Auf Rang 12 der TCS-Liste ist die entsprechende Herausforderung zu finden. In vielen Fällen werden die Chancen und Risiken von Big Data nicht ausreichend verstanden (Schroeck et al. 2012, S. 15).

TCS (2013, S. 33) geht auch auf eine Reihe von Herausforderungen ein, mit denen sich Organisationen bei der Planung und Durchführung konkreter Big-Data-Analysen konfrontiert sehen. Im Hinblick auf die Planung wird einerseits die Identifikation der wichtigsten Investitionsbereiche und andererseits die Organisation von Big-Data-Aktivitäten als Herausforderung wahrgenommen. Die entsprechenden Herausforderungen befinden sich auf den Rängen 9 bzw. 8 der TCS-Liste. Mitverantwortlich für die Probleme bei der Planung der Big-Data-Nutzung dürfte die mangelnde Bekanntheit von Einsatzmöglichkeiten, Technologien und Anbietern sein (Schäfer et al. 2012, S.48). Bei der Durchführung werden die Auswahl der »richtigen« Daten für eine bestimmte Fragestellung und der »richtigen« Verwendung der Analyseergebnisse als besonders herausfordernd angesehen. Die entsprechenden Herausforderungen wurden auf die Ränge 3 bzw. 10 der Herausforderungsliste gesetzt.

KONTROLLE

2.4

Eine zentrale betriebliche Herausforderung im Zusammenhang mit der Nutzung von Cloud Computing stellt die Tatsache dar, dass die Vertrauenswürdigkeit von Anbietern nur schwer feststellbar ist. Rund ein Viertel der Teilnehmer an der IDC-Studie gab an, dass mangelndes Vertrauen eine wesentliche Barriere im

Hinblick auf die Nutzung von Clouddiensten ist (Bradshaw et al. 2012, S. 34). Mit rund 22 % sehen nur knapp weniger Studienteilnehmer ein Problem in der unzureichenden Kontrollierbarkeit von Änderungen. Damit eng verbunden sind Herausforderungen im Hinblick auf Informationssicherheit und Datenschutz. Insbesondere die Angst vor Datenverlust ist im Umfeld von Cloud Computing groß und spielt sowohl bei Organisationen, die bereits Public-Cloud-Dienste nutzen, als auch bei Organisationen, für die das aktuell kein Thema ist, eine wichtige Rolle (KPMG/BITKOM 2013, S. 23 f.). Während 55 % der Vertreter von Organisationen, die bereits Public-Cloud-Dienste nutzen, angaben, dass die Aussage, dass sie Angst vor Datenverlust haben, für sie voll und ganz oder eher zutrifft, gilt das für 79 % der Vertreter von Organisationen, für die der Einsatz von Public-Cloud-Diensten aktuell kein Thema ist.

Darüber hinaus werden die Angst vor einem Verlust an IT-Fachwissen (KPMG/BITKOM 2013) und fehlende Unterstützung vor Ort (Bradshaw et al. 2012) als Barrieren genannt. Rund die Hälfte der Studienteilnehmer der von KPMG und BITKOM durchgeführten Studie gab an, dass die Aussage, dass Angst vor IT-Know-how-Verlust besteht, für sie voll und ganz oder eher zutrifft (KPMG/BITKOM 2013, S. 23 f.). Rund 23 % der im Rahmen der IDC-Studie Befragten sehen in fehlendem lokalen Support eine wichtige Barriere (Bradshaw et al. 2012, S. 34). Bei Widerstand innerhalb der IT-Abteilung, dem Fehlen von Informationen über geeignete Cloudlösungen und Bedenken im Hinblick auf die Vereinbarkeit der Nutzung von Cloud Computing mit der Unternehmenskultur handelt es sich um weitere Herausforderungen, die laut »Cloud-Monitor 2013« im Kontext der Nutzung von Public-Cloud-Diensten relevant sind. Außerdem zweifeln 43 % der Vertreter von Organisationen, für die die Nutzung von Public-Cloud-Diensten derzeit nicht zur Diskussion steht, zumindest eher am Nutzen von Cloud Computing. Selbst bei den Vertretern von Organisationen, die bereits Public-Cloud-Dienste nutzen, ist dieser Wert mit 29 % recht hoch. Im Rahmen der IDC-Studie gaben rund 18 % der Teilnehmer an, dass sie Zweifel am Nutzen von Cloud Computing hegen (Bradshaw et al. 2012, S. 34).

RECHTLICHE HERAUSFORDERUNGEN

3.

Neben technischen und betrieblichen Herausforderungen spielen für Unternehmen und staatliche Stellen und Behörden, die sich mit der Durchführung von Big-Data-Analysen bzw. der Nutzung von Cloud Computing beschäftigen, rechtliche Herausforderungen eine zentrale Rolle. Im Kontext von Big Data und Cloud Computing stellen sich vor allem Fragen, die mit dem Datenschutz, einem möglichen Eigentum an Daten, mit der Ausgestaltung von Verträgen, mit einer möglichen Haftung und mit dem Wettbewerb zusammenhängen. Zahlreiche rechtliche Vorschriften, vor allem aus dem Umfeld des Datenschutzes, stellen



aufgrund der strengen, teils unklaren Haftungssituation hohe Ansprüche an die Datengovernance. Maßnahmen wie die umfassende Überwachung von Datenströmen im Internet, die im Kontext des Themenkomplexes Internetgovernance im Kapitel VI.5.4 diskutiert werden, erschweren die Datengovernance häufig zusätzlich. Besonders problematisch ist, dass hinsichtlich zahlreicher Fragen Rechtsunsicherheit herrscht bzw. sich das anwendbare Recht als wenig geeignet erweist und einer Anpassung an den Stand der Technik bedarf.

Für 45 % der im Rahmen der von KPMG und BITKOM durchgeführten Studie Befragten, die Public-Cloud-Computing bereits nutzen, trifft die Aussage, dass bestehende rechtliche und regulatorische Bestimmungen gegen die Nutzung von Clouddiensten sprechen, voll und ganz oder eher zu (KPMG/BITKOM 2013, S. 23 f.). Im Hinblick auf die Aussage, dass die Rechtslage unklar ist und dies eine Hürde darstellt, stimmten 42 % der Studienteilnehmer voll und ganz oder eher zu. Von den Teilnehmern, für die der Public-Cloud-Einsatz derzeit nicht zur Diskussion steht, sehen 49 % eine Herausforderung in der unklaren Rechtslage. Die Unvereinbarkeit mit rechtlichen und regulatorischen Bestimmungen stellt aus Sicht von 46 % der Studienteilnehmer, die keine cloudbasierten Lösungen nutzen, eine Hürde dar.

Wird für eine Big-Data-Analyse auf Public-Cloud-Dienste zurückgegriffen, ist es nicht selten der Fall, dass der Gerichtsstand im Ausland liegt. Mit 31,7 % gab knapp ein Drittel der im Rahmen der IDC-Studie Befragten an, dass es sich dabei um eine Barriere handelt, die von der Nutzung von cloudbasierten Lösungen abhält (Bradshaw et al. 2012). Knapp ein Viertel der Studienteilnehmer (23,8 %) gab an, dass die Tatsache, dass der Ort der Speicherung und Verarbeitung von Daten unbekannt oder zumindest nicht kontrollierbar ist, eine Barriere darstellt.

DATENSCHUTZ

3.1

Datenschutz spielt als Herausforderung sowohl im Kontext der Durchführung von Big-Data-Analysen (Bange et al. 2013, S. 26; Schäfer et al. 2012, S. 48) als auch bei der Nutzung von Cloud Computing (Bradshaw et al. 2012, S. 34) eine wichtige Rolle. In Deutschland wird der Datenschutz durch das Bundesdatenschutzgesetz (BDSG) geregelt und durch bereichsspezifische Regelungen wie das Telemediengesetz (TMG), das Telekommunikationsgesetz (TKG) und andere ergänzt. Den einfachgesetzlichen Regelungen liegt ein etabliertes verfassungsrechtliches Schutzkonzept zugrunde. Danach formulieren u.a. das sogenannte Recht auf informationelle Selbstbestimmung als Konkretisierung des allgemeinen Persönlichkeitsrechts aus Artikel 2 Abs. 1 Grundgesetz sowie spezielle Verfassungsbestimmungen konkrete Anforderungen an die Verarbeitung von Informationen und Daten. Der Umgang mit Informationen und Daten, die Personen betreffen, unterliegt damit in mehrererlei Hinsicht konkreten gesetzlichen Anforder-

rungen, darunter dem Erforderlichkeitsgrundsatz, dem Zweckfestlegungs- und Zweckbindungsprinzip sowie insbesondere Transparenzanforderungen gegenüber den Betroffenen. Die verfassungsrechtlichen Anforderungen gelten über die Drittwirkung nicht nur für staatliche Stellen, sondern auch für private Unternehmen, also auch für Anbieter von Clouddiensten.

Datenschutz bezieht sich aus gesetzlicher Sicht immer auf personenbezogene Daten (in Deutschland im Sinne des § 3 Abs. 1 BDSG), also Daten, die einer bestimmten oder bestimmbar natürlichen Person zuzuordnen sind. Die Frage der Personenbeziehbarkeit ist gerade im Kontext von Big Data jedoch ein kontrovers diskutierter Aspekt, gerade weil personenbezogene Daten und damit die Sicherung der informationellen Selbstbestimmung in einer Vielzahl von Anwendungsbereichen von Big Data eine zentrale Rolle spielen.

Nach dem BDSG steht die Verarbeitung von personenbezogenen Daten unter einem Erlaubnisvorbehalt (Markl et al. 2013). Dies bedeutet, dass einer der im Gesetz vorgesehenen Rechtfertigungstatbestände vorliegen muss. Im Rahmen einer Anonymisierung oder Pseudonymisierung können Datenverarbeitungen ggf. so gestaltet werden, dass die jeweilige Nutzung zulässig wird (Ulbricht 2012a). Bei Big Data wird bereits durch die Zusammenführung verschiedener, vormals womöglich nichtpersonenbeziehbarer Datenbestände teilweise eine De-anonymisierung (Glockner/Beyers 2012, S. 34) oder Reidentifikation (Markl et al. 2013) möglich. Grundsätzlich ist eine Verarbeitung personenbezogener Daten in der Regel nur dann zulässig, wenn der Betroffene eingewilligt hat (§ 4 BDSG) oder eine sonstige gesetzliche Ermächtigung vorliegt. Privilegierungen bestehen, wenn die Daten öffentlich zugänglich sind (§ 29 Abs. 1 Satz 1 Nr. 2 BDSG) (Ulbricht 2012a). Wenn zu unterschiedlichen Zwecken gesammelte und vorgehaltene Daten auf eine neue Weise analysiert oder mit weiteren verfügbaren Daten zusammengeführt werden, müssen zumeist Einwilligungen der identifizierbar Betroffenen eingeholt werden, weil darin eine – auch verfassungsrechtlich relevante – Zweckänderung der Nutzung liegen kann. Laut Markl et al. (2013, S. 182 f.) wird dadurch die Entwicklung von Big Data deutlich gehemmt. Dennoch ist Ulbricht (2012a) der Ansicht, dass sich Big-Data-Analysen durch ein durchdachtes Zusammenwirken verschiedener Gestaltungsmöglichkeiten durchaus im Einklang mit dem Datenschutzrecht durchführen lassen. Zusätzlich existiert auch das Argument, dass gesetzliche Vorgaben bei entsprechender Umsetzung einen Wettbewerbsvorteil darstellen können, gerade weil das unabdingbare Vertrauen der Verbraucher durch konsequenten Datenschutz gewonnen werden kann.

Aufgrund der steigenden Sensibilität bei Privatpersonen für Fragen des Datenschutzes gibt es zahlreiche unverbindliche Ergänzungen zum Datenschutzgesetz, wie beispielsweise freiwillige Selbstverpflichtungserklärungen von Verbänden und Unternehmen, deren Umsetzung empfohlen wird. Beispielsweise wird es für wichtig gehalten, über die Vorschriften des Datenschutzrechts hinaus klar zu kommu-



nizieren, welche Daten zu welchen Zwecken wie eingesetzt und verarbeitet werden. Außerdem sollte die Möglichkeit bestehen, Daten zu berichtigen oder zu löschen. Ganz grundsätzlich sollten digitale Spuren, die eine Person hinterlässt, nicht dem gehören, der sie aufzeichnet, sondern dem, der sie hinterlässt. Daten sollen nicht länger gespeichert werden als notwendig und mit einem Ablaufdatum versehen werden (The Economist 2010d). Je nach Rechtslage sollte bei den Betroffenen nach einer hinreichenden Aufklärung die Einwilligung (Opt-in) zur jeweiligen Datenerhebung bzw. die Möglichkeit zur Ablehnung (Opt-out) gegeben werden. Allerdings nennen Unternehmen oft technische Schwierigkeiten als Grund für mangelnde Transparenz und Kontrolle im Hinblick auf den Datenschutz (The Economist 2010d).

Transparenz und Kontrolle sind elementare Voraussetzungen, um Big Data nicht zu einer unkontrollierbaren Gefahr für Privatpersonen werden zu lassen. Jedoch warnen verschiedene Gruppen, u.a. auch Industrievertreter, davor, dass es dadurch zu einer Überregulierung im Bereich Datenschutz kommen könnte (Ulbricht 2012a). Allerdings ging die Rechtsprechung, u.a. was die Zweckbindung und die Befristung der Speicherung von Daten betrifft, mit Entscheidungen wie zur europäischen Richtlinie über die Vorratsspeicherung von Daten (2006/24/EG) oder zum Recht auf Vergessen (EuGH, 13. Mai 2014, C-131/12) zuletzt in eine andere Richtung. Erkennbar wird, dass der Europäische Gerichtshof bereit und willens ist, die für den Privatbereich mit der Richtlinie 95/46/EG bereits vor Jahren entstandene Regulierung auf der Grundlage von Artikel 8 EU-Charta auch verfassungsrechtlich zu untermauern. Dies belegen seit dem Urteil Bodil Lindqvist vom 6. November 2003, Rs C-101/01, neben dem genannten Google-Urteil auch die Urteile zum Personenbezug EuGH, 13. Mai 2014, C-131/1, und zur privaten Videoüberwachung (Urteil vom 11. Dezember 2014, C-212/13). Große Hoffnungen auf eine größere Rechtssicherheit für alle Beteiligten, die allerdings mit der Verschiebung bis nach der Europawahl zunächst enttäuscht wurden, liegen aktuell in der Modernisierung des europäischen Datenschutzrechts durch den Ersatz der Datenschutzrichtlinie (95/46/EG) durch die sogenannte Datenschutz-Grundverordnung.

Markl et al. (2013, S. 182 f.) weisen darauf hin, dass die datenschutzrechtlichen Regelungen zur Auftragsdatenverarbeitung (§ 11 BDSG) große Bedeutung für Big Data in der Cloud haben. Cloud Computing stellt einen typischen Fall der Auftragsdatenverarbeitung dar. Den Regelungen zufolge ist der Auftraggeber dafür verantwortlich, dass die Vorschriften des BDSG eingehalten werden und muss dies im Rahmen der technischen und rechtlichen Ausgestaltung der Datenverarbeitung sicherstellen. Aus rechtlicher Sicht ist einerseits die Ausgestaltung entsprechender vertraglicher Vereinbarungen schwierig, andererseits sind Art und Umfang der gesetzlichen Kontrollpflichten weitgehend unklar. Noch herausfordernder wird Big Data in der Cloud aus Datenschutzperspektive bei grenzüberschreitendem Bezug. Wann ein ausländischer Anbieter überhaupt deutsches

Datenschutzrecht beachten muss, ist noch völlig ungeklärt. Auch wenn sich im Kontext von Cloud Computing die Frage stellt, inwiefern der Standort, an dem sich bestimmte Daten befinden oder wo sie erhoben wurden, verlässlich ermittelt werden kann, gilt das Territorialitätsprinzip. Einerseits können Daten auf mehrere Standorte verteilt sein und andererseits kann sich ein einmal ermittelter Standort binnen kurzer Zeit ändern (Markl et al. 2013, S. 177). Die Speicherung von Daten im Ausland führt vor allem deshalb zu rechtlichen Herausforderungen, da sensible Daten in verschiedenen Ländern unterschiedlich gut geschützt werden. Dies wird in entsprechenden Hintergrundpapieren und Stellungnahmen von Zusammenschlüssen der Datenschutzaufsichtsbehörden auf europäischer und nationaler Ebene unterstrichen.⁹

Das Thema Datenschutz wurde auch im Rahmen der Interviews thematisiert. Wie im Kapitel VI.2 dargelegt wurden beispielsweise Themen wie die fehlende Kontrolle über Daten die außer Haus gegeben werden oder die starke Heterogenität der Datenschutzrichtlinien in Europa angesprochen. Ein Experte merkte an, dass die Bemühungen im Hinblick auf eine europäische Datenschutzgrundverordnung zwar in die richtige Richtung gehen, aber es vermutlich noch viele Jahre dauern wird, bis der Bereich Datenschutz in Europa harmonisiert ist. International betrachtet wäre es aus Sicht des Experten wünschenswert, wenn man sich auf wenige große Blöcke mit einheitlicher Gesetzgebung konzentrieren könnte.

Eine Harmonisierung auf europäischer Ebene wird durch die im Europaparlament bereits im Entwurf verabschiedete und im Europäischen Rat diskutierte Datenschutzgrundverordnung angestrebt. Der Trilog wird 2015 aufgenommen. Die Grundverordnung ist eine Reaktion auf neue datenschutzrechtliche Herausforderungen, die Entwicklungen wie Cloud Computing, aber z. B. auch die zunehmende Bedeutung von sozialen Netzwerken mit sich bringen. Im Rahmen einer Studie für das STOA-Panel wurden vier Kernprobleme identifiziert, die Cloud Computing im Hinblick auf den Datenschutz in Europa auf Grundlage der existierenden Datenschutzrichtlinie (95/46) mit sich bringt (Leimbach et al. 2014):

- › **Zuständigkeit und Anwendbarkeit:** Eine zentrale Eigenschaft von Cloud Computing ist, dass der Ort, an dem Daten gespeichert oder verarbeitet werden oder an dem ein Dienst läuft, unwesentlich ist. Aktuelle Datenschutzgesetze verwenden jedoch zur Bestimmung ihrer Zuständigkeit und Anwendbarkeit Konzepte, die am Ort, an dem sich zu schützende Daten physisch befinden, anknüpfen.

⁹ Vergleiche den Entschließungsantrag der Internationalen Datenschutzkonferenz (www.bfdi.bund.de/SharedDocs/Publikationen/Entschliessungssammlung/IntDSK/2012IntDSKCloudComputing.pdf?__blob=publicationFile) sowie das Arbeitspapier der Gruppe »Cloud Computing – Privacy and data protection issues (Sopot Memorandum)«, Sopot (Polen), 23./24. April 2012 (www.datenschutz-berlin.de/attachments/875/Sopot_Memorandum.12.6.12.pdf).

gelungen zum digitalen Vergessen, zur Portabilität von Daten und Diensten und zum selbstverständlichen und lückenlosen Schutz von personenbezogenen Daten (Privacy by Default, Privacy by Design). Die genauen Auswirkungen dieser Regelungen auf die Praxis sind schwer vorherzusagen.

Darüber hinaus enthält die Grundverordnung weitere Regelungen, die möglicherweise auch die Bereitstellung und Nutzung von cloudbasierten Diensten betreffen. Dazu gehören z. B. Regelungen zum digitalen Vergessen, zur Portabilität von Daten und Diensten und zum selbstverständlichen und lückenlosen Schutz von personenbezogenen Daten (Privacy by Default, Privacy by Design). Die genauen Auswirkungen dieser Regelungen auf die Praxis sind schwer vorherzusagen.

EIGENTUM AN DATEN

3.2

Aus juristischer Sicht kann kein »Eigentum« an Daten oder Informationen begründet werden. Dennoch können Daten oder Informationen unter verschiedenen gesetzlichen Voraussetzungen vor dem Zugriff oder der Verwertung durch Dritte geschützt sein. In Deutschland sind neben Nutzungsbedingungen vor allem das Urheberrecht und das ebenfalls im Urheberrechtsgesetz (UrhG) zu verortende Datenbankrecht relevant (Ulbricht 2012b). Im Kontext von Big Data spielt die Frage, ob es eine gesicherte Rechtsposition im Hinblick auf verarbeitete Daten gibt, eine wichtige Rolle. Auch auf insolvenzrechtliche Fragestellungen hätte die Klärung der Eigentumsfähigkeit Einfluss (Markl et al. 2013, S. 165).

Da nur der Datenträger, nicht aber die Daten selbst körperlich sind, können Daten nicht als Sachen im Sinne des § 90 des Bürgerlichen Gesetzbuchs (BGB) qualifiziert werden. Nach § 93 BGB sind Bestandteile, die nicht von einer Sache getrennt werden können, wesentliche Bestandteile. So betrachtet würde das Eigentum an einem Datenträger auch das Eigentum an den darauf gespeicherten Daten umfassen. Im Hinblick auf die tatsächlichen Verhältnisse und insbesondere den Bereich Cloud Computing, wäre dies kaum praktikabel. Auch könnten viele Probleme, die bei vernetzten Datenbeständen auftreten, nicht gelöst werden (Markl et al. 2013, S. 168). Sofern keine legitimierenden Schranken greifen, hängt die Zulässigkeit einer urheberrechtlich relevanten Handlung von der Zustimmung des Rechteinhabers durch Einwilligung oder durch entsprechende Ausgestaltung einer vertraglichen Vereinbarung ab (Markl et al. 2013, S. 168). Eine besondere Herausforderung im Kontext von Big Data stellt die Tatsache dar, dass in vielen Fällen weder legitimierende Schranken greifen noch eine vertragliche Einräumung von Nutzungsrechten mit angemessenem Aufwand möglich ist. In vergleichbaren Konstellationen bediente sich der Bundesgerichtshof (BGH) Markl et al. (2013, S. 176) zufolge der Konstruktion der schlichten konkludenten Einwilligung, die auf den Big-Data-Kontext übertragbar sein könnte. Da gegenüber der Rechtsprechung des BGH im Hinblick auf die Konstruktion



allerdings dogmatische Bedenken geäußert wurden, ist fraglich, ob die Konstruktion überhaupt haltbar ist.

Im Kontext von Big Data ist es nicht unüblich, dass über das Internet verfügbar gemachte Daten systematisch gesammelt und ausgewertet werden. Zahlreiche Internetportale verbieten allerdings in ihren Nutzungsbedingungen den Einsatz entsprechender, auch als Web Crawler bezeichneter, Werkzeuge. Betreiber von Portalen haben auf Grundlage des virtuellen Hausrechts die Möglichkeit, entsprechende Vorgaben für die Nutzung ihrer Internetplattform zu machen, rechtlich verbindlich sind sie allerdings nur, wenn sie im Rahmen einer Registrierung beim entsprechenden Portal vom Datensammler anerkannt wurden (Ulbricht 2012b).

Urheberrechtlicher Schutz wird nur angenommen, wenn ein entsprechendes schutzfähiges Werk vorliegt (in Deutschland § 2 UrhG). Im Rahmen von Big-Data-Analysen werden häufig Texte oder Textfragmente analysiert. Diese sind nur dann durch das Urheberrechtsgesetz geschützt, wenn sie die notwendige Schöpfungshöhe im Sinne der Individualität erreichen (Markl et al. 2013, S. 171; Nordemann et al. 2008). Vor allem bei Textfragmenten sowie bei Datenreihen ist dieses Kriterium in der Regel nicht erfüllt. In Deutschland hat der Hersteller einer Datenbank das ausschließliche Recht, die Datenbank insgesamt oder einen wesentlichen Teil davon zu vervielfältigen, zu verbreiten oder öffentlich wiederzugeben (§ 87b UrhG). Artikel 7 Abs. 1 der europäischen Datenbankrichtlinie (96/9/EG) besagt, dass »wesentlicher Teil« in qualitativer oder in quantitativer Hinsicht verstanden werden kann. Bei der Entnahme weniger Daten im Verhältnis zur Gesamtdatenbank ist durchaus denkbar, dass der Datenbankhersteller rechtlich nicht einschreiten kann (Ulbricht 2012b). Ohne Zustimmung des Datenbankherstellers dürfen zur Datengewinnung also nur unwesentliche Teile einer Datenbank übernommen werden (Markl et al. 2013, S. 173).

Aus dem Schutz von Datenbanken durch das Urheberrecht, kann kein Eigentum an Daten abgeleitet werden. Nur selten beruht der Schutz von Datenbanken auf der schöpferischen Leistung im Hinblick auf die in der Datenbank enthaltenen Elemente (Schutz eines Datenbankwerks als Sonderfall des Sammelwerks nach § 4 Abs. 2 UrhG). Viel häufiger geht es vielmehr um einen Investitionsschutz für denjenigen, der das Investitionsrisiko für die Erstellung einer Datenbank trägt (Sui-generis-Datenbankschutz nach §§ 87 UrhG). Einer Entscheidung des Europäischen Gerichtshofs zufolge werden in diesem Fall allerdings Kosten, die im Rahmen der Datenerhebung angefallen sind, nicht berücksichtigt. Unklarheit herrscht im Hinblick auf die Abgrenzung zwischen Datenerhebung und Datenverwaltung (Markl et al. 2013, S. 173). Das Urheberrecht begründet zwar einen Schutz vor Vervielfältigung, Verbreitung und öffentlicher Wiedergabe, nicht jedoch ein eigentumsähnliches Vollrecht. Durch den Sui-generis-Datenbankschutz sollte ein spezieller Investitionsschutz geschaffen werden, nicht aber eine rechtliche Zuordnung von Daten zu Personen (Markl et al. 2013, S. 169).

Während bei Big-Data-Lösungen, die innerhalb Deutschlands genutzt werden, in der Regel deutsches Urheberrecht zu beachten sein wird, ist die Beurteilung in Fällen mit grenzüberschreitendem Bezug oft schwierig. Vor allem im Bereich des Cloud Computing lassen sich Handlungen nicht immer einem bestimmten Schutzterritorium zuordnen. Die entsprechenden kollisionsrechtlichen Fragestellungen sind noch nicht abschließend geklärt (Markl et al. 2013, S. 171). Bei der Verwendung von cloudbasierten Lösungen ergeben sich in diesem Zusammenhang allerdings noch weitere Herausforderungen. Beispielsweise ist unklar, wer der Eigentümer von Anpassungen ist, die der Nutzer entsprechender Lösungen vorgenommen hat. In der Studie von Bradshaw et al. (2012, S. 34) gab mit 21,3 % immerhin gut jeder Fünfte an, dass das eine Barriere im Hinblick auf die Nutzung von Clouddiensten darstellt.

Im Rahmen der Experteninterviews wurde betont, dass im Kontext von Big Data Urheberrechtsfragen vor allem dann nicht ausreichend beantwortet sind, wenn Daten aggregiert oder wenn neue Daten aus vorhandenen Daten abgeleitet werden. Bei Big-Data-Anwendungen ist das allerdings eher die Regel als die Ausnahme. Ein Experte wies beispielsweise darauf hin, dass im Rahmen eines Forschungsprojekts im Bereich Big Data beschlossen wurde, entsprechende Daten zu löschen, da es keinen Rechtsrahmen gibt. Es wäre unklar gewesen, wem die Daten gehört hätten, die aus der Aggregation und Auswertung von Daten zweier kommerzieller Partner entstanden sind.

Im Kontext von Cloud Computing spielen Urheberrechtsfragen eine geringere Rolle. Die größte Herausforderung stellen Fragen des geistigen Eigentums dar, da das Immaterialgüterrecht, ähnlich wie das Datenschutzrecht, zur Bestimmung seiner Anwendbarkeit am physischen Ort anknüpft (Leimbach et al. 2014). Entsprechende Fragen stellen sich beispielsweise, wenn Nutzer von Clouddiensten auf Basis dieser Dienste neue Anwendungen entwickeln oder wenn Clouddienste vom Nutzer an die eigenen Bedürfnisse angepasst oder anderweitig verändert werden. Noch ist weitgehend unklar, wie Fragen des geistigen Eigentums im Kontext von Cloud Computing gelöst werden können. Im Rahmen einer Studie für das STOA-Panel (Leimbach et al. 2014) wurde die Problematik ausführlich diskutiert.

VERTRAGSGESTALTUNG

3.3

Die Verteilung der Rechte und Pflichten aus einem Vertragsverhältnis liegen nach dem Grundsatz der Vertragsfreiheit in den Händen der beteiligten Parteien. Trotzdem ist die Vertragsgestaltung im Kontext von Big Data aufgrund der zahlreichen technischen Fragen und der Berührungspunkte zum Datenschutz- und Urheberrecht kompliziert (Markl et al. 2013, S. 183). Vertragsrechtliche Fragen müssen jeweils im Einzelfall beurteilt werden. Die grundlegenden Herausforderungen der Vertragsgestaltung im Zusammenhang mit Big Data werden im Kontext der Nutzung von cloudbasierten Lösungen besonders deutlich.



Laut Markl et al. (2013, S.184) ist es nicht abwegig, grundsätzlich von einer mietrechtlichen Beurteilung eines Cloud-Computing-Vertrags auszugehen. In gewisser Weise schuldet der Cloudanbieter die Überlassung und Erhaltung einer Mietsache. Quantitative und qualitative Vereinbarungen erfolgen in der Regel im Rahmen von SLAs. Unklar ist, ob auf das gesetzliche Leitbild des Mietrechts geschlossen werden könnte, wenn bestimmte Aspekte nicht vertraglich geregelt wurden. Im Fall der Verfügbarkeit von Übertragungskapazitäten würde das bedeuten, dass die Mietsache ausnahmslos und uneingeschränkt geschuldet wird. Unklar ist auch, wen das Risiko trifft, wenn die Vertragsdurchführung durch technische Fehler gestört wird (Markl et al. 2013, S.184). Um die Gefahr von Regressansprüchen auszuschließen, werden in der Praxis Haftungsausschlüsse oder Wartungsklauseln in allgemeinen Geschäftsbedingungen geregelt. Während Verträge privatautonom ausgehandelt werden können, unterliegen allgemeine Geschäftsbedingungen in Deutschland den strengen Anforderungen des § 307 BGB. Die Erarbeitung rechtswirksamer Haftungsklauseln stellt für viele Cloudanbieter eine große Herausforderung dar.

Im Kontext von Cloud Computing spielen neben Hauptleistungspflichten auch nachvertragliche Pflichten eine wichtige Rolle. Beispielsweise bietet es sich für Nutzer von Cloudangeboten an, vertraglich sicherzustellen, dass alle Daten nach Beendigung des Vertrags zurückgegeben werden müssen und keine Daten beim Anbieter verbleiben. Unterbleibt das, müsste es bei konsequenter Anwendung mietrechtlicher Vorschriften ein mit dem Wegnahmerecht durch den Mieter (§ 539 Abs. 2 BGB) vergleichbares Recht bestehen. Inwieweit dieses Recht klauselmäßig modifiziert werden darf und ob im Falle eines wirtschaftlichen Schadens Deliktrecht anwendbar ist, bedarf allerdings einer genaueren Untersuchung (Markl et al. 2013, S.185).

Im Fall eines schuldhaften Datenverlusts durch den Cloudanbieter stehen dem Nutzer unterschiedliche Schadenersatzansprüche zu, die sich einerseits durch die Verletzung von Vertragspflichten und andererseits aus dem Deliktrecht oder anderen Rechten ergeben können. Während im Fall, dass die Daten technisch reproduzierbar sind, die Wiederherstellungskosten zur Schadensberechnung zugrunde gelegt werden, wird in Fällen, in denen das nicht möglich ist, die Vermögenseinbuße vom Richter nach freiem Ermessen geschätzt (Markl et al. 2013, S.186). Entsprechend hoch kann die Haftung für schuldhaften Datenverlust im Kontext von cloudbasierten Big-Data-Angeboten ausfallen.

HAFTUNG

3.4

Aufgrund der zunehmenden Bedeutung von Daten in Entscheidungsprozessen und in Anbetracht der daraus entstehenden Gefahren, ist Klarheit im Hinblick auf Fragen der Haftung im Kontext von Big Data unerlässlich. Eine besondere

Herausforderung stellt hier die Tatsache dar, dass es unklar ist, ob es qualitative Mindeststandards für Informationen gibt und welche das sind (Markl et al. 2013, S. 187). Voraussetzung für die Haftung ist, dass ein Fehlverhalten vorliegt. Was ein Fehlverhalten ist bzw. wann Fehlinformationen vorliegen, lässt sich ohne Mindeststandards nur unzureichend klären. Konsens besteht laut Markl et al. (2013, S. 187) zumindest darüber, dass Daten bzw. die daraus abgeleiteten Informationen nachweislich richtig sein müssen.

Zu einer vertraglichen Haftung und damit zu einem Schadenersatzanspruch kommt es in der Regel bei Pflichtverletzungen des Schuldverhältnisses (§§ 280 Abs. 1, 241 Abs. 2 BGB). Markl et al. (2013, S. 188) weisen darauf hin, dass es im Big-Data-Bereich bisher keine höchstrichterliche Entscheidung bezüglich einer vertraglichen Haftung gibt. Inwiefern Entscheidungen in verwandten Bereichen auf Big Data übertragbar sind, ist unklar. Ein deliktischer Schadenersatzanspruch kommt dann in Betracht, wenn Fehlinformation zur Verletzung eines Rechtsgutes im Sinne des § 823 Abs. 1 BGB führt.

WETTBEWERB

3.5

Im Rahmen von Standardverträgen behielten sich viele Anbieter von Clouddiensten vor, Vertragsbedingungen einseitig zu ändern (Bradshaw et al. 2010). Solche Änderungen wurden in der Regel mittels Verweis auf die Veröffentlichung der aktualisierten Bedingungen auf der Webseite des Anbieters kommuniziert. Die Weiternutzung der entsprechenden Dienste wurde als Zustimmung interpretiert.

Vertragsrechtliche Fragen im Zusammenhang mit dem Ende der Nutzung eines Clouddienstes hängen davon ab, ob der Vertrag wie vereinbart endete oder ob er aufgrund eines Vertragsbruchs beendet wurde. Für beide Fälle sollten vertragliche Regelungen beschreiben, wie vorzugehen ist, insbesondere im Hinblick auf die Daten des Cloudnutzers. Drei wesentliche Punkte sollten berücksichtigt bzw. adressiert werden (Bradshaw et al. 2010; Hon et al. 2012):

- › Datenaufbewahrung: Es muss sichergestellt werden, dass der Nutzer auch nach Beendigung des Vertragsverhältnisses für eine bestimmte Zeit auf seine Daten zugreifen kann. Bradshaw et al. (2010) weisen darauf hin, dass Anbieter von Clouddiensten im Hinblick auf die Datenaufbewahrung unterschiedliche Ansätze verfolgen: Sie gewähren eine entsprechende Nachfrist für den Zugriff auf die Daten nach Beendigung des Vertrags, sie löschen die Daten mit dem Ende des Vertrags oder sie verfolgen einen Hybridansatz.
- › Übertragung von Daten und Diensten: Der Nutzer des Clouddienstes möchte seine Daten oder Dienste möglicherweise zu einem anderen Anbieter übertragen. Es gibt eine Vielzahl an Werkzeugen, die Unterstützung bei der Übertragung von Daten oder Diensten von einem Anbieter zu einem anderen in Aus-



sicht stellen. Nutzern muss klar sein, dass nicht alle Schwierigkeiten bei der Übertragung ihren Ursprung bei den Anbietern haben. Auch Anpassungen von Diensten durch den Nutzer können eine Übertragung erschweren.

- › Datenlöschung: In der Regel möchte der Nutzer des Clouddienstes sicherstellen, dass seine Daten beim Anbieter vollständig gelöscht werden. Wichtig ist, dass neben Metadaten auch Daten, die aus Gründen der effizienten Datenverarbeitung repliziert wurden, gelöscht werden.

Aus wettbewerbsrechtlicher Sicht stellt die Tatsache, dass Cloudanbieter Hindernisse aufbauen können, die den Wechsel zu einem anderen Anbieter erschweren (Vendor Lock-in), eine zentrale Herausforderung dar (Leimbach et al. 2014). Die Herausforderung hat sowohl eine rechtliche als auch eine technische Dimension. Aus rechtlicher Sicht ergeben sich Fragen zur Aufbewahrung, Übertragung und Löschung von Daten bzw. Diensten. Aus technischer Sicht gilt es Fragen im Hinblick auf Standards zu klären, die die Interoperabilität fördern würden.

WEITERE RECHTLICHE HERAUSFORDERUNGEN

3.6

Im Zusammenhang mit Cloud Computing stellt auch die Wahl der Gerichtsbarkeit eine rechtliche Herausforderung dar. Nutzer von Public-Cloud-Angeboten müssen davon ausgehen, dass die Daten in verschiedenen Rechenzentren gespeichert und verarbeitet werden. Daher ist es nicht unwahrscheinlich, dass die Verarbeitung von bestimmten Daten in die Zuständigkeit von verschiedenen nationalen Rechtsprechungen fällt. Dies kann zu rechtlichen Unsicherheiten führen, auch wenn die Nutzungsbedingungen auf die Wahl der Gerichtsbarkeit eingehen. In vielen Fällen obliegt die Wahl der Gerichtsbarkeit den Vertragsparteien. Die Wahl der Gerichtsbarkeit wirkt sich nachhaltig auf das Vertragsverhältnis aus. Die Wahrscheinlichkeit, dass Haftungsausschlüsse und -beschränkungen in den Nutzungsbedingungen anerkannt werden, ist beispielsweise bei kalifornischen Gerichten deutlich größer als bei Gerichten in der EU. Bradshaw et al. (2010) fand bei der Untersuchung von 31 Nutzungsbedingungen heraus, dass bei 15 die Wahl auf die Gesetze eines US-Bundesstaates fiel, meistens auf jene von Kalifornien. Bei elf fiel die Wahl auf ein Mitgliedsland der EU und bei fünf Nutzungsbedingungen wurde das lokale Gesetz der Kunden gewählt beziehungsweise das Thema Gerichtsbarkeit nicht angesprochen. Bradshaw et al. (2010) weisen darauf hin, dass Cloudanbieter tendenziell Gerichtsbarkeiten bevorzugen, die für eine kurze Frist für die Einbringung von Beanstandungen stehen und bei denen die Verfahrenskosten relativ hoch sind. Zum Teil schützen Verbraucherschutzvorschriften Kunden aus der EU vor solchen Maßnahmen.

In der Regel bietet EU-Recht große Rechtssicherheit. Die Klärung der Anwendbarkeit von EU-Recht ist vor allem dann wichtig, wenn Clouddienste von Anbietern außerhalb der EU genutzt werden. Viele Anbieter bevorzugen das US-ameri-

kanische Recht. Das Ausmaß an Haftungsausschlüssen und -beschränkungen hängt direkt mit dem Herkunftsland eines Cloudanbieters zusammen (Bradshaw et al. 2010). Bradshaw et al. (2010) weisen darauf hin, dass von den untersuchten Cloudanbietern aus den USA alle versuchten, die Haftung so weit wie möglich zu beschränken. Von den Anbietern aus der EU wurde die Haftung nur in Fällen von höherer Gewalt eingeschränkt. Die Studienergebnisse belegen, dass zahlreiche Anbieter den Haftungsumfang zu begrenzen versuchen. Hon et al. (2012) betonen, dass Nutzer von Clouddiensten, die in der Position sind, um Verträge verhandeln zu können, Haftungsausschlüsse und -beschränkungen zu vermeiden versuchen. Das EU-Recht erlaubt den Ausschluss sowie die Beschränkung der Haftung nicht im selben Ausmaß wie in den USA. Artikel 23 der Datenschutzgrundverordnung adressiert beispielsweise Fragen der Entschädigung von Personen, die durch rechtswidrige Verarbeitung von Daten geschädigt wurden.

Vor allem bei Vertragsbeziehungen mit Privatpersonen ist es wichtig, dass die Sprache, die für die Formulierung von Nutzungsbedingungen verwendet wird, verständlich ist. Vereinfachungen und Vereinheitlichungen könnten hier Abhilfe schaffen. Probleme in dieser Hinsicht können genauso zu Misstrauen gegenüber Cloud Computing führen wie mangelnde Transparenz im Zusammenhang mit Informationssicherheit, Leistungsmessdaten, Prüfungsbefugnissen, der Verwendung von Metadaten, der Identität von Unterauftragnehmern und dem physischen Ort von Daten bzw. Diensten (Leimbach et al. 2014).

GESELLSCHAFTLICHE HERAUSFORDERUNGEN

4.

Im Folgenden sollen die – vor allem aus Big Data erwachsenden – verschiedenen gesellschaftlichen Herausforderungen betrachtet werden. Grundsätzlich gilt es dabei zu bedenken, dass es sich hierbei um eine Vielzahl von Themen handelt, die aufgrund ihrer teilweise gesellschaftsphilosophischen Dimensionen hier nur angerissen werden können. Ein Beispiel hierfür ist die Frage nach der Autonomie des Menschen als handelndes Subjekt. Auffallend dabei ist, dass diese Herausforderungen von den üblichen und vor allem auf Unternehmen ausgerichteten Studien kaum oder gar nicht adressiert werden, sodass auch keine empirischen Belege für deren Relevanz existieren. Jedoch zeigen der Diskurs in den Medien (z. B. Schirmacher 2013) sowie weiterführende Publikationen (z. B. Bunz 2012; Geiselberger/Moorstedt 2013), dass die zugrundeliegenden Fragestellungen durchaus kontrovers diskutiert werden. Ein anderer beachtenswerter Punkt ist, dass die hier aufgezeigten Herausforderungen teilweise aufeinander aufbauen oder sich ergänzen. Dies bedeutet, dass eindeutige Abgrenzungen nicht immer möglich sind und es fließende Übergänge zwischen den einzelnen Aspekten gibt. In diesem Sinn sind die hier entwickelten Kategorien nur Versuche zur Strukturierung der Diskussion, transportieren aber nicht den Anspruch, diese umfassend



abzudecken. Ebenso gilt es aber auch zu berücksichtigen, dass hier verwendete Begriffe und Konzepte in den verschiedenen Anwendungsgebieten teilweise unterschiedliche Bedeutungen haben können. So stehen Souveränität und Autonomie bei Konsumenten für die Möglichkeit, Entscheidungen ohne direkte oder indirekte Beeinflussung treffen zu können. Im Kontext von Technologie steht der Begriff der Souveränität heute vor allem für die Fähigkeit einer Gesellschaft oder Volkswirtschaft, verwendete Technologien zu verstehen und deren Vorgehensweise nachvollziehen und gegebenenfalls verifizieren zu können. Insgesamt zeigt sich, dass die hier thematisierten gesellschaftlichen Herausforderungen nicht durch einzelne Diskussionen oder Maßnahmen zu lösen sind. Vielmehr wird deutlich, dass die zunehmende Digitalisierung des Lebens einen kontinuierlichen Diskurs der resultierenden Fragestellungen sowohl für den Einzelnen als auch für die Gesellschaft als Ganzes erfordert.

VERTRAUEN UND VERLÄSSLICHKEIT

4.1

Sowohl Big Data als auch Cloud Computing erfordern ein generell hohes Vertrauen in eine technologische Lösung, die entsprechenden Anbieter und deren Leistungsfähigkeit.

Für Cloud Computing wird diese Herausforderung am sichtbarsten durch das Konzept der Ortsunabhängigkeit, da ein Nutzer nicht nachvollziehen kann, wo und wie seine Daten gespeichert und verarbeitet werden. Das daraus resultierende Gefühl des Kontrollverlustes ist eine wesentliche Ursache für die Vorbehalte gegenüber Cloud Computing (NTT Communications 2014). Dementsprechend setzt die Nutzung von Clouddiensten ein hohes Maß an Vertrauen in den Anbieter voraus. Dies wird aber umso schwieriger zu erreichen sein, je komplexer die zugrundeliegenden Dienstleistungsstrukturen sind, z. B. wenn ein Anbieter wiederum andere Anbieter von Clouddiensten nutzt, um sein eigenes Angebot bereitzustellen. Alles in allem setzt dies voraus, dass private Konsumenten, Firmen oder Verwaltungseinrichtungen durch verschiedene Maßnahmen wie Zertifizierungen, vertragliche Zusagen oder ähnliches geneigt sind, prinzipiell dem Anbieter und der zugrundeliegenden Technologie zu vertrauen. Ein solches Vertrauen kann jedoch sehr schnell erschüttert werden, wie die Veröffentlichungen um die Abhörpraktiken der NSA und anderer Geheimdienste zeigen. Deutlich wird, dass dies zu Änderungen im Verhalten von Kunden wie dem Verzicht oder dem Ausweichen auf andere Lösungen führen kann (NTT Communications 2014). Gleichzeitig sind weitere Aspekte, wie die Fähigkeit, technologische Entwicklungen zu bewerten, oder die Zuverlässigkeit der Leistungserbringung angesichts der bisherigen Dominanz nur einiger weniger Hersteller aus bestimmten Regionen ein Diskussionsthema geworden, das deshalb im Folgenden thematisiert wird.



Während bei Cloud Computing das Vertrauen in den Anbieter und seine technologischen Fähigkeiten im Vordergrund stehen, spielt bei Big Data das Vertrauen in die Technologie und die zugrundeliegenden Prinzipien eine wesentlich größere Rolle. Von Bedeutung sind dort vor allem statistische Verfahren, die auf Korrelationen als Messgröße für Zusammenhänge setzen. Tatsächlich besteht aber – wie bei den betrieblichen Herausforderungen gezeigt – ein gewisses Maß an begründeter Skepsis, dies als Entscheidungsgrundlage zu nutzen; denn ein blindes Vertrauen in Korrelation kann ebenfalls zu massiven Problemen führen, weil Korrelation und Kausalität nicht das Gleiche sind (z. B. Boyd/Crawford 2011; Mayer-Schönberger/Cukier 2013, S. 70 ff.). Auch wenn zwischen verschiedenen Ereignissen eine hohe Korrelation festgestellt werden kann, müssen diese in keinerlei Zusammenhang stehen – beispielsweise die folgende Beobachtung: »Wenn du Eis isst, hast du ein erhöhtes Risiko zu ertrinken« (Croll 2013). Diese Aussage ist stark korreliert, wenn man den Verbrauch von Speiseeis und die Anzahl an Ertrinkenden in Zusammenhang setzt. Jedoch gibt es einen gemeinsamen Grund für die jeweils steigenden Zahlen, in diesem Fall den Sommer. Ohne Beachtung dieses Aspektes, könnten daraus falsche Schlussfolgerung und Handlungsempfehlungen abgeleitet werden. Ähnliche Probleme können entstehen, falls durch die Größe der einzelnen Datenmengen jederzeit eine Korrelation hergestellt werden kann, eine sogenannte »False Correlation« (Boyd/Crawford 2011). Dass es sich hierbei nicht nur um akademische Gedankenspiele handelt, zeigen erste Beispiele aus dem Big-Data-Umfeld selbst. Diese unterstreichen dabei auch die Schwierigkeiten, solche Probleme zu erkennen und zu bewerten. Das wohl zurzeit am meisten diskutierte Beispiel ist der Google Flu Trend, dessen Ziel es ist, die Entstehung von Grippeepidemien anhand von Suchergebnissen möglichst zeitnah zu erkennen und so Gegenmaßnahmen zu ermöglichen (Ginsberg et al. 2009). Wie eine Überprüfung der Vorhersagen gezeigt hat, neigt diese Anwendung dazu, die tatsächliche Stärke und Ausbreitung teilweise signifikant zu überschätzen. In letzter Konsequenz könnten daher Handlungen, die auf diesen Ergebnissen basieren, zu Fehlallokationen führen, die mindestens monetäre Folgen, aber auch weiter reichende Konsequenzen haben können. Als Gründe werden vermutet, dass die zugrundeliegenden Korrelationen nicht zutreffend sind oder durch Änderungen in der Datenbasis, also den Suchergebnissen von Google, betroffen sind. Da diese Technologien aber wiederum nicht offenliegen, können diese Fragestellungen nicht abschließend beantwortet werden (Lazer et al. 2014).

Festzuhalten bleibt, dass die Nutzung von Big Data und Cloud Computing eine ganze Reihe von Fragen impliziert, die sich aus den grundsätzlichen Eigenschaften der Technologien ergeben. Folglich weisen fast alle weiteren gesellschaftlichen Herausforderungen in unterschiedlicher Art und Weise auf diese beiden grundlegenden Aspekte des Vertrauens und der Verlässlichkeit der eingesetzten Technologien und Anbieter zurück.

VERENGUNG DES SICHTBAREN

4.2

Als Konsequenz des Einsatzes von Big Data kann es zu einer Reihe von Herausforderungen kommen, die am besten unter dem Begriff der »Verengung des Sichtbaren« zusammengefasst werden können. Hierzu zählt das Problem, dass durch Big Data die Informationen und Zusammenhänge für Nutzer stark gefiltert werden und so nur eine beschränkte Sicht auf die Daten möglich ist. Dies tritt in fast allen Anwendungsfeldern in unterschiedlichsten Ausprägungen auf. Die Ursache ist, dass nicht alle Daten verarbeitet werden können. Ein Aspekt dabei sind kapazitative Gründe, die dazu führen, dass Daten ausgewählt bzw. »gefiltert« werden müssen. Dies impliziert die Fragen, wer die Daten auswählt und wie die Auswahl erfolgt. Ein anderer Grund ist, dass nur das, was auch gemessen werden kann, berücksichtigt wird. Dies führt zur generellen Frage nach der Messbarkeit von relevanten Ereignissen oder Interaktionen.

Die erste Herausforderung der »Filterung« hat sehr unterschiedliche Ausprägungen. Gegenwärtig am bekanntesten ist das sogenannte Phänomen der »Filter Bubble« (Pariser 2012). Dieser Ausdruck besagt, dass insbesondere Nutzer von Diensten, die ihre Ergebnisse unter der Verwendung von Big-Data-Technologien optimieren, nur diejenigen Informationen angezeigt bekommen, die durch ihr vorheriges Verhalten als relevant eingestuft wurden. Dies kann dazu führen, dass nur Informationen angezeigt werden, die bereits bestehende Meinungen unterstützen, aber andere Blickwinkel völlig außen vor lassen. Beispiele dafür sind intelligente Suchfunktionen, die schon möglichst vor der vollständigen Eingabe das entsprechende Suchresultat einblenden (Kerr/Earle 2013). Dieses Phänomen wird alternativ auch als Schaffung von »Small Worlds« bezeichnet (Leberecht 2013), wobei sich Menschen immer weiter in kleine, isolierte Welten zurückziehen und keinerlei Interaktion zwischen verschiedenen Gruppen stattfindet.

Filterproblematiken und Verengungen betreffen aber nicht nur Nutzer großer Internetdienste wie Google, Amazon oder Facebook, deren Geschäftsmodelle auf Daten basieren. Auch in anderen Anwendungsbereichen gibt es sehr ähnliche oder im Ergebnis vergleichbare Problemstellungen. So zeigen Beispiele, dass Unternehmen, die ihre Entscheidungen auf der Basis von Datenauswertungen in Kombination mit einer effektiven und transparenten Nutzung dieser Ergebnisse treffen, ihre Wettbewerbsfähigkeit erhöhen können, wenn sie dadurch ihre bestehenden Prozesse und Geschäftsmodelle optimieren (Ross et al. 2013). Jedoch besteht auch die Gefahr, dass durch die Konzentration auf Daten zum operativen Betrieb eine gewisse Filterung stattfindet, bei der unter anderem aufgrund der schier Menge an Daten und Ergebnissen sowie der resultierenden Filterung durch Algorithmen übergeordnete strategische Aspekte aus dem Blick geraten können. Es besteht insbesondere die Gefahr, dass Entwicklungen, die nicht direkt mit dem Kerngeschäft verbunden sind, aber dennoch eine Chance oder Risiko hierfür darstellen, ausgeblendet werden. Langfristig kann dies wiederum Nachteile bedeuten, denn es

droht der Verlust des Zufalls und damit einer Form der Kreativität. Beide sind jedoch zentrale Elemente für Innovationen. Der Grund ist, dass bei der Entwicklung der Analyseverfahren vor allem diejenigen Daten und Erkenntnisse genutzt werden, die schon bekannt sind. Dinge, die nicht bekannt bzw. berücksichtigt werden oder für die keine Daten erhoben werden, fallen so aus dem »Raster«, da hier keine Korrelationen existieren. Zwar kann dies in einem gewissen Umfang adressiert werden, macht aber deutlich, dass Big-Data-Analysen nicht darauf ausgelegt sind, selbst kreativ zu sein (Leberecht 2013).

Ein anderes verwandtes Phänomen kann man in der Wissenschaft finden, das sogenannte »Overfitting« (Pentland 2012a; The Economist 2013b). Gerade in sehr datenintensiven Wissenschaftsbereichen wie der Pharmaforschung gibt es dieses Problem schon länger. Hier werden die Vielzahl der verfügbaren Daten und mögliche Korrelationen so lange betrachtet und bewertet, bis deren Ergebnis perfekt zu bestehenden Daten und Annahmen passen, auch auf die Gefahr hin, den tatsächlichen Zusammenhang nicht zu erkennen (Pentland 2012a; The Economist 2013b).

Daneben gibt es noch weitere Aspekte, die diese Verengung begünstigen, nämlich die technischen Grenzen der Informationsverarbeitung sowie der Zugang zu und die Verfügbarkeit von Daten. Die technischen Limitationen führen einerseits dazu, dass nicht alle Informationen gespeichert werden können. Andererseits können auch nicht alle vorhandenen Informationen gleichzeitig verarbeitet werden. Zwar können cloudbasierte Lösungen diese Herausforderungen von Big Data insbesondere für kleinere Anwender wie Mittelständler oder Privatpersonen teilweise lösen, aber es bleibt die Frage nach dem Zugang zu allen maßgeblichen Daten für eine datenbasierte Entscheidung. Der Zugang zu Daten kann aus verschiedenen Gründen eingeschränkt sein. So wird zwar in der gegenwärtigen Diskussion der Eindruck erweckt, dass große Informationsmengen auf Knopfdruck bereitstünden. Dies mag auch, was die reine Menge an Informationen betrifft, richtig sein, jedoch sind oftmals gerade die für Unternehmen relevanten Daten über Personen oder andere Organisationen aus verschiedensten, zum Teil durchaus berechtigten Gründen nicht zugänglich. Um die Potenziale gerade für kleinere Anwender zu heben, müssen hier noch Wege gefunden werden, den Zugang zu Daten jenseits von Open Data zu erhöhen.

Zusammenfassend zeigt dies, dass verschiedene Herausforderungen im Zusammenhang mit der Filterung und Verengung des danach noch Sichtbaren existieren. Dazu zählt, nach welchen Prinzipien und Prioritäten Daten ausgewählt und analysiert werden. Ein Beispiel für diese Problematik ist, dass Daten durch die Wahl der Datenquellen schon verzerrt sein können oder einen Bias enthalten. So sind beispielsweise viele Inhalte – wie Twitter-Meldungen – automatisch gefiltert, sodass einige Aspekte grundsätzlich fehlen können (Boyd/Crawford 2011). Andere Beispiele zeigen, dass die Einbeziehung spezieller Datenbanken – wie

Statistiken zu Verbrechen oder Gesundheit – dazu führt, dass die Daten von vornherein einen Bias enthalten, da einzelne Bevölkerungsgruppen dort überrepräsentiert sind (Doctorow 2014). Auch die Probleme von Google Flu zeigen, dass Big-Data-Analysen und darauf aufbauende Maßnahmen solche Verzerrungen erzeugen und verstärken können (Lazer et al. 2014). Hinzu kommt, dass Daten durchaus kontextabhängig sein können, wodurch ihre Verwendung in anderen Kontexten zu vollkommen irreführenden oder irrelevanten Ergebnissen führen kann (Ang et al. 2013; Boyd/Crawford 2011). Ebenso spielt eine Rolle, welche Fragen von wem an welche Daten gestellt werden. Denn es zeigt sich, dass durch die Fragestellung und die damit verknüpften Erwartungen die Ergebnisse beeinflusst werden. Denn schließlich sind Daten, anders als oft unterstellt, nur so neutral und objektiv, wie derjenige, der sie auswertet (Boyd/Crawford 2011). Dieses erkenntnistheoretische Problem, das schon lange in der Wissenschaftstheorie diskutiert wird, wird in der gegenwärtigen Diskussion um Big Data weitgehend ausgeblendet.

AUTONOMIE UND BEEINFLUSSUNG

4.3

Die Verengung des Sichtbaren ist eng verbunden mit zwei weiteren zentralen gesellschaftlichen Herausforderungen: einem möglichen Verlust von Autonomie sowie einer möglichen Beeinflussung und Manipulation. Beide ergeben sich aus den Möglichkeiten insbesondere von Big Data, das Verhalten von Gruppen und Individuen vorherzusagen und beeinflussen zu können.

Aus entsprechenden Verhaltensvorhersagen leitet sich eine Reihe von Anwendungsszenarien ab, welche die zugrundeliegende Problematik verdeutlichen. Am bekanntesten sind die heute schon in der Anwendung befindlichen Empfehlungssysteme. Mit deren Hilfe werden Nutzern von Onlineshops Kaufempfehlungen gegeben – basierend auf ihrem vergangenen Kauf- und Suchverhalten sowie dem Verhalten vergleichbarer Kunden. Auf diese Weise sollen mögliche weitere Bedürfnisse angesprochen werden, um so Verkauf und Umsatz zu erhöhen. Dies ist vergleichbar mit den Systemen zum Vorschlag von Suchbegriffen und ähnlichen anderen Anwendungen, wie z. B. die Schaltung von Werbeanzeigen (Konstan/Riedl 2012). Prinzipiell kann dies aus zwei Blickwinkeln betrachtet werden. Einerseits ist es für Konsumenten bequem, relevante Angebote zu erhalten und nicht danach suchen zu müssen. Andererseits wird durch die getroffenen Vorschläge die Wahlmöglichkeit begrenzt und damit die Handlungsautonomie eingeschränkt, da hier bewusst ein Set von Optionen vorgeschlagen wird, das auf den Käufer, seine früheren Präferenzen, seine ökonomischen Ressourcen und Lebensumstände zugeschnitten ist. Vergleichbar zur Diskussion um Journalismus versus nutzergenerierte Nachrichten (Bunz 2012, S. 85 ff.) könnte hier argumentiert werden, dass auch Fachgeschäfte eine solche Auswahl treffen. Diese



ist jedoch weniger selektiv und individuell und damit wesentlich weniger in die Handlungsautonomie eingreifend.

Auch bei der Verwendung von datenbasierten Analysen in Unternehmen kann es zur Einschränkung der Handlungsautonomie kommen. So sind viele Systeme darauf ausgelegt, Handlungsoptionen auf Basis der erkannten Korrelationen vorzuschlagen. Neben der Gefahr, dass durch das Vertrauen auf Korrelationen wesentliche Kausalzusammenhänge übersehen werden, können durch diese vermeintlich evidenzbasierten Vorschläge andere Handlungsalternativen ausgeblendet werden (Manovich 2012). Dieser Aspekt spielt auch in der Wissenschaft eine nicht unwesentliche Rolle. Durch die Fokussierung auf evidenzbasierte Daten würde in gewisser Weise die Autonomie der Wissenschaft eingeschränkt, denn es könnten nur noch Dinge erforscht werden, für die Daten vorliegen und die messbar sind. Zwar ist die damit verbundene erste Euphorie über ein neues Paradigma in der Wissenschaft, wo unter anderem das »Ende der Theorie« prophezeit wurde (Anderson 2008), verflogen, jedoch bleibt der Trend zu evidenzbasierten Erkenntnissen bestehen. Dies beinhaltet damit weiterhin die Gefahr, dass alternative Ansätze zur Wissensgenerierung marginalisiert werden können. Dies würde die Autonomie der Wissenschaft und damit die Möglichkeiten des wissenschaftlichen Erkenntnisgewinns einschränken (Boyd/Crawford 2011; Manovich 2012; Pentland 2012a).

Problematisch sind aber vor allem die Möglichkeiten zur Beeinflussung der gesellschaftlichen Entwicklung, die sich aus den Möglichkeiten der detaillierten Verhaltensvorhersage ergeben. Ein Beispiel dafür ist der Wahlkampf von Barack Obama im Jahr 2011. Hier wurden zum ersten Mal Big-Data-Technologien in einem Wahlkampf mit dem Ziel eingesetzt, die beschränkten Ressourcen möglichst sinnvoll zu allokalieren. Dazu wurden Wähler aufgrund verschiedener Datensätze nach bestimmten Kriterien (Unterstützung für Obama, Möglichkeit eines Meinungswechsels sowie Wahrscheinlichkeit der Wahlteilnahme) bewertet. Dadurch wurden vor allem solche Wähler identifiziert, die möglicherweise noch für eine Wahlentscheidung zugunsten von Obama zu gewinnen waren. Diese wurden dann gezielt durch persönliche Kontakte (Telefon, Hausbesuche) angesprochen (Issenberg 2012; Quack 2013). Tatsächlich hat diese gezielte und damit individualisierte Kampagne Wirkung gezeigt; ob sie aber wahlentscheidend war, kann nicht belegt werden, da eine Wahlentscheidung von vielen Faktoren abhängt. Dennoch zeigt dieses Beispiel sowohl den Trend zu Individualisierung auf Basis von Daten als auch die Möglichkeit auf, individuelle Ansprachen mit gezielten Anreizen zu verbinden. Ein zweiter, bisher jedoch noch nicht möglicher Schritt wäre die Gestaltung und Platzierung von Anreizen in der Weise, dass sie kaum noch bewusst wahrgenommen werden könnten. Dies birgt die große Gefahr, dass es zu einer unterschweligen Beeinflussung von Verhalten und damit zu einer nachhaltigen Beeinflussung und Manipulation gesellschaftlicher Wil-



lensbildungsprozess kommen könnte. Prinzipiell ist die zugrundeliegende Methodik – die Steuerung durch Anreize – ein schon lange genutztes Prinzip. Während ein solches Vorgehen in einigen Fällen – wie etwa bei der Förderung eines gesunden Lebensstils durch Anreizprogramme von Krankenkassen – aus gesellschaftlicher Sicht sinnvoll sein mag, um die Kostenentwicklung im Gesundheitssystem zu begrenzen, stellt es in anderen Fällen – wie beispielsweise der politischen Willensbildung – einen Eingriff in demokratische Prozesse dar. Deswegen sind Kritiker grundsätzlich gegen diese Art der gesellschaftlichen Einflussnahme (Morozov 2013, S. 302 ff.) und sehen diese »präemptive Steuerung« insbesondere im Zusammenhang mit Big Data besonders kritisch (Morozov 2011).

Insgesamt ist dennoch nicht davon auszugehen, dass diese Möglichkeiten der Beeinflussung die Menschen zu willenlosen, steuerbaren Objekten machen, da das menschliche Verhalten zu komplex ist. Notwendig ist aber ein kontinuierlicher Diskurs, um die gesellschaftspolitische Sensibilität für eine solche mögliche Entwicklung aufrechtzuerhalten. Zudem sollte diskutiert werden, in welchem Umfang und in welchen Anwendungsgebieten der Einsatz solcher Möglichkeiten gesellschaftlich gewünscht und akzeptiert ist und wo dessen Grenzen liegen.

DISKRIMINIERUNG UND TEILHABE

4.4

Eine zentrale gesellschaftliche Herausforderung im Zusammenhang mit Big Data ist die Möglichkeit, datenunterstützte Entscheidungen zur Diskriminierung von Personen zu nutzen. Dies ist aus verschiedenen Gründen bedenklich. Der erste ist, dass solche Diskriminierungen in der Regel auf Eigenschaften einer Person wie Geschlecht, Alter oder Zugehörigkeit zu bestimmten Gruppen basiert und somit dem Gleichheitsgrundsatz widersprechen (Weichert 2013). Dadurch können die Entwicklungsmöglichkeiten von Personen oder Gruppen eingeschränkt werden und zu einer steigenden gesellschaftlichen Segregation beitragen. Dies ist besonders bedenklich, da neben der Problematik von Korrelation und Kausalitäten bei solchen Analysen auch Fehlertoleranzen von mehreren Prozent auftreten (Narayanan/Shmatikov 2009). Daraus resultierend können sich falsche Einschätzungen ableiten.

Diskriminierung durch vorhandene Daten ist im Bereich von Big Data die bekannteste Art der Diskriminierung. Hierunter wird verstanden, dass mit der bestehenden Datenlage und deren Analyse Vorhersagen für einzelne Personen oder Personengruppen getroffen werden, die diese nicht beeinflussen können. Im Gegenzug zur Einflussnahme geht es hierbei darum, dass nicht nur Handlungen der Personen beeinflusst und gesteuert werden sollen, sondern dass aus diesen Vorhersagen Maßnahmen abgeleitet werden, welche die Entfaltungs- und Handlungsspielräume beschränken und ohne Wissen oder Einspruchsmöglichkeit der Betroffenen umgesetzt werden. Beispiele für massive Eingriffe sind die sogenann-

ten No-Fly-Listen, die zu deutlichen Einschränkungen der Reisefähigkeit führen, sowie die Kreditvergabe, bei der von privaten Firmen ein sogenanntes Scoring, das aus verschiedensten Datensätzen zu Wohnort, Bezahlgewohnheiten und vielem mehr besteht, erstellt wird, welches darüber entscheidet, ob und zu welchen Bedingungen Personen einen Kredit erhalten. Parallelen dazu weist die Missbrauchserkennung bei Kreditkarten auf, wo die Software auf Basis von Daten zu Nutzungsgewohnheiten ungewöhnliche Transaktionen verweigert. So sinnvoll dies insgesamt sein kann, so problematisch kann dies für einzelne Betroffene sein. Besonders heikel ist dabei, dass die Verfahren nicht offengelegt werden müssen und es nur schwer oder gar nicht möglich ist, Widerspruch einzulegen (Weichert 2013). Im Extremfall wird hierdurch die Unschuldsvermutung ausgehebelt (Kerr/Earle 2013), eine Kernidee der modernen Rechtslehre. Prinzipiell bliebe dann eine Person durch eventuell falsche, aber auf jeden Fall kaum nachvollziehbare Bewertungen dauerhaft diskriminiert. Daher ist die Möglichkeit zur Einsicht und Korrektur eine zentrale Forderung.

Ein anderer Fall von Diskriminierung kann entstehen, wenn keinerlei oder nur sehr geringe Daten über Individuen vorhanden sind. Diese Form der Diskriminierung wird in der bisherigen Literatur eher wenig beachtet. In einer hoch vernetzten Welt, in der eine Vielzahl an Daten gesammelt und verarbeitet bzw. genutzt wird, können Gruppen oder Einzelpersonen ohne Daten aus dem »Raster« fallen. Eine Konsequenz ist, dass bei einer komplett auf Daten basierenden Entscheidungsfindung, sei es im politischen oder wirtschaftlichen Bereich, diejenigen, die keine Daten generieren, ignoriert beziehungsweise nicht erfasst werden und hierdurch höchstwahrscheinlich gravierende Nachteile erleiden (Lerman 2013). Beispiele sind fehlende Daten über Personen, die dazu führen, dass eine Gruppe von Personen beispielsweise bei der Ausrichtung des öffentlichen Gesundheitswesens benachteiligt wird, oder dass Personen aus strukturell schwachen Regionen nicht ausreichend an Partizipationsprozessen teilnehmen können und deren Belange nicht berücksichtigt werden. Prinzipiell können neben Unterschieden in den Nutzungsgewohnheiten, wie sie beispielsweise mit verschiedenen Altersgruppen oder regionalen Aspekten wie der Verfügbarkeit von ausreichenden Internetanschlüssen assoziiert werden, auch weitere Faktoren wie Einkommens- oder Bildungsunterschiede Gründe darstellen. Vergleichbare Problemstellungen werden schon seit Längerem unter dem Begriff »Digitale Kluft« beziehungsweise dem Widerpart »Digitale Teilhabe« diskutiert.

VULNERABILITÄT UND ABHÄNGIGKEIT

4.5

Gesellschaftliche Vulnerabilität durch die Verwendung komplexer IT-Systeme ist in den letzten Jahren durch verschiedene Ereignisse wie das Auftreten bestimmter »Internetangriffe« (z. B. Stuxnet, Duqu) (Wessling 2013) sowie die darauf



folgenden Tests und Berichte zu einem wichtigen Thema geworden (Stahl 2013). Insbesondere die nachgewiesene Verletzbarkeit von kritischen Infrastrukturen – wie etwa die Energieversorgung – durch IT-Angriffe stand dabei im Vordergrund. Die Angriffsmöglichkeiten werden durch die steigende Vernetzung durch Smart-City- oder Smart-Grid-Konzepte noch steigen. Aber auch die Verletzbarkeit von Industrieanlagen ist im Zuge dieser Diskussion immer deutlicher geworden, da die zunehmende Integration von Systemen als Folge der Flexibilisierung von Prozessen schwerwiegende Rückwirkungen durch den Ausfall von Zulieferern haben können. Zwar handelt es sich in erster Linie um eine technische beziehungsweise betriebliche Herausforderung (Kap. VI.1 u. VI.2), doch in einer globalisierten Wirtschaftswelt, in der nicht nur einfache Verwertungs- und Wertschöpfungsketten, sondern komplexe Wertschöpfungsnetzwerke existieren (Iansiti/Levien 2004), kann dies auch gravierende Folgen für die Gesellschaft als Ganzes haben.

Mit dem Einsatz von Cloud Computing und Big Data wird sich dieser Trend zur Integration und Flexibilisierung sowie einer damit verbundenen Zunahme von Komplexität weiter fortsetzen. Ein Beispiel für die resultierenden Folgen sind die Wechselwirkungen bei Angeboten von Diensten in Cloud-Computing-Umgebungen, die wiederum auf andere Dienste aufbauen, wodurch sehr komplexe Strukturen entstehen, bei denen kaum noch erkennbar ist, welche Abhängigkeiten und Lieferbeziehungen bestehen. Dadurch ergeben sich neue Schwächen und Angriffspotenziale. Ein Beispiel dafür ist, dass mit einer weiteren Integration und zunehmender Komplexität der Wertschöpfungsnetzwerke die Gefahr besteht, dass selbst die Ausfälle nicht direkt vorgelagerter, aber dennoch wichtiger Teile des Netzwerkes Kaskadeneffekte erzielen könnten, die dann eine gesamte Wirtschaft und Gesellschaft empfindlich treffen. Eine andere Gefahr besteht darin, dass die zunehmenden Datenströme missbraucht werden können. Dies betrifft sowohl die Möglichkeit, die enthaltenen privaten Informationen von Kunden zu missbrauchen, als auch die Absicht, durch Datenmanipulationen falsche Reaktionen hervorzurufen, die nicht nur einzelne Firmen, sondern im Beispiel der Smart Cities ganze Stadtviertel betreffen könnten. Die technischen Grundlagen der Angriffe sind zwar bekannt und resultieren entweder aus der relativen Neuartigkeit dieser Technologien, wie beispielsweise im Fall der Verletzbarkeit von NoSQL-Datenbanken, oder aus den inhärenten Konzepten von Cloud Computing und Big Data, wie z. B. der Multilevelstruktur, oder der Größe der Daten und den damit einhergehenden Problemen. Die zunehmende Komplexität der Anwendungen, insbesondere bei der Kombination von Cloud Computing und Big Data, führt jedoch dazu, dass existierende Sicherheitskonzepte aufgrund des hohen Grades der Komplexität der Anwendung nur noch schwer oder fast gar nicht anzuwenden sind.

Die Diskussion um mögliche gesellschaftliche Vulnerabilitäten führte insbesondere im Zusammenhang mit der NSA-Affäre zu einer neu aufgelebten Debatte über »technologische Souveränität«. Der Begriff wird schon seit längerem benutzt, um die technologische Abhängigkeit von anderen Ländern und Regionen zu beschreiben, bei der sowohl ökonomische als auch sicherheitspolitische Gründe im Vordergrund stehen. Problematisch an diesem Begriff ist, dass er nur sehr vage und sehr unterschiedlich definiert ist. Jedoch kam es in letzter Zeit unter der zunehmenden Erkenntnis der Verletzbarkeit von IT-basierten Systemen und Infrastrukturen auch in Europa und insbesondere in Deutschland zu einer Renaissance dieser Idee, teilweise auch unter der Bezeichnung IT-Souveränität. So nutzte 2010 Innenminister de Maizière in seinen 14 Thesen zur Netzpolitik den Begriff der technologischen Souveränität, den er durch die Forderung nach strategischen, nationalen Kernkompetenzen in der IT- und Internettechnologie sowie einer starken IT-Wirtschaft zu fassen versucht (Maizière 2010). Es folgten darauf seitens des Innenministeriums verschiedene Initiativen, die später nicht mehr weiterverfolgt wurden. Dennoch wurde das Thema der technologischen Souveränität auch im Umfeld der Internet-Enquetekommission diskutiert und tauchte dann wieder in dem als Reaktion auf die NSA-Praktiken veröffentlichten Acht-Punkte-Plan (Bundesregierung 2013) der letzten Bundesregierung auf. Zumeist wird der Begriff dabei nicht konzeptionell aufgearbeitet, es stehen aber immer wieder Ideen, wie die Beherrschung oder Kompetenz von Schlüsseltechnologien sowie eine starke nationale IT-Wirtschaft als Garant dieser Souveränität, im Fokus. Vergleichbare Vorstellungen gibt es auch auf der europäischen Ebene, insbesondere im Zusammenhang mit den sogenannten Key Enabling Technologies (KET) (EK 2011) sowie im Zug der Reaktionen auf die NSA-Aktivitäten wie dem Beschluss des Europäischen Rates zur Digitalen Agenda vom Oktober 2013 (ER 2013), ohne jedoch den Begriff der Souveränität direkt zu verwenden.

Jenseits des Problems, dass der Begriff für industriepolitische Zwecke missbraucht werden kann, stellt sich jedoch die Frage, ob der Fokus auf eine nationale Kompetenz und Beherrschung von Technologie sowie einer starken nationalen IT-Wirtschaft wirklich geeignet ist, die damit angesprochenen Herausforderungen zu lösen. Denn in Zeiten globalisierter Verwertungs- und Wertschöpfungsnetzwerke sowie den dazugehörigen Waren- und Datenströmungen erscheint diese Ideen wenig sinnvoll, da auch der Ausfall anderer Regionen oder Nationen durch Kaskadeneffekte dafür sorgen würde, dass Deutschland kurz- oder mittelfristig davon ebenfalls betroffen wäre. Gründe sind beispielsweise fehlende Zulieferungen für die Produktion oder ausbleibende Daten- und Zahlungsströme in der Finanzbranche. Daher ist es notwendig, die Idee einer nationalen Souveränität weiter in Richtung globaler Lösungsansätze zu entwickeln, zumal in einer globalen, arbeitsteiligen und sehr verzweigten IT-Industrie offen ist, ob überhaupt eine nationale technologische Souveränität erreicht werden kann.

STRUKTURELLE HERAUSFORDERUNGEN

5.

Im letzten Abschnitt sollen strukturelle Problemstellungen adressiert werden. Darunter fallen Fragestellungen, die sich ähnlich wie die gesellschaftlichen Herausforderungen auf die Gesellschaft oder Wirtschaft als Ganzes beziehen. Jedoch handelt es sich dabei in der Regel um Rahmenbedingungen, die einen direkten oder indirekten Einfluss auf mögliche Realisierung der Potenziale von Cloud Computing und Big Data haben. Diese können in der Regel durch entsprechende Maßnahmen angegangen werden. Dies kann auf regionaler, nationaler, europäischer oder gar globaler Ebene notwendig sein, sodass hier vor allem Lösungsansätze unter Einbeziehung aller relevanten Akteure möglich sind. Zu diesen Herausforderungen zählen die Verfügbarkeit von Humankapital und Infrastrukturen sowie die digitale Wettbewerbsfähigkeit und die Internetgovernance.

HUMANKAPITAL

5.1

Die Entwicklung einer ausreichenden Fachkräftebasis ist aus mehreren Gründen eine wichtige Herausforderung. Insbesondere im Kontext von Big Data spielt der Mangel an geeigneten Fachkräften eine wichtige Rolle. Big Data wie auch Data-Warehousing setzen durch die Anwendung von statistischen Verfahren Fachwissen voraus, das nicht in allen Unternehmen vorhanden ist (Ortega 2013). Laut TCS (2013, S. 33) stellt die Rekrutierung von Spezialisten (Data Scientists), die mit großen Datenmengen umgehen können, eine zentrale Herausforderung dar (Rang 5 der TCS-Herausforderungsrangliste). 11 % der Teilnehmer an der TDWI-Studie gaben an, dass vor allem Hadoop-Experten sehr schwer zu finden sind (Russom 2011). Aber auch mittel- und langfristig spielt die Entwicklung des verfügbaren Humankapitals eine wesentliche Rolle als Grundbedingung für die Wettbewerbsfähigkeit Deutschlands in einer digitalisierten Welt. Dies bezieht sich zum einen auf eine ausreichende Zahl von Fachkräften in der IT-Branche selbst, zum anderen aber auch auf Fachkräfte und Nutzer in den Anwenderindustrien. Gerade diese Anwenderindustrien – wie beispielsweise Automobil- und Maschinenbau –, die heute schon zu einem großen Teil IT-basierte Produkte entwickeln, sind besonders wichtig für die Wettbewerbsfähigkeit in Deutschland.

Wohl ist in der Vergangenheit schon eine Vielzahl von Maßnahmen in Angriff genommen worden, die dieses Problem beheben sollen. Dazu gehören neben Aktionen zur Steigerung der Attraktivität der sogenannten MINT-Fächer (Mathematik, Informatik, Naturwissenschaft und Technik) in Schulen beispielsweise auch Maßnahmen zur Steigerung des Frauenanteils in solchen Berufen. Angesichts der Breite dieser Maßnahmen stellt sich eher die Frage nach neuen kreativen Wegen zur Erschließung weiterer Personengruppen. Beispielfähig werden hier ältere Arbeitnehmer oder Studienabbrecher genannt, die entweder durch Weiter-

bildung gewonnen werden könnten oder denen durch die Anerkennung erbrachter Studienleistungen ein schneller Zugang zu einer formalen Ausbildung ermöglicht werden kann. Einzelne Pilotprojekte gab es dazu schon, nun müsste eine Ausweitung dieser Maßnahmen geprüft werden.

Ein anderer wesentlicher Aspekt sind die durch Cloud Computing und Big Data ausgelösten Änderungen in den Qualifikationsanforderungen. Für Cloud Computing sind diese schon untersucht worden (z. B. Laugesen et al. 2012). Deren Erkenntnisse werden auch teilweise schon umgesetzt. Beispiele dafür sind die geänderten Zielsetzungen im eSkills-Programm der Europäischen Kommission oder die neuen Angebote zu Cloud Computing innerhalb des BMWi-geförderten e-Kompetenznetzwerks (e-Business-Lotsen). Anders sieht die Situation bei Big Data aus, wo bisher kein einheitliches Verständnis der im Umfeld von Big Data relevanten neuen Profile besteht. Angemerkt sei hier, dass bei diesen Profilen neben den technologischen Möglichkeiten und Potenzialen auch die damit verbundenen Problemstellungen berücksichtigt werden sollten, sowohl auf Entwickler- als auch auf Anwenderseite. Dies reicht dabei von der Einbeziehung neuer Ansätze wie »Security by Design« bzw. »Privacy by Design« in die Curricula für Entwickler bis hin zur Schulung von Konsumenten bzw. Bürgern, die sie befähigt, mögliche Konsequenzen der Nutzung solcher Technologien zu erkennen.

Im Hinblick auf Big Data prognostizieren Manyika et al. (2011) einen großen Bedarf an Datenanalyseexperten. Die Anforderungen an Data Scientists sind hoch. Neben Methodenkenntnissen in Datenanalyseverfahren und technischen Kompetenzen im Bereich des skalierbaren Datenmanagements sind auch praktische Kompetenzen in der Systemimplementierung erforderlich (Markl et al. 2013). Zusätzlich benötigen Data Scientists Fachkompetenzen in der jeweiligen Anwendungsdomäne. Wie im Kapitel VI.2.2 bereits angedeutet wird vermutet, dass die derzeit entstehenden Lehr- und Studienangebote aufgrund der Vielschichtigkeit der Anforderungen nicht in der Lage sein werden, Data Scientists in ausreichender Anzahl zu qualifizieren. Neben Ausbildungsmaßnahmen sind auch technische Maßnahmen erforderlich, die die Anforderungen an Data Scientists reduzieren. Im Prinzip sollte ein Businessanalyst mit den Big-Data-Lösungen umgehen können. Zentral ist in diesem Zusammenhang beispielsweise eine Abfragesprache, die es erlaubt, deklarativ zu formulieren, was mit den Daten passieren soll, ohne sich um die Infrastruktur im Hintergrund sorgen zu müssen.

INFRASTRUKTUR/BREITBANDVERSORGUNG

5.2

Sowohl die Durchführung von Big-Data-Analysen als auch die Nutzung von Cloud Computing setzen die Verfügbarkeit und Funktionstüchtigkeit technischer Infrastruktur voraus. Neben der Energieversorgung spielen das Internet und die dafür erforderliche Netzwerkinfrastruktur eine wesentliche Rolle. Im Rahmen



der IDC-Studie wurden langsame Internetverbindungen von 18 % der Befragten als große Herausforderung für das Cloud Computing genannt (Bradshaw et al. 2012, S. 34). Auch andere Studien, insbesondere zu Cloud Computing, unterstreichen die Bedeutung von ausreichender Bandbreite sowie deren Verlässlichkeit (z. B. Fielder et al. 2012, S. 76; Schubert/Jeffery 2012, S. 5).

In den letzten Jahren sind dabei Fortschritte innerhalb der EU und der einzelnen Mitgliedstaaten zu beobachten. Das ist insofern wenig verwunderlich, als der Ausbau der Breitbandversorgung ein zentraler Punkt der Digitalen Agenda ist. Auch in Deutschland genießt das Thema mit der Breitbandstrategie eine hohe Priorität, die sich auch im Koalitionsvertrag der derzeitigen Regierung widerspiegelt (CDU/CSU/SPD 2013, S. 34 ff.). Jedoch lassen sich bei genauerer Betrachtung einige Probleme erkennen. So schneidet Deutschland in einzelnen Untersuchungen recht gut ab (BITKOM 2014), während es in anderen Studien nur im Mittelfeld liegt (EK 2013a; ITU 2013). Ursache dafür sind unter anderem die verwendeten Statistiken und Breitbanddefinitionen. Doch auch jenseits der Frage, wie viele Haushalte eines Landes mit welcher Bandbreite erreicht werden, verdeutlicht ein genauerer Blick in die Daten größere Herausforderungen. Insbesondere in ländlichen Gegenden ist der Ausbau deutlich weniger fortgeschritten als in stark urbanisierten Regionen. Dieses Problem lässt sich auch in Deutschland beobachten (EK 2013a, S. 24 f.). Dadurch ergibt sich ein starkes Ungleichgewicht, insbesondere bei der Nutzung der Potenziale von Cloud Computing für Privatpersonen und auch KMU im ländlichen Raum.

Neben diesem regionalen Ungleichgewicht ist jedoch die verwendete Technologie für den Ausbau ebenfalls ein kritischer Faktor. Zwar ist der Anteil an Next Generation Access (NGA), welcher alle Technologien umfasst, die einen Durchsatz von 30 Mbit/s und mehr ermöglichen, in Europa im Jahr 2012 weiter gewachsen und liegt in Deutschland bei rund 66 %. Dies ist ungefähr 10 % höher als der EU-Durchschnitt; führend sind jedoch vor allem die nordischen Länder, die einen noch deutlich höheren Anteil aufweisen. Auffällig ist, dass der Anteil von FTTB/H (Fiber to the Building/Home) in Deutschland nur bei rund 2,5 % liegt, deutlich unter dem EU-Durchschnitt von 12 % und weit hinter führenden Ländern wie Schweden (46 %), Dänemark (43 %) oder Finnland (35 %) (EK 2013a, S. 21 f. u. S. 55). Dies ist insofern interessant, da die anderen Anschlussarten, die unter NGA fallen (z. B. VDSL 2), bisher eine Grenze bei 100 Mbit/s oder weniger aufweisen. Wegen der stark ansteigenden Nachfrage durch Videodienste und vielem anderen mehr ist es offen, ob diese Bandbreite angesichts des generellen Trends für eine weitere verstärkte Nutzung von Cloud Computing und Big Data in der Zukunft ausreicht (z. B. Börnsen 2012). Dementsprechend ergibt sich die Herausforderung, dass die derzeitige Bandbreite für eine datenintensive Nutzung von Cloud Computing, vor allem durch die zusätzlichen Anforderungen durch Big Data, in der Zukunft unzureichend ist. Folglich sollte mit Blick

auf die eingesetzten Anschlusstechnologien die existierende Breitbandstrategie für den weiteren Ausbau überdacht werden. Vor allem der Ausbau von FTTB/H-Technologien dürfte hierbei eine Schlüsselrolle spielen, jedoch benötigt dies hohe Investitionen. Die Telekommunikationsanbieter haben diese Kosten bisher gescheut und nutzen wahrscheinlich aus diesem Grund so stark wie möglich die existierende Infrastruktur, wie z. B. DSL und ISDN.

Daher ist bei einer Neuregelung des Ausbaus insbesondere die Frage der Finanzierung beziehungsweise der Aufteilung der Kosten zu beachten. Falls dies rein aufseiten der Anbieter geschieht, kann sich dies in höheren Preisen für die Nutzer selbst oder in Form von Abgaben durch die Serviceanbieter niederschlagen. Beides ist problematisch; im ersten Fall kann dies zu einer Aufteilung der Gesellschaft in digitale Bürger erster und zweiter Klasse führen, und damit wahrscheinlich zu einer geringeren Nutzung von Cloud Computing. Im zweiten Fall kann dies zu Auswirkungen auf die Wettbewerbsfähigkeit insbesondere neuer und kleiner Anbieter führen, die nicht in der Lage sind, zusätzliche Zahlungen für bessere Übermittlungsqualität aufzubringen. Diese auch unter dem Begriff der Netzneutralität geführten Diskussionen sind jedoch teilweise sehr kontrovers (z. B. EFI 2011; Heng 2011), und zugleich fehlen noch aussagekräftige empirische Daten, welche die unterschiedlichen Ansichten belegen könnten. Aus diesem Grund lassen sich die Folgen solcher Finanzierungsmodelle nicht abschließend beurteilen. Dennoch unterstreichen diese Problemstellungen die Notwendigkeit, dass neue Strategien für den Breitbandausbau auch die Frage der Finanzierung berücksichtigen müssen. Hierzu ist es sinnvoll, auch die Erfahrungen anderer Länder in Europa (z. B. Schweden) oder außerhalb Europas (Südkorea) zu studieren und diese Ergebnisse in eine Revision der Breitbandstrategie einfließen zu lassen (TAB 2012, S. 51 ff.).

DIGITALE WETTBEWERBSFÄHIGKEIT

5.3

Unter digitaler Wettbewerbsfähigkeit wird hier die Fähigkeit eines Standortes (Region oder Land) verstanden, in neu entstehenden Märkten – wie beispielsweise Cloud Computing oder Big Data – relevante, globale Akteure hervorzubringen. Insbesondere in Europa und speziell in Deutschland wird dies immer wieder unter der Fragestellung diskutiert, warum es kein deutsches Google, Facebook etc. gibt. Tatsächlich zeigt neuere wissenschaftliche Forschung, dass es in Europa einen Mangel an innovativen, schnell wachsenden Unternehmen auch im Hochtechnologiebereich gibt. Dies wird auch als Überalterung von europäischen Firmen bezeichnet, d. h., die Mehrzahl der europäischen Firmen ist im Durchschnitt älter als Firmen in den USA. Das wird auch als einer der Gründe für das langsamere Produktivitätswachstum von Europa im Vergleich zu den USA betrachtet (Ark et al. 2003; Philippon/Véron 2008).



Grundsätzlich lässt sich diese Herausforderung aber nicht auf ein einzelnes Problem zurückführen, sondern vielmehr spielen hier mehrere Faktoren, die sich teilweise gegenseitig beeinflussen, eine Rolle. Ein erster Faktor ist die fehlende Marktgröße, welche die Wachstumsmöglichkeiten von Firmen beschränkt. Dies führt dazu, dass innovative Unternehmen wie beispielsweise StudiVZ zwar in Deutschland zeitweise Marktführer sein können, aber auf lange Sicht meistens den global auftretenden Konkurrenten unterliegen, da ihnen die Internationalisierung schwerer fällt. Ursache ist, dass die Märkte für Software und Internetdienste sehr stark von Netzwerkeffekten geprägt sind. Dies bedeutet, dass sich der Nutzen dadurch erhöht, dass möglichst viele dasselbe Produkt oder dieselbe Dienstleistung erwerben bzw. in Anspruch nehmen. Eine wesentliche Konsequenz ist, dass Firmen in möglichst vielen und großen Märkten vertreten sein müssen, um diese Effekte zu erzielen. Da aber der deutsche Markt im Vergleich zu den USA relativ klein ist, fällt dies den deutschen Unternehmen schwerer. Sie müssen schneller den internationalen Markt anvisieren, was gerade in Europa durch unterschiedliche rechtliche Rahmenbedingungen wie beispielsweise Datenschutz, Steuerrecht etc. erschwert wird. Hinzu kommen weitere Faktoren wie Sprachbarrieren oder kulturelle Unterschiede. Dieses als Marktfragmentierung bezeichnete Problem, das in der Literatur schon lange als ein wesentlicher Nachteil europäischer IT-Unternehmen diskutiert wird (Mowery 1996; Steinmueller 2004), ist Gegenstand zahlreicher Maßnahmen durch die Europäische Union. So ist der Single Digital Market in fast allen Strategien der Europäischen Kommission ein zentraler Bestandteil. Dementsprechend sind hier auch viele Initiativen wie die Richtlinie zum E-Commerce oder die Etablierung eines einheitlichen Zahlungsraumes (Single Payment Area, SEPA) verortet. Weitere geplante Schritte sind die Vereinheitlichung des Datenschutzes durch die Datenschutzgrundverordnung oder die Harmonisierungen in anderen Bereichen wie Verbraucherschutz. Doch trotz der zahllosen Initiativen stellt die Internationalisierung in Europa insbesondere neugegründete Unternehmen sowie KMU noch vor viele Probleme, da sie in der Regel nicht über die erforderlichen Strukturen und Ressourcen wie eine Rechtsabteilung verfügen, um zum Beispiel unterschiedliche Zulassungsbedingungen oder Umsatzsteuerregelungen zu meistern. Zwar wird argumentiert, dass dies für amerikanische Unternehmen kein Problem darstellt. Dies trifft teilweise zu, verkennt aber, dass beispielsweise kleine Unternehmen, die nur von den USA aus operieren und keine Niederlassungen in Europa unterhalten, nur dem amerikanischen Rechtsrahmen unterliegen und Fragen wie Umsatzsteuerverrechnung in Europa für sie nur eine geringe Rolle spielen. Falls aber amerikanische Firmen in Europa Niederlassungen gründen, haben sie meistens schon auf dem amerikanischen Markt eine gewisse Größe erlangt, die es ihnen ermöglicht, die notwendigen Strukturen und Ressourcen für die Bewältigung und teilweise Ausnutzung der unterschiedlichen Rahmenbedingungen in Europa aufzubauen.

All dies unterstreicht, dass eine weitere Harmonisierung in Europa notwendig ist. Dies geht auch deutlich aus der Ankündigung des Europäischen Rates vom Oktober 2013 hervor, die einen einheitlichen europäischen Markt für digitale Produkte und Dienstleistungen bis 2015 ankündigt. Inwieweit eine solche Harmonisierung aber möglich ist, bleibt offen, und es gilt aus deutscher Sicht, diesen Prozess weiterhin zu begleiten und zu unterstützen. Abschließend sei noch angemerkt, dass eine gewisse Diversität nicht nur ein Problem darstellt, sondern auch Chancen bietet. Ein Beispiel dafür ist der Erfolg von Skype, das ursprünglich erfunden wurde, um die großen Preisdifferenzen im europäischen Telekommunikationsmarkt zu umgehen.

Ein weiterer wichtiger Faktor ist das geringe Niveau der unternehmerischen Aktivitäten in Europa und insbesondere in Deutschland im Vergleich zu den USA und anderen Weltregionen (Aumasson et al. 2010, S.184 f.; Leimbach/Wydra 2012, S.6 ff.). Jedoch haben Studien gezeigt, dass diese Zahlen über die Zeit stark variieren (Veugelers/Cincera 2010). Ein Beispiel dafür ist gerade die deutsche IT-Branche, die in den letzten Jahren sehr viel Aufmerksamkeit für ihren »Gründungsboom« erhielt. So stiegen die Zahlen zeitweise an, aber zum einen blieben sie dennoch unter dem Niveau der anderen Länder und zum anderen finden sich schon wieder Anzeichen einer Abkühlung (Leimbach/Wydra 2012, S.6 ff.; Müller et al. 2012). Neuere Studien auf europäischer Ebene identifizieren folgende Ursachen für die fehlenden Aktivitäten und den daraus resultierenden Mangel an schnell wachsenden Unternehmen in der Software- und IT-Dienstleistungsbranche: die mangelnde Kommunikation zwischen den Akteuren des Innovationssystems, insbesondere zwischen der Wirtschaft und der Forschung; dass der Staat nicht als Mittler zwischen diesen Akteuren handelt; der fehlende Wettbewerb zwischen jungen und alten Firmen sowie der Mangel an Wagnis- und Wachstumskapital (Veugelers et al. 2012, S.9 ff.). Zusammen genommen führen diese Gründe dazu, dass junge, innovative und erfolversprechende Unternehmen entweder scheitern, Wachstumsbeschränkungen erreichen oder von bestehenden Unternehmen übernommen werden, jedoch keine weitere Wachstumspotenziale ausschöpfen können.

Die mangelnde Kommunikation zwischen den Akteuren sowie die Rolle des Staates zeigen sich in unterschiedlichen Punkten. Ein Indikator ist das geringe Niveau von FuE-Ausgaben der Unternehmen, wo sich bei Software- und Internettechnologien den USA gegenüber ein deutliches Defizit zeigt (Turlea et al. 2010, S.75; Turlea et al. 2011, S.55). Ein Grund dafür ist gerade die geringe Größe der Unternehmen in Deutschland und Europa, die kaum eigene FuE-Aktivitäten erlaubt. Zudem gibt es kaum Verbindungen zu Universitäten und anderen Forschungseinrichtungen. Zwar wird dies durch Förderprogramme der verschiedenen Ministerien adressiert, aber eine deutlich größere Hebelwirkung könnte der Staat auf anderen Wegen erzielen. Wesentlich wären dabei die staat-



lichen Beschaffungen, die in Deutschland und anderen EU-Mitgliedstaaten ungefähr 20 % des Marktvolumens ausmachen (Aumasson et al. 2010, S.231 ff.). Dieses Potenzial könnte eingesetzt werden, um gewünschte Entwicklungen zu verstärken. Aktuelle Beispiele aus den USA ist die »Cloud first policy«, die vorsieht, dass staatliche Einrichtungen bei der Beschaffung zuerst eine Lösung auf Cloud-Computing-Basis überprüfen müssen, oder das Big-Data-Programm, welches darauf ausgerichtet ist, dass staatliche Einrichtungen Big-Data-Lösungen kaufen können, um so Markimpulse zu setzen. Gerade letzteres verweist auf einen weiteren Hebel, der in den USA verstärkt eingesetzt wird, nämlich die vor-kommerzielle oder innovative Beschaffung. Diese schafft einerseits weitere Verbindungen zwischen Unternehmen und Forschung und bietet andererseits die Möglichkeit, insbesondere neugegründeten und kleinen Unternehmen die Probleme der Startphase zwischen Prototyp und Produkt (valley of death) zu überbrücken (Wessner 2009). Ein klassisches Beispiel ist das Programm »Small Business Innovation Research« (SBIR), bei dem staatliche Einrichtungen innovative Lösungen bei kleinen Unternehmen beschaffen können. Die Unternehmen werden dabei nicht nur finanziell, sondern durch die Vermittlung zu Venture-Capital-Gebern unterstützt (Wessner 2009). Ansätze zur innovativen Beschaffung befinden sich in Deutschland aber erst im Anfangsstadium. Auf europäischer Ebene sind mit der European Cloud Plattform und den verbundenen Projekten (Cloud for Europe) erste Maßnahmen im Rahmen des neuen Horizont-2020-Forschungsprogramms implementiert worden.

Die andere große Herausforderung ist der Mangel an Finanzkapital, welches für die Gründung und das Wachstum von Firmen notwendig ist. Hierbei werden zwei Aspekte genannt: die eingeschränkten Möglichkeiten zur externen Finanzierung durch Banken sowie der allgemeine Mangel an Venture Capital. Durch die Finanzkrise stand zeitweise der erste Aspekt im Vordergrund, jedoch ist der allgemeine Mangel an Venture Capital (VC) ein altbekanntes Problem. Sowohl in absoluten wie auch in relativen Zahlen (per Arbeitnehmer) sind die VC-Investitionen in Europa und insbesondere in Deutschland in der Software- und IT-Dienstleistungsbranche niedriger als in den USA (Schleife et al. 2012, S. 32 f.). Die dafür genannten Gründe sind sehr unterschiedlich. Genannt werden unter anderem die gesetzlichen Regulierungen, die Größe des VC-Marktes oder auch ein Mangel an attraktiven Investitionsmöglichkeiten (Fransman 2011). Dass der Mangel tatsächlich einen negativen Einfluss auf die Wettbewerbs- und Leistungsfähigkeit des europäischen IT-Sektors hat, legen neuere Studienergebnisse nahe (Veugelers et al. 2012, S.25 ff.). Sie zeigen auch, dass sich dieser Mangel nicht nur in der Gründungsphase, sondern gerade auch in der Wachstumsphase von schnell wachsenden Unternehmen negativ auswirkt (Veugelers/Cincera 2010). Auf nationaler und europäischer Ebene gab es in den letzten Jahren eine Vielzahl von Initiativen, die dieses Problem zu adressieren versuchen. Bisher ist aber noch nicht ersichtlich, wie erfolgreich und nachhaltig diese sind.

All diese Probleme sind nicht neu und nicht unbekannt. Es gab bereits in der Vergangenheit eine Vielzahl von Initiativen, an der Situation selbst hat sich dabei bis jetzt nicht viel geändert. Dies unterstreicht, dass nicht nur regulative oder finanzielle Aspekte, sondern auch andere Faktoren wie kulturelle Besonderheiten eine Rolle spielen (Fransman 2011). Dass sich dieses Problem trotzdem erfolgreich angehen lässt, zeigt das Beispiel Israel. Dort sind seit den 1960er Jahren verschiedene Programme entwickelt worden, um die genannten Punkte wie Verbindung von Forschung und Unternehmen, VC-Kapital oder innovative Beschaffung anzugehen. Dies hat dazu geführt, dass Israel heute insbesondere im Bereich IT ein erfolgreiches, wenn auch hoch spezialisiertes Land ist (Breznitz 2006, 2011). Dies bedeutet jedoch nicht, dass solche Maßnahmen ohne Weiteres zu kopieren sind. Vielmehr zeigen die hier diskutierten Punkte, dass vor allem ein ganzheitlicher und koordinierter Ansatz über Politikfelder hinweg, der sowohl den Lebenszyklus eines Unternehmens als auch die komplette Wertschöpfungskette in Betracht zieht, diese Probleme adressieren kann.

INTERNETGOVERNANCE

5.4

Die Enthüllungen verschiedener geheimdienstlicher Praktiken (Snowden-Enthüllungen) in Verbindung mit technologischen Entwicklungen wie Cloud Computing und Big Data haben dazu geführt, dass die Debatte über die Steuerung (Governance) eines globalisierten Internets wieder stark an Bedeutung gewonnen hat. Ein wesentlicher Grund ist, dass das Vertrauen in Datenschutz und Datensicherheit für die Akzeptanz von Cloud Computing und Big Data durch Firmen, staatliche Einrichtungen sowie einzelne Bürger ein zentraler Aspekt ist (Bradshaw et al. 2012; Catteddu 2011; Catteddu/Hogben 2009; KPMG 2013a u. 2013b; Tene/Polonetsky 2013), welches durch die enthüllten Eingriffe staatlicher Stellen beeinträchtigt wurde. Folglich ergeben sich daraus Fragen an die bisherigen Verwaltungsstrukturen des Internets.

Bisher wurde das Internet von verschiedenen Institutionen, Regeln und technologischen Bedingungen getrieben, die aus historisch gewachsenen Gründen von den USA dominiert werden (z. B. Hafner/Lyon 1998). Selbst die Reformen zu Beginn des Jahrtausends, wie die Etablierung von ICANN und IGF im Rahmen der Diskussion um die Informationsgesellschaft, haben daran nicht viel geändert. Die Governance der technischen Struktur des Internets, z. B. über das IP-Adresssystem, das Domain Name System (DNS) oder die Weiterentwicklung des TCP/IP-Protokolls, werden vor allem von Organisationen wahrgenommen, die in den USA angesiedelt sind oder in denen die USA einen wesentlichen Einfluss haben (DeNardis 2009). Dies erstreckt sich auch auf die Technologien, welche die Sicherheit und Übermittlung von Informationen betreffen. Diese wurden und werden noch heute oft in Kooperationen zwischen amerikanischen Firmen sowie



amerikanischen Behörden wie der DARPA oder NIST entwickelt. Die resultierenden Probleme wie undurchsichtige Interessenlagen und Beeinflussungen verdeutlicht das Beispiel des durch die NSA kompromittierten Standards für Generierung von Zufallszahlen (Ermert 2013). Darüber hinaus ist sowohl ein Großteil der Server- und Netzwerkinfrastrukturen inklusive der wesentlichen Datenverbindungen, welche die Grundlagen für das Internet bilden, in den USA beheimatet beziehungsweise verlaufen über die USA und sind daher von der Gesetzgebung der USA abhängig (Bowden 2013).

Bis zu den Snowden-Enthüllungen wurde nur selten Kritik an der Dominanz der USA bei der Internetgovernance geäußert. Dabei standen dann vor allem wirtschaftliche Aspekte und weniger Sicherheitsaspekte im Vordergrund. Letztere wurden allenfalls außerhalb der westlichen Hemisphäre thematisiert, meist im Zusammenhang mit der Kontrolle über Inhalte und damit verbundene Meinungsbildungsprozesse. In Europa standen vor allem ökonomische Aspekte im Vordergrund der Diskussion, wie insbesondere die Diskussion um das Fehlen von globalen, europäischen Software- und Internetfirmen verdeutlicht (Kap. VI.5.3). In Bezug auf die Datensicherheit sah sich Europa nach der Datenschutzrichtlinie von 1995 in einer Sonderrolle, da diese einen höheren Datenschutzstandard als in anderen Ländern/Regionen der Welt bot. Zusätzlich wurde mit den USA das Safe-Harbor-Abkommen getroffen, welches die Geltung des strengeren europäischen Datenschutzes auch für europäische Daten in den USA garantieren sollte. Firmen mit Sitz in den USA, die sich diesem Abkommen anschlossen, sollten, vom US-amerikanischen Handelsministerium überwacht, den gleichen Datenschutz für Daten aus Europa garantieren. Zwar geriet das Abkommen immer wieder in die Kritik, da die Umsetzung und Kontrolle als schwach bewertet wurden. Insbesondere der Umgang amerikanischer Internetkonzerne wie Facebook oder Google führten zu Diskussionen, aber dies stellte die Fortführung des Abkommens nicht ernsthaft infrage (z. B. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein 2010). Erst die Snowden-Enthüllungen, die unter anderem zeigten, wie die Daten dieser Firmen von der NSA und anderen Geheimdiensten genutzt wurden, haben die resultierenden Probleme und Herausforderungen so offensichtlich gemacht, dass sie nun zum Gegenstand einer ernsthaften öffentlichen Diskussion geworden sind.

Einen ersten Schritt, diese Probleme zu adressieren, stellt die neue Datenschutzgrundverordnung dar, die zurzeit verhandelt wird und eine Vielzahl von datenschutzrechtlichen Aspekten für Cloud Computing durch neue Instrumente und eine europäische Vereinheitlichung des Rechtsrahmens adressiert (Kap. VI.3). Diese stärken vor allem die Rechte gegenüber privatwirtschaftlichen Organisationen. Die Snowden-Enthüllungen haben jedoch verdeutlicht, dass sich die weiterführenden Probleme in Europa längerfristig nicht allein lösen lassen. Denn die enthüllten Tatsachen haben das Vertrauen in Regelungen wie das Safe-Harbour-

Abkommen erschüttert und grundsätzlich die Fähigkeit der EU und ihrer Mitgliedstaaten, die Daten ihrer Bürger vor ausländischen Firmen und Behörden zu schützen, infrage gestellt (Bowden 2013). Eine Konsequenz ist, dass Themen wie Internetgovernance oder Data Sovereignty, also die Kontrolle über Daten, inzwischen auch in Europa und insbesondere Deutschland diskutiert werden. Hierbei gilt jedoch zu beachten, dass, obwohl die USA und ihre Partner am stärksten die Internetüberwachung betreiben und der gesetzliche Datenschutz in der EU und insbesondere in Deutschland deutlich stärker ausgeprägt ist, auch dies nicht die völlige Sicherheit der Daten garantieren kann. Grund dafür sind vor allem die Geheimdienste der europäischen Mitgliedstaaten, die ähnliche Operationen durchführen oder gar mit den USA eng zusammenarbeiten, wie das Beispiel Großbritannien zeigt. Die eingesetzten Spionageanwendungen selbst sind dabei meist technologisch auf geringerem Kenntnisstand, und es stehen in der Regel weniger Ressourcen zur Verfügung, jedoch ist die Gesetzeslage meist sehr ähnlich, wenn nicht gar vergleichbar mit den USA (Heumann/Scott 2013; Maxwell/Wolf 2012).

Eine erste Reaktion auf die Abhängigkeit von der amerikanisch-geprägten Infrastruktur für das Internet ist die Wiederbelebung des Strebens nach der Internationalisierung sowohl der Governancestrukturen als auch der technischen Infrastrukturen für das Internet, welche auf den ursprünglichen Idealen zur Offenheit und Transparenz des Internets beruhen. Aus diesem Grund haben sich führende Vertreter der wichtigsten Organisationen wie beispielsweise ICANN, IANA oder IETF, die verantwortlich für die Koordination des Internets und seiner Infrastrukturen sind und deren Entwicklungsgemeinschaft repräsentieren, sich für diese Internationalisierung ausgesprochen und im gleichen Atemzug Bedenken ob der Überwachungsaktivitäten geäußert (ICANN 2013). Ebenfalls hat die Internet Engineering Task Force (IETF) auf einem ihrer letzten Treffen damit begonnen, Protokolle und Best-Practice-Ansätze zu überarbeiten und zu verbessern.¹⁰ Eine erste politische Konsequenz dieser Entwicklung ist, dass die Obama-Regierung vor Kurzem beschlossen hat, die Aufsicht über die Internet Assigned Numbers Authority (IANA), welche als Verwalter aller IP-Adressen und Domainnamen eine zentrale Organisation der Internetgovernance ist, abzugeben (Ermert 2014).

Jedoch zeigen die verschiedenen Diskussionen um die Ausgestaltung dieser neuen Verwaltungsstruktur auch, dass dies nicht zwangsläufig zu einer Verbesserung im Sinne von Offenheit und Transparenz führen muss. So versuchen unterschiedliche Staaten aus unterschiedlichen Gründen, Einfluss auf die Steuerung zu nehmen. Die Motive sind dabei sehr unterschiedlich und reichen von Kontrolle über Inhalte bis zu ökonomischen Interessen. Zusammen mit Tendenzen wie der Abspaltung nationaler Netze, Diskussionen um territorial begrenztes Routing (z.B. Schengen-Routing), Sperren und Zensuren zeigt sich deutlich, dass im

10 www.ietf.org/proceedings/88/technical-plenary.htm



VI. HERAUSFORDERUNGEN

Moment verschiedenste Vorstellungen um die Deutungshoheit konkurrieren. Eine Konsequenz könnte sein, dass es das Internet, wie es ursprünglich konzipiert wurde und noch heute in den Idealen der Entwicklergemeinschaft fortlebt, in Zukunft so nicht mehr geben wird (z.B. Gaycken 2014). Ein Beispiel dafür ist, dass weitere Bemühungen um die internationale Harmonisierung von Datenschutz oder Netzsicherheit, wie sie beispielsweise von Brasilien und Deutschland betrieben werden, bisher nicht auf ungeteilte Zustimmung stoßen. Auch in dieser Frage stehen vielfältige strategische Erwägungen der einzelnen Nationen im Vordergrund.

RESÜMEE:

HANDLUNGS- UND FORSCHUNGSFELDER

VII.

Die Phänomene Big Data und Cloud Computing unterscheiden sich in vielerlei Hinsicht. Während der Begriff Big Data als Synonym für die effiziente Analyse besonders großer Datenmengen verwendet wird, beschreibt Cloud Computing einen Ansatz zur bedarfsgerechten Bereitstellung von IT-Ressourcen. Auch wenn Big-Data-Analysen ohne Cloud Computing durchgeführt werden können und Cloud Computing neben Big-Data-Analysen zahlreiche andere Anwendungsfelder hat, wird der Kombination von Big Data und Cloud Computing großes Potenzial zugeschrieben. Am augenscheinlichsten ist, dass Unternehmen, staatliche Stellen und Behörden, wissenschaftliche Einrichtungen wie auch Privatpersonen durch Cloud Computing die Möglichkeit bekommen, große Datenmengen zu analysieren, ohne vorher eine entsprechende IT-Infrastruktur aufbauen zu müssen. Organisationen können sich bei Big Data in der Cloud grob zwischen IaaS- und SaaS-Angeboten entscheiden. Bei der Nutzung von IaaS-Angeboten stellt der Cloudanbieter nur die Infrastruktur zur Verfügung, und der Nutzer betreibt die zur Analyse großer Datenmengen nötige Software selbst. Bei der Nutzung von SaaS-Angeboten wird die gesamte Big-Data-Lösung vom Cloudanbieter bereitgestellt. Für Privatpersonen kommen in der Regel nur SaaS-Angebote infrage.

Im Hinblick auf alle Bedarfsfelder der Hightech-Strategie der deutschen Bundesregierung können Anwendungsbeispiele genannt werden, in denen Big Data entweder bereits heute eine große Rolle spielt oder zukünftig eine große Rolle spielen wird. Beispielsweise kann Big Data nicht nur die Vorhersage von Wetterereignissen oder Straftaten verbessern, sondern auch das Zusammenspiel von verschiedenen Akteuren im Straßenverkehr oder in der industriellen Fertigung optimieren. In hochkomplexen Kontexten, wie sie z. B. aus dem Gesundheitswesen bekannt sind, kann Big Data entscheidungsunterstützend wirken. Während Big Data ganz allgemein Potenzial zur Steigerung von Kundenorientierung, Produktivität und Innovationsfähigkeit sowie zum Aufbau von strategischen Wettbewerbsvorteilen zugeschrieben wird, verspricht man sich von cloudbasierten Umsetzungen zusätzlich niedrigere Kosten, höhere Flexibilität und weniger Bedarf an der Beschäftigung eigener, einschlägig qualifizierter IT-Fachkräfte. Auch Privatpersonen können, z. B. im Zusammenhang mit Citizen Science, von Big Data profitieren. Erwartet wird, dass Big Data und Cloud Computing zu Produktivitätssteigerungen, Innovationen, zusätzlichen Unternehmensgründungen sowie zu mehr Wachstum und Beschäftigung führen.

Mithilfe von Big Data in der Cloud sollen die Vorteile beider Entwicklungen kombiniert werden. Während sich bei einigen Big-Data-Anwendungen die Nutzung von Clouddiensten anbietet, ist das bei anderen nicht der Fall. Ein wesent-

licher Punkt ist, dass es sich insbesondere bei großen Datenmengen empfiehlt, die Daten dort zu analysieren, wo sie anfallen, weil der Transport ein massives technisches Problem darstellt. Dementsprechend wäre die cloudbasierte Analyse großer Datenmengen insbesondere dann sinnvoll, wenn die Daten im Netz entstehen oder dort bereits aggregiert wurden. Darüber hinaus bietet sich die Nutzung von Clouddiensten vor allem dann an, wenn im Vorfeld eines Big-Data-Vorhabens vorübergehend mit vorhandenen Daten experimentiert werden soll oder wenn bei einem Analysevorhaben mit vorübergehenden Anforderungsspitzen zu rechnen ist.

Derzeit besteht in den meisten Anwendungsbereichen kein Konsens darüber, wie weit die Analyse von Daten und ein datenbasiertes Entscheiden gehen sollen. Zahlreiche Anwendungsmöglichkeiten versprechen sowohl Nutzen als auch Gefahren. Während z. B. die Prävention von Verbrechen durch erhöhte Polizeipräsenz an bestimmten Punkten häufig als sinnvoll betrachtet wird, wird – im Hinblick auf die Möglichkeit auf vorzeitige Entlassung aus dem Vollzug – die Berechnung der Rückfallwahrscheinlichkeit einzelner Straftäter auf Grundlage von Daten über andere Straftäter eher kritisch gesehen. Grundsätzliche Bedenken bestehen insbesondere bezüglich der zunehmenden Ökonomisierung von Privatheit, also der Bereitschaft von Privatpersonen, für eine Gegenleistung auf Privatheit zu verzichten.

Sowohl die Durchführung von Big-Data-Analysen als auch die Nutzung von Cloud Computing sind mit zahlreichen Herausforderungen verbunden. Neben technischen und betrieblichen Herausforderungen spielen rechtliche, gesellschaftliche und strukturelle Herausforderungen eine Rolle. Durch die Kombination von Big Data und Cloud Computing verstärken sich manche Herausforderungen, teils werden sie abgemildert. Die zentralen technischen Herausforderungen von Big Data betreffen den Aufbau der Analysearchitektur und den Umgang mit dem Datenmaterial. Einerseits ist davon auszugehen, dass die Nutzung cloudbasierter Lösungen den Aufbau der Analysearchitektur aus Sicht des Nutzers vereinfacht, da der Cloudanbieter für die Bereitstellung eines wesentlichen Teils der Infrastruktur verantwortlich ist. Andererseits kommen neue Herausforderungen mit Bezug auf die Interoperabilität von Diensten hinzu, da die genutzten Cloudlösungen untereinander und mit Inhouselösungen integriert werden müssen. Die betrieblichen Herausforderungen betreffen vor allem die Bestimmung der Wirtschaftlichkeit der Durchführung von Big-Data-Analysen und die eingeschränkte Verfügbarkeit von Experten mit einschlägigem Fachwissen, aber auch Widerstände bei Planung und Durchführung. Die Nutzung von Clouddiensten führt dazu, dass Nutzer deutlich weniger Kontrolle über die analysierten Daten und die Analyseergebnisse haben. Gleichzeitig kann angenommen werden, dass sich die Nutzung von Clouddiensten überwiegend positiv auf die Wirtschaftlichkeit auswirkt. Hier unterscheiden sich allerdings cloudbasierte Big-Data-Anwendungen nicht wesentlich von anderen Anwendungen, die in die Cloud verlagert wer-



den. Im Hinblick auf rechtliche Herausforderungen zeigt sich, dass sowohl bei Big Data als auch bei Cloud Computing hinsichtlich zahlreicher Fragen Rechtsunsicherheit herrscht bzw. sich das anwendbare Recht als wenig passgenau erweist und einer Anpassung an den Stand der Technik bedarf. Betroffen sind vor allem die Bereiche Datenschutz, Eigentum an Daten, Vertragsgestaltung, Haftung und Wettbewerb. Die gesellschaftlichen Herausforderungen durch Big Data sind vielfältig und betreffen Bereiche wie Vertrauen und Verlässlichkeit, die Verengung des Sichtbaren, Autonomie und Beeinflussung, Diskriminierung und Teilhabe, aber auch Vulnerabilität und Abhängigkeit. Die Nutzung von cloud-basierten Lösungen hat dagegen kaum größere gesellschaftliche Auswirkungen. Die Entwicklung einer ausreichenden Fachkräftebasis stellt die zentrale strukturelle Herausforderung im Kontext von Big Data dar. Durch die Nutzung von Clouddiensten kommen Herausforderungen in den Bereichen Infrastruktur-/Breitbandversorgung und Internetgovernance dazu. Die Sicherstellung der digitalen Wettbewerbsfähigkeit ist sowohl im Hinblick auf Big Data als auch im Hinblick auf Cloud Computing eine zentrale Herausforderung.

Großunternehmen sowie auf Big Data bzw. Cloud Computing ausgerichtete Start-ups kommen mit den Herausforderungen in der Regel am besten zurecht. Sie haben gegenüber mittelständischen Unternehmen meist strukturelle Vorteile, wie z. B. die für den wirtschaftlichen Betrieb einer Private Cloud notwendige Größe sowie besseren Zugriff auf die erforderliche Expertise. Big Data in der Cloud kann sowohl auf Basis von Public-Cloud-Angeboten als auch auf Basis von Private-Cloud-Angeboten realisiert werden, deren Aufbau sich jedoch meist nur für Großunternehmen rechnet. Die Nutzung einer Public Cloud erlaubt KMU oft erst die Ausschöpfung des Potenzials von Big Data in der Cloud, erfordert aber die Bewältigung zusätzlicher Herausforderungen im Bereich Informationssicherheit. Unter anderem aufgrund der Tatsache, dass Public-Cloud-Angebote überwiegend von ausländischen Anbietern angeboten werden, ist die Rechtslage teils schwer einschätzbar.

Insgesamt stammen sowohl die starken Anbieter im Bereich Cloud Computing als auch die Anbieter der zentralen Technologien im Bereich Big Data überwiegend aus den USA. Vor allem im Big-Data-Umfeld fällt auf, dass einerseits die angebotenen Lösungen noch nicht in jeder Hinsicht ausgereift sind und dass andererseits Open-Source-Lösungen eine große Bedeutung haben. Im Hinblick auf die mangelnde Reife der angebotenen Lösungen wird vor allem kritisiert, dass weiterführende, fortschrittliche Analysen, wie sie z. B. für Cluster- oder Netzwerkanalysen erforderlich sind, nur sehr eingeschränkt möglich sind und dass die Nutzung der meisten Lösungen nur mit umfassendem Fachwissen möglich ist.

Experten in den Bereichen Big Data und Big Data in der Cloud sind allerdings auf dem Arbeitsmarkt kaum vorhanden. Es wird davon ausgegangen, dass dieser Mangel nicht allein durch das Angebot spezialisierter Lehr- und Studiengänge

bewältigt werden kann; die Anforderungen sind zu vielschichtig. Von Technologieanbietern wird deshalb erwartet, dass sie die Anforderungen an die Nutzer reduzieren. Hilfreich wären in diesem Zusammenhang neben einer benutzerfreundlichen Abfragesprache auch Verbesserungen bei der Darstellung von Analyseergebnissen. Dennoch steigen die Anforderungen an die Informationskompetenz der Nutzer cloudbasierter Big-Data-Dienste.

Im Folgenden werden zentrale Aussagen und wichtige Erkenntnisse dieser Vorstudie noch einmal aufgegriffen. Das Hauptaugenmerk liegt dabei auf sieben Themen mit dazugehörigen Fragestellungen, von denen angenommen wird, dass sie zukünftig im Kontext von Big Data in der Cloud für Wirtschaft, Politik, Forschung und Gesellschaft besonders relevant sein werden. Im Zusammenhang mit diesen Themen und Fragestellungen werden zentrale Handlungs- und Forschungsbedarfe erläutert sowie mögliche Optionen skizziert.

ANWENDUNGSFÄLLE UND EINSATZSZENARIEN

Für Big Data und Big Data in der Cloud sind vielfältige Anwendungsfälle denkbar. Vor allem wissenschaftliche Einrichtungen setzen sich schon lange mit der Analyse sehr großer Datenmengen auseinander. Die Identifikation von lohnenswert erscheinenden Anwendungen stellt in diesem Bereich kein großes Problem dar. Die Schwierigkeiten liegen viel mehr in der Umsetzung. Beispielsweise ist aufgrund von beschränkten Mitteln der Aufbau einer angemessenen Analysearchitektur oft nicht möglich. Big Data erlaubt Unternehmen nicht nur die Unterstützung und Automatisierung von Entscheidungen im Geschäftsalltag, sondern auch die Umsetzung von neuen Geschäftsmodellen und Dienstleistungen (z. B. Echtzeit-Cross- und -Up-Selling in E-Commerce und stationärem Vertrieb). Die Chancen und Risiken von Big Data in der Cloud sind genauso wie die Beschaffenheit geeigneter Einsatzszenarien in vielen Organisationen, insbesondere der öffentlichen Verwaltung in Deutschland und Europa, weitgehend unbekannt – obwohl Big Data in der Cloud eine sinnvolle Ergänzung zu den existierenden Open-Government-Data-Angeboten bieten könnte. Die Nutzung von Cloud-diensten könnte einerseits die Zusammenarbeit zwischen unterschiedlichen Teilen der öffentlichen Verwaltung verbessern und andererseits durch mehr Transparenz und einfachere Teilhabe zu mehr Bürgernähe führen.

Folgende Handlungs- und Forschungsfelder sind zu nennen:

- > Förderung des Erfahrungsaustausches: Damit potenzielle Nutzer die Relevanz von cloudbasierten Lösungen im Hinblick auf bestimmte Big-Data-Vorhaben besser einschätzen können, sollte der Austausch von Erfahrungen im Hinblick auf Chancen und Risiken von Big Data in der Cloud einerseits und Best Practices andererseits gefördert werden. Big-Data-Workshops mit Fokus auf cloudbasierten Lösungen und umfassender Berichterstattung könnten beispielsweise von Branchenverbänden in Kooperation mit Experten aus For-



- schungseinrichtungen oder Unternehmensberatungen durchgeführt werden. Derartige Veranstaltungen wären vor allem für KMU von hoher Relevanz.
- › Einsatz in der öffentlichen Verwaltung: Für Big Data im Allgemeinen und Big Data in der Cloud im Speziellen gibt es nicht nur in der Wissenschaft und in Unternehmen vielversprechende Anwendungsgebiete, sondern vor allem auch in der öffentlichen Verwaltung. Der erfolgreiche Einsatz von Big Data in der öffentlichen Verwaltung könnte zu einer effizienteren Verwaltung und mehr Bürgernähe führen, zur Unterstützung deutscher und europäischer Anbieter beitragen und helfen, die Potenziale von Big Data in der Cloud anderen potenziellen Nutzern zu verdeutlichen. Einerseits sollten die Anwendungspotenziale von Big Data und Big Data in der Cloud im öffentlichen Sektor genauer untersucht werden. Ein zentrales Element entsprechender Aktivitäten könnte ein Vergleich unterschiedlicher nationaler Big-Data-Strategien sein. Andererseits sollten erste verbindliche Ziele im Hinblick auf Angebote im Kontext von Open Government Data im Rahmen der E-Government-Initiative festgeschrieben werden.
 - › Berücksichtigung der Interessen von Privatpersonen: Privatpersonen spielen im Kontext von Big Data eine große Rolle, primär, da sie Organisationen bei der Datengewinnung helfen bzw. deren Analyseergebnisse nützen. Allerdings sind Privatpersonen auch häufig von Big-Data-Analysen betroffen, ohne dass sie es wollen oder darüber ausreichend informiert werden. Personenbezogene Daten spielen im Hinblick auf zahlreiche Anwendungsgebiete von Big Data eine zentrale Rolle. Um die Akzeptanz von Big Data bei Privatpersonen, insbesondere in Zeiten der Massenüberwachung durch Geheimdienste, aufrechtzuerhalten, sollten Maßnahmen getroffen werden, die gewährleisten, dass die Interessen von Privatpersonen berücksichtigt werden. Beispielsweise könnte der Aufgabenbereich von Datenschutzbeauftragten entsprechend erweitert oder ein entsprechender Ombudsrat initiiert werden.

TECHNISCHE INFRASTRUKTUR

Obwohl die für Big Data und Big Data in der Cloud erforderliche Technologie im Prinzip vorhanden ist, gibt es doch bei zahlreichen Anwendungsszenarien Probleme, die aufwendige individuelle Anpassungen erforderlich machen. Einerseits scheinen die vorhandenen Lösungen nicht flexibel genug zu sein, und andererseits werden weiterführende, fortschrittliche Analysen nicht ausreichend unterstützt. Ein zentrales Problem besteht darin, dass existierende Lösungen zur effizienten Verarbeitung großer Datenmengen das Konzept der Iteration nicht unterstützen. Es ist nicht ohne Weiteres möglich, bestimmte Berechnungen wiederholt auf eine Datenmenge anzuwenden, bis ein bestimmter Fixpunkt erreicht ist. Genau das ist aber beispielsweise für Cluster- oder Netzwerkanalysen erforderlich. Darüber hinaus führen unausgereifte Benutzerschnittstellen dazu, dass nur ein sehr eingeschränkter Expertenkreis als Nutzer infrage kommt. Nutzer müssen sich beispielsweise vielfach mit Fragen der Systemprogrammierung aus-

einandersetzen und entscheiden können, wie mit den Daten verfahren werden soll. Insbesondere für KMU stellt somit der aktuelle Stand der Technologie eine große Hürde dar.

Folgende Handlungs- und Forschungsfelder sind zu nennen:

- > Unterstützung von fortschrittlichen Analysen: Um weiterführende, fortschrittliche Analysen zu ermöglichen, ist beispielsweise die native Unterstützung des Konzepts der Iteration zwingend erforderlich. Existierende Lösungen behelfen sich in der Regel mit Maßnahmen, die einerseits zu Einschränkungen bei der Effizienz führen und andererseits spezielle Kenntnisse der Nutzer im Bereich der Programmierung erfordern. Defizite bei der Unterstützung von komplexen Berechnungen auf großen Datenmengen stellen ein zentrales Problem im Kontext von Big Data dar und machen beispielsweise die effiziente Durchführung von Cluster- oder Netzwerkanalysen schwierig. Intensive Forschung in diesem Bereich könnte durch die Ausrichtung von Wettbewerben gefördert werden, bei denen die Leistungsfähigkeit verschiedener Lösungsansätze verglichen wird.
- > Erweiterung des möglichen Nutzerkreises: Die Nutzung von Ergebnissen aus Big-Data-Analysen einerseits und die Durchführung von entsprechenden Analysen andererseits sollten einem möglichst großen Personenkreis möglich sein. Hier sind Anstrengungen vor allem im Bereich der Visualisierung notwendig. Eine Angleichung der Benutzerschnittstellen von Big-Data-Lösungen an jene, die aus dem BI-Umfeld bekannt sind, könnte ein erster Schritt sein. Darüber hinaus scheint eine möglichst standardisierte, deklarative Abfragesprache notwendig. Eine für die vorhandene Analysearchitektur optimale Umsetzung einer Abfrage sollte systemimmanent und autonom vonstattengehen können. Die Entwicklung entsprechender Ansätze sollte in enger Abstimmung mit potenziellen Nutzern erfolgen.
- > Ausbau der Netzwerkinfrastruktur: Der weitere Ausbau der Netzwerkinfrastruktur wird bei der Frage, inwieweit das Potenzial von Big Data in der Cloud genutzt werden kann, eine entscheidende Rolle spielen. Es gilt die aktuellen Strategien zum Netzausbau zu überprüfen, da Deutschland sowohl im Hinblick auf die Verfügbarkeit von FTTB/H als auch im Hinblick auf LTE deutlich unter dem EU-Durchschnitt liegt. Für Anbieter gibt es derzeit wirtschaftlich wenig Anreize, in weniger dicht besiedelten Gebieten zu investieren. Ein Ansatz zur Lösung des Stadt-Land-Gefälles könnte die Verpflichtung der Anbieter zur Bereitstellung einer Mindestbandbreite für alle Kunden durch den Gesetzgeber sein.

ÖKONOMIE DER DATEN

Der Wert von Big Data für Organisationen ist schwer quantifizierbar, unabhängig davon, ob entsprechende Analysen auf Basis von Cloud Computing genutzt werden sollen oder nicht: einerseits, weil die Erfahrungen mit Big Data noch sehr

limitiert sind, und andererseits, weil geeignete Ansätze zur Quantifizierung des Wertes von Daten und Analyseergebnissen fehlen bzw. in der Praxis nicht angewendet werden. Entsprechend schwer haben es Big-Data-Vorhaben, die sich bei beschränkten Mitteln gegen andere mögliche IT-Investitionen durchsetzen müssen. Dies führt dazu, dass Organisationen bei Investitionen in Big Data eher zurückhaltend sind. Die Bestimmung des Wertes von Daten an sich ist sowohl beim Ankauf von Daten für eigene Analysen als auch beim Verkauf von Daten an andere wichtig. Während solche Fragestellungen primär für Unternehmen und andere Organisationen relevant sind, betrifft die Ökonomisierung von Privatheit in erster Linie Privatpersonen. Diese sind zunehmend bereit, für eine Gegenleistung auf Privatheit zu verzichten. Ein Beispiel aus dem Bereich der materiellen Güter sind Community-Marktplätze zur Buchung und Vermietung von privaten Unterkünften. Die starke Nutzung von »kostenlosen« Diensten wie Gmail von Google zeigt, dass das Prinzip auch auf immaterielle Güter wie persönliche Daten übertragbar ist.

Folgende Handlungs- und Forschungsfelder sind zu nennen:

- › Bewertung von Investitionen in Daten und Analysearchitekturen: Mit den sich rasant entwickelnden technischen Möglichkeiten der Datenanalyse steigt der Wert von Daten als Wirtschaftsgut an. Die Nutzung dieses Gutes wird zunehmend als ein wesentlicher Faktor für die Wettbewerbsfähigkeit von Unternehmen und Ländern gesehen. Problematisch ist, dass der Wert von Daten und Analyseergebnissen schwer quantifizierbar ist. Dies erschwert nicht nur den An- und Verkauf von Daten, sondern auch die Berechnung eines ROI für Investitionen in Analysearchitekturen für Big Data. Im Hinblick auf Analysearchitekturen gilt es auch die Kostenseite zu berücksichtigen. Auch wenn Open Source im Kontext von Big Data eine relativ große Rolle spielt, ist die Durchführung von Big-Data-Analysen gleichwohl mit Kosten verbunden. Welche Kosten berücksichtigt werden sollten, wie hoch sie sind und wie sich die Nutzung von Cloudlösungen darauf auswirkt, ist jedoch weitgehend unklar. Die Entwicklung von Ansätzen, die helfen, Investitionen in Daten oder Analysearchitekturen zu bewerten, sollte gezielt gefördert werden.
- › Wiederverwendung von Daten: Die Wiederverwendung von Daten in unterschiedlichen Kontexten ist eine zentrale Komponente im Rahmen eines funktionierenden Marktes für Daten. Sie steigert nicht nur den Wert von Daten für Anbieter, sondern vergrößert auch das Angebot an Daten für potenzielle Nutzer. Allerdings ist die Wiederverwendung von Daten, beispielsweise über Datenmarktplätze, aus verschiedenen Gründen nicht ohne Weiteres möglich. Einerseits fehlen klare Anwendungsfälle und zugrundeliegende Geschäftsmodelle, die sowohl für Anbieter und Nutzer von Daten als auch für Datenvermittler attraktiv sind. Andererseits gibt es eine Reihe ungeklärter Fragen im Hinblick auf rechtliche (z. B. Datenschutz, Urheberrecht), technische (z. B. Datenhaltung, Anonymisierung) sowie ökonomische und gesellschaftliche As-

pekte. Die Erlangung eines umfassenden Verständnisses der Funktionsweise von Märkten für Daten sollte vorangetrieben werden.

- > Folgen der Ökonomisierung von Privatheit: Vieles deutet darauf hin, dass die Ökonomisierung von Privatheit in Zukunft eine noch größere Rolle spielen wird, sowohl im Hinblick auf materielle als auch im Hinblick auf immaterielle Güter wie persönliche Daten. Privatpersonen sind einerseits bereit, bewusst oder unbewusst, für eine Gegenleistung auf Privatheit zu verzichten, andererseits ist die Angst vor einem zunehmenden Verlust von Privatheit gerade in Deutschland groß. Nicht unwahrscheinlich ist, dass der Anteil derer, die unbewusst auf Privatheit verzichten und denen die Folgen ihres Handelns nicht klar sind, groß ist. Diese Vermutung sollte überprüft und, sofern erforderlich, entsprechende Maßnahmen zur Förderung des Bewusstseins gesetzt werden.

WETTBEWERBSFÄHIGKEIT

Die Auswertung umfassender Datenbestände wird in Zukunft immer mehr zu einem Wettbewerbsfaktor werden, sowohl auf nationaler, als auch auf internationaler Ebene. Auf Nutzerseite heutzutage vor allem dann, wenn es um das Verständnis von Kundenverhalten geht. Die entsprechenden internen Daten, vor allem auch im CRM-Bereich, befinden sich immer öfter in einer Cloud. Externe Daten werden ohnehin großteils aus dem Internet bezogen. Aufgrund dessen sind cloudbasierte Big-Data-Lösungen hier besonders relevant. Auf der Anbieterseite stellt die Wettbewerbsfähigkeit auf internationaler Ebene eine zentrale Herausforderung dar. Nur wenn es gelingt, die erforderlichen Rahmenbedingungen zu schaffen, werden sich einschlägig spezialisierte IT-Unternehmen langfristig in Deutschland bzw. der EU niederlassen.

Folgende Handlungs- und Forschungsfelder sind zu nennen:

- > Berücksichtigung der Entwicklung in anderen Ländern und Regionen: Die Wahrscheinlichkeit ist groß, dass die Rolle von Big Data als Wettbewerbsfaktor über die nächsten Jahre weiter zunehmen wird. Damit Organisationen aus Deutschland nicht ins Hintertreffen geraten, sollte die internationale Entwicklung genau beobachtet werden. Relevant sind sowohl Erfolgsgeschichten als auch verschiedene nationale Maßnahmen zur Förderung von Big Data und Cloud Computing. Deutschland sollte allerdings nicht nur versuchen, erfolgreiche Maßnahmen, angepasst an die eigenen Gegebenheiten, umzusetzen, sondern in vielversprechenden Bereichen eine Vorreiterrolle anstreben. In dieser Hinsicht sollten Bereiche identifiziert werden, in denen Deutschland gute Voraussetzungen hat, eine entsprechende Rolle einzunehmen.
- > Gründungs- und Wachstumsmöglichkeiten von neuen IT-Unternehmen: Neben rechtlichen Herausforderungen werden vor allem eine fehlende Unterstützung bei Gründungsvorhaben sowie mangelnde Möglichkeiten der Wachstumsfinanzierung als Gründe für das Fehlen von global agierenden IT-Unternehmen

in Deutschland und Europa genannt. Daher sollte untersucht werden, ob und wie Instrumente der innovativen Beschaffung, steuerliche Anreize oder Vereinfachungen für VC-Investitionen geeignet sind, die Gründungs- und Wachstumsmöglichkeiten von IT-Unternehmen zu fördern.

KOMPETENZ UND FACHKRÄFTE

Die Vielfalt der an Data Scientists gestellten Anforderungen macht es insbesondere für KMU und die öffentliche Verwaltung schwierig, geeignetes Personal zu finden. Von Data Scientists werden nicht nur Kenntnisse in Datenanalysetechnologien verlangt, sondern ebenso in skalierbarem Datenmanagement. Zusätzlich sind Kenntnisse in der jeweiligen Anwendungsdomäne unverzichtbar. Data Scientists werden vor allem im Kontext von Unternehmen und in der öffentlichen Verwaltung benötigt. Über alle Einsatzbereiche hinweg spielt das Thema Informationskompetenz, also der selbstbestimmte, souveräne, verantwortliche und zielgerichtete Umgang mit Informationen, eine zentrale Rolle.

Folgende Handlungs- und Forschungsfelder sind zu nennen:

- › Entwicklung einer benutzerfreundlichen Abfragesprache: Die Auseinandersetzung mit Fragen der Systemprogrammierung ist für Data Scientists zentral. Wichtig wäre ein einheitlicher Ansatz in der Interaktion mit Big-Data-Lösungen auf Basis einer Sprache, in der deklarativ formuliert werden kann, wie mit den Daten zu verfahren ist, ohne zugleich Probleme der Infrastruktur berücksichtigen zu müssen. Eine entsprechende Entwicklung gab es in den 1970er und 1980er Jahren mit SQL bereits im Bereich der relationalen Datenbanken. Ziel sollte sein, dass jeder Businessanalyst grundsätzlich auch mit Big-Data-Lösungen arbeiten kann. Die Entwicklung entsprechender Ansätze sollte in enger Abstimmung mit potenziellen Nutzern erfolgen.
- › Überarbeitung der Curricula für die Ausbildung von Data Scientists: Zusammen mit der Überarbeitung des Anforderungsprofils und der Umsetzung technologischer Verbesserungen sollten die bestehenden Curricula für die Ausbildung von Data Scientists entsprechend überarbeitet werden. Wichtig wäre auch die Einbeziehung ethischer Aspekte. Nicht jede Entscheidung, die datengetrieben getroffen werden kann, sollte auch so getroffen werden.
- › Erhöhung der Informationskompetenz: Neben der Schaffung eines an die Anforderungen des Marktes angepassten Ausbildungsangebots spielt auch die Erhöhung der Informationskompetenz in der Gesellschaft eine wichtige Rolle. Beispielsweise muss der Unterschied zwischen Korrelation und Kausalität genauso bekannt sein, wie es ein Verständnis dafür geben muss, welche Gefahr von der Überanpassung von Modellen ausgeht. Trotz der zahlreichen Potenziale ist ein kritischer Umgang mit Analyseergebnissen unumgänglich. Aber auch Privatpersonen müssen die Folgen einer Verwendung ihrer Daten im Rahmen von Big-Data-Analysen besser nachvollziehen können. Die Vermitt-

lung von Informationskompetenz sollte nicht erst beim Studium, sondern bereits in der Schule stärker eingebunden werden.

RECHTLICHE RAHMENBEDINGUNGEN

In verschiedenen Bereichen ist die Rechtssituation den Anforderungen von Big Data und Cloud Computing nicht gewachsen. Das Einhalten der Vorschriften und damit die Sicherstellung der Compliance werden für Organisationen immer aufwendiger. Im Kontext von Big Data in der Cloud ist das besonders relevant, da die Nutzung von Cloud Computing, vor allem von Public Clouds, die Rechtslage deutlich verkompliziert. Da liegt daran, dass unter anderem durch die Stärke von US-amerikanischen Unternehmen im Public-Cloud-Computing-Bereich ein grenzüberschreitender Bezug sich kaum vermeiden lässt. Besonders relevant sind beispielsweise Urheber-, Datenschutz- und Haftungsrecht.

Folgende Handlungs- und Forschungsfelder sind zu nennen:

- > Untersuchung des Anpassungsbedarfs auf Ebene der Gesetzgebung: Zu klären ist, inwieweit bestimmte Gesetze wie das BDSG oder das UrhG an die Entwicklungen in den Bereichen Big Data und Cloud Computing angepasst werden können. Teils erscheinen die vorhandenen Regelungen nicht mehr zeitgemäß, teils wurde der Anwendungsbereich aufgrund von grenzüberschreitender Datenverarbeitung unklar.
- > Untersuchung der Auswirkungen der EU-Datenschutzgrundverordnung: Die Datenschutzgrundverordnung, die aktuell im Europaparlament und im Europäischen Rat diskutiert wird, hat eine Harmonisierung sowie eine Modernisierung der nationalen Datenschutzgesetzgebungen in der EU zum Ziel. Big Data spielte im Gegensatz zu Cloud Computing in der Entwurfsphase der Grundverordnung keine wesentliche Rolle. Untersucht werden sollten deshalb die möglichen Auswirkungen eines Inkrafttretens der Grundverordnung auf die Durchführung von Big-Data-Analysen.
- > Eigentum an Daten: Während sich Daten zum zentralen Wirtschaftsgut entwickeln, kann aus juristischer Sicht kein »Eigentum« an Daten oder Informationen begründet werden. Auch wenn Gesetze, in Deutschland z. B. das UrhG, Daten unter bestimmten Voraussetzungen in beschränktem Ausmaß schützen, reichen entsprechende Lösungen für viele kompliziertere Anwendungsszenarien, vor allem mit grenzüberschreitendem Bezug, nicht aus. Um wachstumsbremsende Auswirkungen zu vermeiden, sollte Klarheit im Hinblick auf Eigentum an Daten geschaffen werden.
- > Harmonisierungen auf EU-Ebene: Zu klären ist, welche Harmonisierungen auf EU-Ebene durchsetzbar sind. Die vielen unterschiedlichen nationalen Regelungen bedeuten für die EU im Vergleich zu großen Wirtschaftsräumen wie den USA, Indien oder China einen Nachteil.



SICHERHEIT UND TECHNISCHER DATENSCHUTZ

Bei Überlegungen im Bereich Big Data in der Cloud spielt Sicherheit eine zentrale Rolle. Die Verarbeitung von personenbezogenen oder wettbewerbskritischen Daten im Rahmen von Public-Cloud-Diensten wird jedoch in der Regel aus Gründen der Sicherheit als nicht durchführbar angesehen.

Folgende Handlungs- und Forschungsfelder sind zu nennen:

- › Maßnahmen zur Gewährleistung der Informationssicherheit: Informationssicherheit spielt im Kontext von Big-Data-Analysen vor allem dann eine besondere Rolle, wenn Dienste aus einer Public Cloud genutzt werden und wenn Daten analysiert werden, die entweder personenbezogen oder wettbewerbskritisch sind. Das Problem liegt vor allem im Verlust der Kontrolle über die Daten in der Cloud. Einerseits zeigen die Snowden-Enthüllungen, wie unsicher Daten vor allem bei US-amerikanischen Anbietern vor dem Zugriff durch Geheimdienste sind, andererseits gibt es in Europa einen Mangel an Alternativen. Die Bemühungen, Europa für große und innovative Anbieter im Public-Cloud-Bereich attraktiv zu machen, sollten weiter intensiviert werden. Die technologische Souveränität von Deutschland und Europa sollte gestärkt werden, um die Abhängigkeit von anderen Märkten zu reduzieren.
- › Sicherheit und Datenschutz als explizite Anforderungen: Allerdings können auch die Ergebnisse von Big-Data-Analysen Bedenken im Hinblick auf die Sicherheit begründen. Es sind durchaus Einsatzszenarien denkbar, in denen fehlerhafte Daten oder fehlerhafte Algorithmen fatale Konsequenzen haben könnten. Insbesondere bei automatisierten Entscheidungsprozessen im Kontext von kritischen Infrastrukturen (z. B. bei Smart-City- oder Industrie-4.0-Lösungen) muss den Fragen der Sicherheit ausreichend Aufmerksamkeit gewidmet werden. Erkenntnisse aus dem Qualitätsmanagement fordern, Sicherheitsfragen bereits während der Entwicklung von Lösungen zu berücksichtigen. Diesbezüglich sind Prinzipien wie »Security by Design« und »Privacy by Design« auch im Kontext von Big Data relevant. Die Verankerung solcher Prinzipien in Projektkonzeptionen sollte durch geeignete Maßnahmen gefördert werden. Beispielsweise könnte der Aufgabenbereich von Sicherheits- bzw. Datenschutzbeauftragten entsprechend erweitert werden.





LITERATUR

- Abolhassan, F. (2013): Big Data und Cloud Computing gehören zusammen. www.zdnet.de/88165297/big-data-und-cloud-computing-gehoren-zusammen (17.2.2014)
- Acquisti, A. (2010): From the Economics to the Behavioral Economics of Privacy: A Note. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C. et al. (eds.): *Ethics and Policy of Biometrics*. Berlin/Heidelberg, S.23–26
- Acquisti, A., Krontiris, I., Langheinrich, M., Sasse, M.A. (2013): »My Life, Shared« – Trust and Privacy in the Age of Ubiquitous Experience Sharing. In: *Dagstuhl Reports* 3(7), S. 74–107
- Andersen, R. (2012): How Big Data Is Changing Astronomy (Again). www.theatlantic.com/technology/archive/2012/04/how-big-data-is-changing-astronomy-again/255917/ (20.2.2014)
- Anderson, C. (2008): The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. www.wired.com/science/discoveries/magazine/16-07/pb_theory (10.2.2014)
- Anderson, C. (2012): *Makers. The new industrial revolution*. New York
- Ang, C.S., Bobrowicz, A., Schiano, D.J., Nardi, B. (2013): Data in the wild. In: *interactions* 20(2), S. 39–43
- Appelquist, J. (2005): *Informationsteknik och organisatorisk förändring. Teknik, organisation och produktivitet inom svensk banksektor 1975–2003*. Stockholm
- Ark, B. van, Inklaar, R., McGuckin, R.H. (2003): The Contribution of ICT-Producing and ICT-Using Industries to Productivity Growth: A Comparison of Canada, Europe and the United States. In: *International Productivity Monitor* 6, S. 56–63
- Attiya, H., Welch, J. (2004): *Distributed computing. Fundamentals, simulations, and advanced topics*. Hoboken
- Aumasson, A., Bonneau, V., Leimbach, T., Gödel, M. (2010): *Economic and Social Impact of Software & Software-Based Services*. SMART 2009/0041. D5 – Final Report. Europäische Kommission, <http://cordis.europa.eu/fp7/ict/ssai/docs/study-sw-report-final.pdf> (20.2.2014)
- Bange, C. (2012): Big Data Plattformen für polystrukturierte Daten – neue Chancen und Herausforderungen. 7. Deutsche Oracle BI + DWH Konferenz. www.oracledwh.de/downloads/03_DWH_Community_und_Konferenzen_Vortraege/TB_DWH_Konferenz_2012_Koenigswinter/Carsten_Bange_BARC%20Big%20Data%20Plattformen.pdf
- Bange, C., Grosser, T., Janoschek, N. (2013): *Big Data Survey Europe. Nutzung, Technologie und Budgets europäischer Best Practice Unternehmen*. Business Application Research Center (BARC), <http://public.dhe.ibm.com/common/ssi/ecm/de/iml14371dede/IML14371DEDE.PDF> (20.2.2014)
- Barret, V. (2011): Dropbox: The Inside Story Of Tech's Hottest Startup. www.forbes.com/sites/victoriabarret/2011/10/18/dropbox-the-inside-story-of-techs-hottest-startup/4/ (21.2.2014)
- Beath, C., Becerra-Fernandez, I., Ross, J., Short, J. (2012): Finding Value in the Information Explosion. In: *MIT Sloan Management Review* 53(4), S. 18–20



LITERATUR

- Behörden Spiegel (2013): Big Data – Big Chance für die Verwaltung? Bonn, www.behoerden-spiegel.de/icc/Internet/sub/c0e/c0e51fee-f724-d316-f60c-9157b988f2ee,,,aaaaaaaa-aaaa-aaaa-bbbb-000000000003&uMen=1f75009d-e07d-f011-4e64-494f59a5fb42.htm (20.2.2014)
- Bergen, P., Sterman, D., Schneider, E., Cahall, B. (2014): Do NSA's Bulk Surveillance Programs Stop Terrorists? New America Foundation, www.newamerica.net/sites/newamerica.net/files/policydocs/Bergen_NAF_NSA%20Surveillance_1_0_0.pdf (20.2.2014)
- Best, J. (2013): IBM Watson: The inside story of how the Jeopardy-winning supercomputer was born, and what it wants to do next. TechRepublic. www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/ (20.2.2014)
- Birkenheuer, G., Blunk, D., Breuers, S., Brinkmann, A., dos Santos Vieira, I., Fels, G. et al. (2011): MoSGrid – a molecular simulation grid as a new tool in computational chemistry, biology and material science. In: Journal of Cheminformatics 3(Suppl 1), S. P14
- BITKOM (Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V.) (2012): Big Data im Praxiseinsatz. Szenarien, Beispiele, Effekte. [www.bitkom.org/files/documents/BITKOM_LF_big_data_2012_online\(1\).pdf](http://www.bitkom.org/files/documents/BITKOM_LF_big_data_2012_online(1).pdf) (20.2.2014)
- BITKOM (2013): Umsatz mit Cloud Computing steigt auf fast 8 Milliarden Euro. www.bitkom.org/de/presse/8477_75301.aspx (20.2.2014)
- BITKOM (2014): Breitbandausbau: Deutschland in der Spitzengruppe. www.bitkom.org/de/presse/8477_78434.aspx (2.4.2014)
- Bleyl, F. (2013): Intelligent verknüpft. In: Pictures of the Future, S. 27
- Böckle, R. (2014): Warum Private und hybride Clouds das Rennen machen. www.computerwoche.de/a/warum-private-und-hybride-clouds-das-rennen-machen,3041043 (20.2.2014)
- Böhler, T.M. (2012): Industrie 4.0 – Smarte Produkte und Fabriken revolutionieren die Industrie. Produktion. www.produktion.de/automatisierung/industrie-4-0-smarte-produkte-und-fabriken-revolutionieren-die-industrie/ (10.3.2014)
- Bond-Graham, D., Winston, A. (2013): All Tomorrow's Crimes: The Future of Policing Looks a Lot Like Good Branding. www.sfweekly.com/2013-10-30/news/predpol-sfpd-predictive-policing-compstat-lapd/full/ (6.3.2014)
- Borja, F. de (2012): Cloud Computing in Europe Lagging Behind. CloudTimes. <http://cloudtimes.org/2012/09/11/cloud-computing-europe/> (20.2.2014)
- Börnsen, A. (2012): Breitbandversorgung 2020. Entwicklungen, Ziele und Förderinstrumente. Friedrich Ebert Stiftung <http://library.fes.de/pdf-files/wiso/09070.pdf> (8.4.2014)
- Bose, R. (2009): Advanced analytics: opportunities and challenges. In: Industrial Management & Data Systems 109(2), S. 155–172
- Bowden, C. (2013): The US surveillance programmes and their impact on EU citizens' fundamental rights. www.europarl.europa.eu/meetdocs/2009_2014/documents/libedv/briefingnote_/briefingnote_en.pdf (3.4.2014)
- Boyd, D., Crawford, K. (2011): Six Provocations for Big Data. Microsoft Research. Redmond

- Boyd, D., Crawford, K. (2012): Critical questions for big data. In: *Information, Communication & Society* 15(5), S. 662–679
- Bradshaw, D., Folco, G., Cattaneo, G., Kolding, M. (2012): Quantitative Estimates of the Demand for Cloud Computing in Europe and the Likely Barriers to Take-up. SMART 2011/0045. D4 – Final Report. Europäische Kommission, <http://cordis.europa.eu/fp7/ict/ssai/docs/study45-d2-interim-report.pdf>
- Bradshaw, S., Millard, C., Walden, I. (2010): Contracts for Clouds: Comparison and Analysis of the Terms and Conditions of Cloud Computing Services. Legal Studies Research Paper No. 63/2010. Queen Mary School of Law, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1662374 (1.4.2014)
- Brewer, E.A. (2000): Towards Robust Distributed Systems. 19th ACM Symposium on Principles of Distributed Computing. ACM. www.cs.berkeley.edu/~brewer/PODC2000.pdf
- Breznitz, D. (2006): The Israeli Software Industry. In: Arora, A., Gambardella, A. (Hg.): From underdogs to tigers. The rise and growth of the software industry in Brazil, China, India, Ireland, and Israel. Oxford, S. 72–98
- Breznitz, D. (2011): Innovation and the state. Political choice and strategies for growth in Israel, Taiwan, and Ireland. New Haven/London
- Brynjolfsson, E. (1993): The productivity paradox of information technology. In: *Communications of the ACM* 36(12), S. 66–77
- Brynjolfsson, E., Hitt, L., Heekyung, K. (2011): Strength in Numbers: How does data-driven decision-making affect firm performance? In: Proceedings of the 2011 International Conference on Information Systems. Paper 13, http://www.a51.nl/storage/pdf/SSRN_id1819486.pdf
- Brynjolfsson, E., Hofmann, P., Jordan, J. (2010): Cloud computing and electricity: Beyond the utility model. In: *Communications of the ACM* 53(5), S. 32–34
- Brynjolfsson, E., McAfee, A. (2012): Race Against the Machine: how the revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy. Lexington
- Brynjolfsson, E., McAfee, A. (2014): The second machine age. Work, progress, and prosperity in a time of brilliant technologies. New York
- Bryson, S., Kenwright, D., Cox, M., Ellsworth, D., Haimes, R. (1999): Visually exploring gigabyte data sets in real time. In: *Communications of the ACM* 42(8), S. 82–90
- Bundesregierung (2013): Deutschland ist ein Land der Freiheit. www.bundesregierung.de/Content/DE/Artikel/2013/07/2013-07-19-bkin-nsa-sommerpk.html (22.7.2013)
- Bunz, M. (2012): Die stille Revolution. Wie Algorithmen Wissen, Arbeit, Öffentlichkeit und Politik verändern, ohne dabei viel Lärm zu machen. Berlin
- Burns, L.D. (2013): Sustainable mobility: A vision of our transport future. In: *Nature* 497(7448), S. 181–182
- Buyya, R., Yeo, C.S., Venugopal, S. (2008): Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities. In: Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications, Los Alamitos, S. 5–13



LITERATUR

- Byczkowski, T. (2013): Autonome Autos. Computer am Steuer. www.zeit.de/zeitwissen/2013/03/autonomes-auto-google-fahrzeugindustrie (18.2.2014)
- Carr, N.G. (2001): *The digital enterprise. How to reshape your business for a connected world.* Boston
- Carr, N.G. (2005): The end of corporate computing. In: MIT Sloan Management Review 46(3), S. 66–73
- Cassell, J. (2013): The »Fourth Industrial Revolution« Brings Challenges to Big Data. <http://news.thomasnet.com/IMT/2014/02/11/the-fourth-industrial-revolution-brings-challenges-to-big-data/> (10.3.2014)
- Castro, D. (2013): How much will PRISM cost the U.S. Cloud Computing industry? www2.itif.org/2013-cloud-computing-costs.pdf (20.2.2014)
- Cattaneo, G., Kolding, M., Bradshaw, D., Folco, G. (2012): Quantitative Estimates of the Demand for Cloud Computing in Europe and the Likely Barriers to Take-up. SMART 2011/0045. D2 – Interim Report – Statistical Annex. Europäische Kommission, <http://cordis.europa.eu/fp7/ict/ssai/docs/study45-d2-interim-statisticalannex.pdf> (21.2.2014)
- Catteddu, D. (2011): Security & Resilience in Governmental Clouds. Making an informed decision. European Network and Information Security Agency (ENISA), www.enisa.europa.eu/activities/risk-management/emerging-and-future-risk/deliverables/security-and-resilience-in-governmental-clouds/at_download/fullReport (3.4.2014)
- Catteddu, D., Hogben, G. (2009): An SME perspective on Cloud Computing. Survey. European Network and Information Security Agency (ENISA), www.enisa.europa.eu/activities/risk-management/files/deliverables/cloud-computing-sme-survey/at_download/fullReport (3.4.2014)
- CDU/CSU, SPD (2013): *Deutschlands Zukunft gestalten, Koalitionsvertrag zwischen CDU, CSU und SPD.* Berlin
- Chen, H., Chiang, R.H.L., Storey, V.C. (2012): Business intelligences and analytics: From big data to big impact. In: MIS Quarterly 36(4), S. 1165–1188
- Chellappa, R.K. (1997): Intermediaries in Cloud-Computing: A New Computing Paradigm. INFORMS Annual Meeting, October 26-29, Dallas
- Cloud Industry Forum (2012a): *Cloud UK. UK Cloud Adoption and Trends for 2013.* High Wycombe
- Cloud Industry Forum (2012b): *USA Cloud Adoption and Trends 2012.* High Wycombe
- Cowen, T. (2011): *The great stagnation. How America ate all the low-hanging fruit of modern history, got sick, and will (eventually) feel better.* New York
- Cox, M., Ellsworth, D. (1997): Application-controlled demand paging for out-of-core visualization. In: Dill, J., Gershon, N. (eds.): *Proceedings of the IEEE Symposium on Information Visualization.* Los Alamitos, S. 235–244
- CPR (Det Centrale Personregister) (2009): *Udviklingen på CPR-området i de seneste 20-25 år frem til 2009.* <https://cpr.dk/media/165161/4396.pdf> (20.2.2014)
- Croll, A. (2013): The vanishing cost of guessing. <http://radar.oreilly.com/2013/08/the-vanishing-cost-of-guessing.html> (1.4.2014)



- Davenport, T.H. (2006): Competing on analytics. In: Harvard Business Review 84(1), S. 98–107
- Dean, D., Digrande, S., Field, D., Lundmark, A., O'Day, J., Pineda, J., Zwillenberg, P. (2012): The Connected World. The \$4.2 Trillion Opportunity. The Internet Economy in the G-20. Boston Consulting Group, https://publicaffairs.linx.net/news/wp-content/uploads/2012/03/bcg_4trillion_opportunity.pdf (2.4.2014)
- DeNardis, L. (2009): Protocol politics. The globalization of Internet governance. Cambridge
- Dignan, L. (2011): Cloud computing market: \$241 billion in 2020. www.zdnet.com/blog/btl/cloud-computing-market-241-billion-in-2020/47702 (20.2.2014)
- Dillon, T., Wu, C., Chang, E. (2010): Cloud computing: Issues and challenges. In: Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications. Los Alamitos, S. 27–33
- DIW (Deutsches Institut für Wirtschaftsforschung) (2008): An Economic Assessment of ICT Adoption and its Impact on Innovation and Performance. http://ec.europa.eu/enterprise/archives/e-business-watch/studies/special_topics/2007/documents/Study_10-2008_ICT-Impact.pdf (2.4.2014)
- DIW econ (2010): Economic Implications of Cloud Computing. A Report for Microsoft. www.amcham.de/fileadmin/user_upload/TIM/DIW_econ_Cloud_Computing_v1_0_2_.pdf (2.4.2014)
- Doctorow, C. (2014): Chicago PD's Big Data: using pseudoscience to justify racial profiling. <http://boingboing.net/2014/02/25/chicago-pds-big-data-using.html> (7.4.2014)
- Dumbill, E. (2012): What is big data? An introduction to the big data landscape. <http://radar.oreilly.com/2012/01/what-is-big-data.html> (20.2.2014)
- Eckert, M., Bry, F. (2009): Complex Event Processing (CEP). In: Informatik-Spektrum 32(2), S. 163–167
- Edwards, L. (2014): Military robot truck convoy autonomously drives without weak fleshy humans. www.pocket-lint.com/news/126984-military-robot-truck-convoy-autonomously-drives-without-weak-fleshy-humans (17.2.2014)
- EFI (Expertenkommission Forschung und Innovation) (2011): Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands 2011. www.e-fi.de/fileadmin/Gutachten/2011_deu.pdf (2.4.2014)
- EGI (European Grid Infrastructure) (2013): EGI in numbers. www.egi.eu/infrastructure/operations/egi_in_numbers (20.2.2014)
- Eicher, T.S., Strobel, T. (2009): Information Technology and Productivity Growth – German Trends and OECD Comparisons. Cheltenham
- EITO (European Information Technology Observatory) (2012a): ICT Market Report 2012/13. Berlin
- EITO (2012b): Trendkongress diskutiert neue Entwicklungen der digitalen Welt. www.eito.com/press/Press-Releases-2012/Trendkongress-diskutiert-neue-Entwicklungen-der-digitalen-Welt (20.2.2014)
- EITO (2013): Big Data in Europe: Evolution and Revolution. Berlin



LITERATUR

- EIU (Economist Intelligence Unit) (2012): The Deciding Factor: Big Data & Decision Making. www.capgemini.com/sites/default/files/resource/pdf/The_Deciding_Factor_Big_Data__Decision_Making.pdf (20.2.2014)
- EK (Europäische Kommission) (2009): IKT-Infrastrukturen für die e-Wissenschaft. EC COM 2009/108. <http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52009DC0108&from=EN> (9.4.2014)
- EK (2010): Riding the wave. How Europe can gain from the rising tide of scientific data. http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707 (20.2.2014)
- EK (2011): High-Level Expert Group on Key Enabling Technologies. Final Report. http://ec.europa.eu/enterprise/sectors/ict/files/kets/hlg_report_final_en.pdf (3.4.2014)
- EK (2012a): Unleashing the Potential of Cloud Computing in Europe. EC COM 2012/529. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0529:FIN:EN:PDF> (10.3.2014)
- EK (2012b): Unleashing the Potential of Cloud Computing in Europe – What is it and what does it mean for me? http://europa.eu/rapid/press-release_MEMO-12-713_en.htm (20.2.2014)
- EK (2013a): Broadband coverage in Europe in 2012. Mapping progress towards the coverage objectives of the Digital Agenda: final report. Luxembourg
- EK (2013b): Horizon 2020, Work Program 2014-2015. Section 16: Science with and for society. European Commission Decision C (2013)8631 of 10 December 2013. Brüssel
- EK (2013c): Horizon 2020, Work Program 2014-2015. Section 4: European research infrastructures (including e-Infrastructures). European Commission Decision C (2013)8631 of 10 December 2013. Brüssel
- Elsener, R. (2013): Gesundheitsnetz in den USA. In: Pictures of the Future, S. 41–43
- EMC (2012): Big Data-as-a-Service. A Market and Technology Perspective. www.emc.com/collateral/software/white-papers/h10839-big-data-as-a-service-perspt.pdf (20.2.2014)
- Epstein, E.A., Schor, M.I., Iyer, B.S., Lally, A., Brown, E.W., Cwikilik, J. (2012): Making Watson fast. In: IBM Journal of Research and Development 56(3.4), S. 15.1–15.12
- Ermert, M. (2013): NSA-Skandal: IETF bezweifelt Vertrauenswürdigkeit der NIST. www.heise.de/netze/meldung/NSA-Skandal-IETF-bezweifelt-Vertrauenswuerdigkeit-der-NIST-2042800.html (3.4.2014)
- ER (Europäischer Rat) (2013): European Council, 24/25 October 2013 Conclusions, EUCO 169/13. www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/139197.pdf (17.4.2014)
- Ermert, M. (2014): USA wollen Kontrolle über Internetverwaltung abgeben. www.heise.de/newsticker/meldung/USA-wollen-Kontrolle-ueber-Internetverwaltung-abgeben-2146449.html (3.4.2014)
- Ester, M., Sander, J. (2000): Knowledge discovery in databases. Techniken und Anwendungen. Berlin/New York

- Etro, F. (2009): The Economic Impact of Cloud Computing on Business Creation, Employment and Output in Europe. An application of the Endogenous Market Structures Approach to a GPT innovation. In: *Review of Business and Economics* 54(2), S. 179–208
- Etro, F. (2011): The Economics of Cloud Computing. In: *The IUP Journal of Managerial Economics* 9(2), S. 7–22
- Experton Group (2013): Experton Group Big Data Vendor Benchmark 2013. www.experton-group.de/press/releases/pressrelease/article/experton-group-big-data-vendor-benchmark-2013.html (20.2.2014)
- Feinleib, D. (2013): The Big Data Landscape. The Big Data Group. www.slideshare.net/bigdatalandscape/the-big-data-landscape (20.2.2014)
- Ferrucci, D., Lally, A. (1999): UIMA: An architectural approach to unstructured information processing in the corporate research environment. In: *Natural Language Engineering* 10(3-4), S. 327–348
- Few, S. (2012): Criteria for Evaluating Visual EDA Tools. www.perceptualedge.com/articles/visual_business_intelligence/evaluating_visual_eda_tools.pdf (20.2.2014)
- Fielder, A., Brown, I., Weber, V., McSpedden-Brown, N. (2012): Cloud computing. Study. Europaparlament. Brüssel/Straßburg
- Fitze, U. (2013): Smart Cities: Urban Pioneers with Great Business Potential. In: *Pictures of the Future*, S. 34–35
- Forbes Insights, Cisco (2013): Collaborating in the Cloud. <http://images.forbes.com/forbesinsights/StudyPDFs/cisco-cloud-report.pdf> (8.4.2014)
- Forrester (2012): Personal Cloud Services emerge to orchestrate our Mobile Computing Lives. www.sugarsync.com/media/sugarsync-forrester-report.pdf (21.2.2014)
- Foster, I. (2002): What is the grid? – A three point checklist. In: *GRIDtoday* 1(6) www.mcs.anl.gov/~itf/Articles/WhatIsTheGrid.pdf
- Foster, I., Zhao, Y., Raicu, I., Lu, S. (2008): Cloud computing and grid computing 360-degree compared. In: *Proceedings of the Grid Computing Environments Workshop*. Los Alamitos, S. 1–10
- Fransman, M. (2011): The Evolving ICT Industry in Asia and the Implications for Europe. Europäische Kommission, http://ftp.jrc.es/EURdoc/JRC63986_TN.pdf (3.4.2014)
- Friend, Z. (2013): Predictive Policing: Using Technology to Reduce Crime. www.fbi.gov/stats-services/publications/law-enforcement-bulletin/2013/April/predictive-policing-using-technology-to-reduce-crime (18.2.2014)
- Furrier, J. (2012): Big Data Is Big Market & Big Business – \$50 Billion Market by 2017. www.forbes.com/sites/siliconangle/2012/02/17/big-data-is-big-market-big-business (20.2.2014)
- Galvani, A.P., May, R.M. (2005): Epidemiology: Dimensions of superspreading. In: *Nature* 438(7066), S. 293–295
- Gantz, J., Reinsel, D. (2012): *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*. IDC, Framingham
- Garfinkel, S.L. (1999): *Architects of the information society. Thirty-five years of the Laboratory for Computer Science at MIT*. Cambridge



LITERATUR

- Gartner (2009): Gartner Highlights Five Attributes of Cloud Computing. www.gartner.com/newsroom/id/1035013 (20.2.2014)
- Gartner (2012a): Gartner Says Cloud Adoption in Europe Will Trail U.S. by At Least Two Years. www.gartner.com/newsroom/id/2032215 (20.2.2014)
- Gartner (2012b): Gartner Says Nearly One Third of Organizations Use or Plan to Use Cloud Offerings to Augment Business Intelligence Capabilities. www.gartner.com/newsroom/id/1903814 (20.2.2014)
- Gartner (2012c): Gartner Says Worldwide Cloud Services Market to Surpass \$109 Billion in 2012. www.gartner.com/newsroom/id/2163616 (20.2.2014)
- Gartner (2014): By 2016, 25 Percent of Large Global Companies Will Have Adopted Big Data Analytics For At Least One Security or Fraud Detection Use Case. www.gartner.com/newsroom/id/2663015 (18.2.2014)
- Gater, C. (2011): Grid Use Cases. European Grid Infrastructure (EGI), https://documents.egi.eu/public/RetrieveFile?docid=643&version=1&filename=egi_inspire_useslides.pdf (20.2.2014)
- Gaycken, S. (2014): Freies Netz – eine Illusion. www.sueddeutsche.de/digital/zukunft-des-internets-freies-netz-eine-illusion-1.1921160 (3.4.2014)
- Geiselberger, H., Moorstedt, T. (Hg.) (2013): Big Data. Das neue Versprechen der Allwissenheit. Berlin
- Gilbert, S., Lynch, N. (2002): Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. In: ACM SIGACT News 33(2), S. 51–59
- Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L. (2009): Detecting influenza epidemics using search engine query data. In: Nature 457(7232), S. 1012–1014
- Global Pulse (2012): Big Data for Development: Challenges & Opportunities. www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf (26.2.2014)
- Glockner, H., Beyers, B. (2012): Big Data 2030. In: Zukunftsmanager (6), S. 31–34
- Golliez, A., Aschwanden, C., Bretscher, C., Bernstein, A. et al. (2012): Open Government Data Studie Schweiz. Berner Fachhochschule, www.itopia.ch/repository/Publikationen/OGD_Studie_Schweiz_Juni_2012.pdf (1.4.2014)
- Gomes, L. (2011): The Challenges of Big Data on the Smart Grid. MIT Technology Review. www.technologyreview.com/news/424088/the-challenges-of-big-data-on-the-smart-grid/ (20.2.2014)
- Graumann, S., Speich, A. (2011): Monitoring-Report Deutschland Digital 2011. Kurzfassung. Bundesministerium für Wirtschaft und Energie (BMWi), www.bmwi.de/Dateien/BMWi/PDF/IT-Gipfel/ikt-monitoring-kurz,property=pdf,bereich=bmwi2012,sprache=de,rwb=true.pdf (21.2.2014)
- Greif, B. (2014): Amazon-Patent sieht Paketversand noch vor der Bestellung vor. www.zdnet.de/88181594/amazon-patent-sieht-paketversand-noch-bestellung (2.4.2014)
- Haarmann, T. (2013): Citizen Science: Zocken für die Forschung. www.zeit.de/2013/38/citizen-science-eyewire-seti (20.2.2014)
- Hackmann, J. (2014): NSA-Skandal schadet der US-Industrie. www.computerwoche.de/a/nsa-skandal-schadet-der-us-industrie,2553863 (20.2.2014)



- Hafner, K., Lyon, M. (1998): Where wizards stay up late. The origins of the Internet. New York
- Hand, E. (2010): Citizen science: People power. In: *Nature* 466(7307), S. 685–687
- Harris, D. (2013): Why Apple, eBay, and Walmart have some of the biggest data warehouses you've ever seen. <http://gigaom.com/2013/03/27/why-apple-ebay-and-walmart-have-some-of-the-biggest-data-warehouses-youve-ever-seen> (20.2.2014)
- Harris, R. (2012): International Council for Science (ICSU) and the Challenges of Big Data in Science. In: *Research Trends* (30), S. 11–12
- Harvard School of Public Health (2012): Using cell phone data to curb the spread of malaria. www.hsph.harvard.edu/news/press-releases/cell-phone-data-malaria (20.2.2014)
- Hauck, M. (2013): Mayer-Schönberger über Big Data – »Daten sammeln ist richtig«. www.sueddeutsche.de/digital/viktor-mayer-schoenberger-ueber-big-data-daten-sammeln-ist-richtig-1.1817452 (20.2.2014)
- Helios, A. (2013): Big Data – Big Deal? In: *Zukunftsmanager* (5), S. 15–17
- Heng, S. (2011): Netzneutralität. Innovation und Differenzierung keine Antipoden. www.dbresearch.com/PROD/DBR_INTERNET_EN-PROD/PROD0000000000277416.pdf (2.4.2014)
- Henschen, D. (2012): 13 Big Data Vendors To Watch In 2013. www.informationweek.com/software/information-management/13-big-data-vendors-to-watch-in-2013/240144124 (20.2.2014)
- Heumann, S., Scott, B. (2013): Law and Policy in Internet Surveillance Programs: United States, Great Britain and Germany. In: *Impulse* (25), S. 1–17
- Hey, T., Tansley, S., Tolle, K. (eds.) (2009): The fourth paradigm. Data-intensive scientific discovery. Microsoft, Redmond
- Higgins, J.K. (2012): Federal Cloud Adoption, Part 2: Raining Contracts. www.ecommercetimes.com/story/75125.html (20.2.2014)
- Hilbert, M., Lopez, P. (2011): The World's Technological Capacity to Store, Communicate, and Compute Information. In: *Science* 332(6025), S. 60–65
- Hintemann, R., Fichter, K. (2012): Energieverbrauch und Energiekosten von Servern und Rechenzentren in Deutschland. Aktuelle Trends und Einsparpotenziale bis 2015. Borderstep Institut, [www.bitkom.org/files/documents/Kurzstudie__Borderstep_I_Rechenzentren\(1\).pdf](http://www.bitkom.org/files/documents/Kurzstudie__Borderstep_I_Rechenzentren(1).pdf) (10.3.2014)
- Hogan, O., Mohamed, S., McWilliams, D., Greenwood, R. (2010): The Cloud Dividend: Part One. The economic benefits of cloud computing to business and the wider EMEA economy. France, Germany, Italy, Spain and the UK. Centre for Economics and Business Research (CEBR), <http://uk.emc.com/collateral/microsites/2010/cloud-dividend/cloud-dividend-report.pdf> (11.3.2014)
- Hogan, O., Mohamed, S., McWilliams, D., Greenwood, R. (2011): The Cloud Dividend: Part Two. The economic benefits of cloud computing to business and the wider EMEA economy. Comparative analysis of the impact on aggregated industry sectors. Centre for Economics and Business Research (CEBR), www.globb tv.com/microsite/35/Adjuntos/CLOUD-DIVIDEND-REPORT.PDF (11.3.2014)
- Hon, W.K., Millard, C., Walden, I. (2012): Negotiating Cloud contracts: Looking at Clouds from both sides now. In: *Stanford Technology Law Review* 16(1), S. 79–129



LITERATUR

- Hoorens, S., Elixmann, D., Cave, J., Li, M.S., Cattaneo, G. (2012): Towards a competitive European Internet industry. A socio-economic analysis of the European Internet industry and the Future Internet Public-Private Partnership. RAND Europe, www.rand.org/content/dam/rand/pubs/technical_reports/2012/RAND_TR1262.pdf (2.4.2014)
- Hottelet, U. (2014): Deutsche Cloud-Anbieter punkten mit Datenschutz. www.compu-woche.de/a/deutsche-cloud-anbieter-punkten-mit-datenschutz,2553692 (20.2.2014)
- HP (2012): Big security for big data. <https://ssl.www8.hp.com/ww/en/secure/pdf/4aa4-4051enw.pdf> (18.2.2014)
- Huber, B., Kurnikowski, A., Müller, S., Pozar, S. (2013): Die wirtschaftliche und politische Dimension von Open Government in Österreich. WU Wien, www.data.gv.at/wp-content/uploads/2012/03/die_wirtschaftliche_und_politische_dimension_von_open_government_data_in_oesterreich_final.pdf (1.4.2014)
- Iansiti, M., Levien, R. (2004): Strategy as Ecology. In: Harvard Business Review 82(3), S. 68–81
- IBM (o. J.): Watson hilft Ärzten, Krebs zu bekämpfen. <http://www-05.ibm.com/de/watson/gesundheitswesen.html> (10.3.2014)
- IBM (2012): Smarter, More Competitive Cities. Forward-thinking cities are investing in insight today. <http://public.dhe.ibm.com/common/ssi/ecm/en/pub03003usen/PUB03003USEN.PDF> (21.2.2014)
- IBM (2013a): Memorial Sloan-Kettering Cancer Center. IBM Watson helps fight cancer with evidence-based diagnosis and treatment suggestions. www-03.ibm.com/innovation/us/watson/pdf/MSK_Case_Study_IMC14794.pdf (21.2.2014)
- IBM (2013b): WellPoint, Inc. IBM Watson enables more effective healthcare preapproval decisions using evidence-based learning. www-03.ibm.com/innovation/us/watson/pdf/WellPoint_Case_Study_IMC14792.pdf (21.2.2014)
- ICANN (Internet Corporation for Assigned Names and Numbers) (2013): Montevideo Statement on the Future of Internet Cooperation. www.icann.org/en/news/announcements/announcement-07oct13-en.htm (3.4.2014)
- IDC (Central Europe GmbH) (2012a): Cloud Computing in Deutschland 2012. Deployment-Modelle und Management, Integration, Security und Compliance im Fokus. www.kaspersky.com/de/downloads/pdf/idc_executive_brief_mc_cloud_computing_2012_kaspersky.pdf (21.2.2014)
- IDC (2012b): IDC Forecasts Public IT Cloud Services Spending Will Approach \$100 Billion in 2016, Generating 41 % of Growth in Five Key IT Categories. www.idc.com/getdoc.jsp?containerId=prUS23684912 (20.2.2014)
- IDC (2013a): IDC Forecasts Worldwide Public IT Cloud Services Spending to Reach Nearly \$108 Billion by 2017 as Focus Shifts from Savings to Innovation. www.idc.com/getdoc.jsp?containerId=prUS24298013 (20.2.2014)
- IDC (2013b): IDC Forecasts Worldwide Spending on Hosted Private Cloud Services to Surpass \$24 Billion in 2016. www.idc.com/getdoc.jsp?containerId=prUS23972413 (20.2.2014)
- Inmon, W.H. (2005): Building the data warehouse. Indianapolis

- Interxion (2013): Big Data – Beyond the Hype. www.interxion.com/Documents/Whitepapers%20and%20PDFs/Big%20Data/BigData_Beyond_the_Hype_eng.pdf (21.2.2014)
- Issenberg, S. (2012): How President Obama's campaign used big data to rally individual voters. MIT Technology Review. Cambridge www.technologyreview.com/featured-story/509026/how-obamas-team-used-big-data-to-rally-voters/ (20.2.2014)
- ITU (International Telecommunication Union) (2013): The State of Broadband 2013. Universalizing Broadband. www.broadbandcommission.org/documents/bb-annual-report2013.pdf (2.4.2014)
- Jackson, W. (2012): How models got a complex storm like Sandy mostly right. <http://gcn.com/articles/2012/10/30/how-models-got-a-complex-storm-sandy-mostly-right.aspx> (27.2.2014)
- Jacobs, A. (2009): The pathologies of big data. In: Communications of the ACM 52(8), S. 36–44
- Janssen, M., Charalabidis, Y., Zuiderwijk, A. (2012): Benefits, Adoption Barriers and Myths of Open Data and Open Government. In: Information Systems Management 29(4), S. 258–268
- Johnson, M.E. (2012): Repost This Hurricane Sandy: Big Data Predicted Big Power Outages. www.informationweek.com/big-data/big-data-analytics/hurricane-sandy-big-data-predicted-big-power-outages/d/d-id/1107329? (28.2.2014)
- Johnson, S. (2013): Silicon Valley: NSA spying could prove costly to Bay Area and other Internet businesses, experts say. www.mercurynews.com/business/ci_24437687/nsa-spying-could-prove-costly-bay-area-and (20.2.2014)
- Kalenda, F., Pößneck, L. (2011): Bericht: »Bundes-Cloud« für deutsche Firmen und Behörden in Planung. www.zdnet.de/41558883/bericht-bundes-cloud-fuer-deutsche-firmen-und-behoerden-in-planung (20.2.2014)
- Kamel Boulos, M.N., Sanfilippo, A.P., Corley, C.D., Wheeler, S. (2010): Social Web mining and exploitation for serious applications: Technosocial Predictive Analytics and related technologies for public health, environmental and national security surveillance. In: Computer Methods and Programs in Biomedicine 100(1), S. 16–23
- Kaplan, R.S., Norton, D.P. (1997): Balanced scorecard. Strategien erfolgreich umsetzen. Stuttgart
- Kaplan, R.S., Norton, D.P. (2009): Der effektive Strategieprozess. Erfolgreich mit dem 6-Phasen-System. Frankfurt a.M.
- Katragkou, E., Zanis, P., Kioutsioukis, I., Tegoulas, I., Melas, D., Krüger, B.C., Coppola, E. (2011): Future climate change impacts on summer surface ozone from regional climate-air quality simulations over Europe. In: Journal of Geophysical Research 116(D22)
- Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F. (2010): Mastering the Information Age. Solving Problems With Visual Analytics. Goslar
- Kerr, I., Earle, J. (2013): Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy. In: Stanford Law Review Online 66, S. 65–72
- Klessmann, J., Löhe, M.G., Müller, L.-S. (2014): Digitale Teilhabe. Fraunhofer FOKUS; Kompetenzzentrum Öffentliche IT. www.oeffentliche-it.de/documents/18/21941/OeFIT-Expertise+Digitale+Teilhabe



LITERATUR

- Kommune21 (2012): Big Data. Kommunen auf dem Weg. www.kommune21.de/meldung_13367_Kommunen+auf+dem+Weg.html (20.2.2014)
- Kommune21 (2013): Big Data. Ungehobener Schatz? www.kommune21.de/meldung_15670_Ungehobener+Schatz%3F.html (20.2.2014)
- Kommune21 (2014): Big Data. Reichlich Potenzial. www.kommune21.de/meldung_17987_Reichlich+Potenzial.html (26.2.2014)
- Konkel, F. (2012): Sandy shows storm-prediction progress. <http://fcw.com/articles/2012/11/05/sandy-hurricane-center.aspx> (27.2.2014)
- Konstan, J.A., Riedl, J. (2012): Deconstructing recommender systems. How Amazon and Netflix predict your preferences and prod you to purchase. <http://spectrum.ieee.org/computing/software/deconstructing-recommender-systems> (8.4.2014)
- Korolov, M. (2013): 15 most powerful Big Data companies. Network World. Framingham www.networkworld.com/slideshow/114134/#slide1 (20.2.2014)
- Kosinski, M., Stillwell, D., Graepel, T. (2013): Private traits and attributes are predictable from digital records of human behavior. In: Proceedings of the National Academy of Sciences 110(15), S. 5802–5805
- KPMG (2012): Exploring the Cloud. A Global Study of Governments' Adoption of Cloud. www.kpmg.com/ES/es/ActualidadNovedades/ArticulosyPublicaciones/Documents/Exploring-the-Cloud.pdf (21.2.2014)
- KPMG (2013a): Breaking through the cloud adoption barriers. KPMG Cloud Providers Survey. www.kpmg.com/Global/en/IssuesAndInsights/ArticlesPublications/cloud-service-providers-survey/Documents/cloud-service-providers-survey.pdf (21.2.2014)
- KPMG (2013b): The cloud takes shape. Global cloud survey: the implementation challenge. Zug
- KPMG, BITKOM (2013): Cloud-Monitor 2013. Cloud-Computing in Deutschland – Status quo und Perspektiven. Zug/Berlin
- Krüger-Brand, H.E. (2013): Big Data. Big Problem? In: Deutsches Ärzteblatt 110(38), S. 1728
- Küll, U. (o. J.): Im Rausch der Geschwindigkeit: Big Data und Echtzeitanalysen revolutionieren Business Intelligence. heise online. Hannover www.heise.de/microsites/big-data-so-beherrschen-sie-die-datenflut/big-data-und-echtzeitanalysen-revolutionieren-business-intelligence/150/304/893/1 (20.2.2014)
- LaMonica, M. (2013): Additive Manufacturing. MIT Technology Review. Cambridge www.technologyreview.com/featuredstory/513716/additive-manufacturing/ (2.4.2014)
- Landrock, H. (2013a): Big Data Leader 2013. Big-Data-Anbieter im Vergleich. T-Systems, Experton Group, Ismaning
- Landrock, H. (2013b): Der Anbietermarkt für Big Data Lösungen – ambitioniert, aber noch unreif. Erste Eindrücke vom Big Data Vendor Benchmark 2013. isi Medien, München
- Laney, D. (2001): 3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group. Stamford <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (21.2.2014)

- Laugesen, N.S., Lauritzen, J.R., Carpenter, G., Ellegaard, C.E., Bucher, M., Stabe, M. (2012): Cloud Computing, Cyber Security and Green IT. The impact on e-Skills requirements. Europäische Kommission, http://ec.europa.eu/enterprise/sectors/ict/files/eskills/e-skills_and_cloud_computing_final_report_en.pdf (1.4.2014)
- Layo, I. (2012): Gartner: »Personal Cloud« to Replace Traditional Business IT Solutions by 2014. <http://cloudtimes.org/2012/03/22/personal-computers-will-be-traded-for-personal-cloud-by-2014/> (20.2.2014)
- Lazer, D., Kennedy, R., King, G., Vespignani, A. (2014): Big data. The parable of Google Flu: traps in big data analysis. In: *Science (New York)* 343(6176), S. 1203–1205
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D. et al. (2009): Social science. Computational social science. In: *Science (New York)* 323(5915), S. 721–723
- Leberecht, T. (2013): Why Big Data will never beat business intuition. <http://management.fortune.cnn.com/2013/06/20/big-data/> (7.4.2014)
- Leimbach, T., Hallinan, D., Bachlechner, D., Weber, A., Jaglo, M., Hennen, L. et al. (2014): Potential and Impacts of Cloud Computing Services and Social Network Websites. Europaparlament/STOA. www.europarl.europa.eu/stoa/cms/cache/offoffice/home/publications/studies;jsessionid=879DFA2884B2E3276831366FA5CF8B29?reference=IPOL-JOIN_ET%282014%29513546 (31.3.2014)
- Leimbach, T., Kayser, V., Weber, A., Jaglo, M., Hennen, L., Nentwich, M. et al. (2013): Foundation of Cloud Computing. Deliverable No. 2 of the STOA Project »Potential and Impacts of Cloud Computing Services and Social Network Sites«. Europaparlament/STOA. Brüssel/Straßburg
- Leimbach, T., Wydra, S. (2012): Software-Atlas Deutschland 2012. Fraunhofer ISI, www.softwareag.com/de/images/Fraunhofer%20ISI%20-%20Software-Atlas%20Deutschland%202012%20Web_tcm17-104249.pdf (2.4.2014)
- Leimeister, S., Böhm, M., Riedl, C., Krcmar, H. (2010): The Business Perspective of Cloud Computing: Actors, Roles and Value Networks. In: *Proceedings of the 2010 European Conference on Information Systems*, Paper 56
- Lerman, J. (2013): Big Data and its Exclusions. In: *Stanford Law Review Online* 66, S. 55–63
- Liao, W., Wang, Y., Pan, E. (2012): Single-machine-based predictive maintenance model considering intelligent machinery prognostics. In: *The International Journal of Advanced Manufacturing Technology* 63(1-4), S. 51–63
- Liebram, C. (2013): Wie Apps auch Hobbysportlern helfen können. www.welt.de/gesundheit/article120835881/Wie-Apps-auch-Hobbysportlern-helfen-koennen.html (20.2.2014)
- Little, G., Segal, R., Huston, C. (2013): Worldwide Cloud Professional Services 2013 Vendor Analysis. http://idcdocserv.com/242401e_IBM (21.2.2014)
- Liu, F., Tong, J., Mao, J., Bohn, R.B., Messina, J.V., Badger, M.L., Leaf, D.M. (2011): NIST Cloud Computing Reference Architecture. Recommendations of the National Institute of Standards and Technology. National Institute of Standards and Technology (NIST), www.disa.mil/Services/DoD-Cloud-Broker/~media/Files/DISA/Services/Cloud-Broker/nist-cloud-ref-architecture.pdf (21.2.2014)



LITERATUR

- Lynch, C. (2009): Jim Gray's Fourth Paradigm and the Construction of the Scientific Record. In: Hey et al. 2009, S. 177–183
- Mainzer, K. (2014): Die Berechnung der Welt. Von der Weltformel zu Big Data. München
- Maizière, T. de (2010): 14 Thesen zu den Grundlagen einer gemeinsamen Netzpolitik der Zukunft. Bundesministerium des Inneren, www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/OED_Verwaltung/Informationsgesellschaft/thesen_netzpolitik.pdf?__blob=publicationFile (3.4.2014)
- Mangelsdorf, J. (2012): Supercomputing the Climate: NASA's Big Data Mission. www.csc.com/cscworld/publications/81769/81773-supercomputing_the_climate_nasa_s_big_data_mission (20.2.2014)
- Manovich, L. (2012): Trending: The Promises and the Challenges of Big Social Data. In: Gold, M.K. (ed.): Debates in the digital humanities. Minneapolis, S. 460–475
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A. (2011): Big data: The next frontier for innovation, competition, and productivity. www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation (21.2.2014)
- Markl, V., Hoeren, T., Krmar, H. (2013): Innovationspotenzialanalyse für die neuen Technologien für das Verwalten und Analysieren von großen Datenmengen (Big Data Management). Finale Studienergebnisse, Bundesministerium für Wirtschaft und Energie, Berlin
- Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., Ghalsasi, A. (2011): Cloud computing — The business perspective. In: Decision Support Systems 51(1), S. 176–189
- Maxwell, W., Wolf, C. (2012): A Global Reality: Governmental Access to Data in the Cloud. A comparative analysis of ten international jurisdictions. [www.hoganlovells.com/files/News/c6edc1e2-d57b-402e-9cab-a7be4e004c59/Presentation/NewsAttachment/a17af284-7d04-4008-b557-5888433b292d/Revised%20Government%20Access%20to%20Cloud%20Data%20Paper%20\(18%20July%202012\).pdf](http://www.hoganlovells.com/files/News/c6edc1e2-d57b-402e-9cab-a7be4e004c59/Presentation/NewsAttachment/a17af284-7d04-4008-b557-5888433b292d/Revised%20Government%20Access%20to%20Cloud%20Data%20Paper%20(18%20July%202012).pdf) (3.4.2014)
- Mayer-Schönberger, V., Cukier, K. (2013): Big data. A revolution that will transform how we live, work and think. London
- McKendrick, J. (2013): »Big data as a service« is here, but is anybody ready? www.zdnet.com/big-data-as-a-service-is-here-but-is-anybody-ready-7000013257/ (20.2.2014)
- Mearian, L. (2014): By 2018, cars will be self-aware. www.computerworld.com/s/article/9245206/By_2018_cars_will_be_self_aware (18.2.2014)
- Mohamed, S., Ismail, O., Hogan, O. (2012): Data Equity. Unlocking the value of Big Data. Centre for Economics and Business Research, www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf (2.4.2014)
- Mohamed, S., Ismail, O., Hogan, O. (2013): Data Equity – Ireland. Unlocking the value of Big Data. Centre for Economics and Business Research, www.cebr.com/wp-content/uploads/2013/06/Cebr_Value-of-BigData_Ireland_lores.pdf (2.4.2014)
- Monroy, M. (2012): »Big Data« auch beim BKA. www.heise.de/tp/artikel/37/37695/1.html (6.3.2014)
- Morozov, E. (2011): The net delusion. The dark side of internet freedom. New York

- Morozov, E. (2013): To save everything, click here. The folly of technological solutionism. New York
- Mowery, D.C. (1996): The international computer software industry. A comparative study of industry evolution and structure. New York
- Müller, B., Egel, J., Höwer, D., Licht, G., Murmann, M. (2012): Hightech-Gründungen in Deutschland. Gründungsdynamik im ITK-Sektor. Zentrum für Europäische Wirtschaftsforschung, http://ftp.zew.de/pub/zew-docs/gutachten/BITKOM_Bericht2012.pdf (3.4.2014)
- Murdoch, T.B., Detsky, A.S. (2013): The Inevitable Application of Big Data to Health Care. In: JAMA 309(13), S. 1351
- Narayanan, A., Shmatikov, V. (2009): De-anonymizing Social Networks. In: Proceedings of the 30th IEEE Symposium on Security and Privacy, S. 173–187
- Neumann, A., Baginski, M., Czub, J. (2010): How Do Sterols Determine the Antifungal Activity of Amphotericin B? Free Energy of Binding between the Drug and Its Membrane Targets. In: Journal of the American Chemical Society 132(51), S. 18266–18272
- Nordemann, W., Nordemann, A., Nordemann, J.B. (2008): Urheberrecht. Kommentar zum Urheberrechtsgesetz, Verlagsgesetz, Urheberrechtswahrnehmungsgesetz. Stuttgart
- North, K. (2010): The NoSQL alternative. Low-cost, high-performance database options make gains. www.informationweek.com/development/architecture-design/the-nosql-alternative/224900559 (20.2.2014)
- NTT Communications (2014): NSA After-shocks. How Snowden has changed ICT decision-makers' approach to the Cloud. www.eu.ntt.com/fileadmin/NTT-Europe/media/PDF/Research_Reports/NTT_Research_Report_NSA_Aftershocks_FINAL_1.pdf (2.4.2014)
- n-tv (2013): Der Kampf ums autonome Fahren. Autobauer bremsen Google aus. www.n-tv.de/auto/Autobauer-bremsen-Google-aus-article11970166.html (18.2.2014)
- OECD (Organisation for Economic Co-operation and Development) (2004): The Economic Impact of ICT. Measurement, Evidence and Implications. <http://browse.oecdbookshop.org/oecd/pdfs/free/9204051e.pdf> (2.4.2014)
- OECD (2013a): New Sources of Growth: Knowledge-Based Capital. Key Analyses and Policy Conclusions. www.oecd.org/sti/inno/knowledge-based-capital-synthesis.pdf (9.4.2014)
- OECD (2013b): Supporting Investment in Knowledge Capital, Growth and Innovation. Paris
- Olofson, C.W. (2012): Managing Big Data: SQL or NoSQL? Technology Assessment. IDC. <http://laser.inf.ethz.ch/2013/material/breitman/additional%20reading/IDC%20Big%20Data%20NoSQL%20or%20SQL%20data.pdf> (21.2.2014)
- OnRamps (o.J.): How does big data work? <https://onramps.instructure.com/courses/723227/wiki/how-does-big-data-work> (20.2.2014)
- Ortega, I. (2013): Big-Data-System oder Data Warehouse? www.computerwoche.de/a/big-data-system-oder-data-warehouse,2516320 (20.2.2014)
- Parent, M. (2013): Automated Vehicles: Autonomous or Connected? In: Proceedings of the 2013 IEEE 14th International Conference on Mobile Data Management, S. 2



LITERATUR

- Pariser, E. (2012): Filter Bubble. Wie wir im Internet entmündigt werden. München
- Pease, A.F. (2013): Das lebende Labor. www.siemens.com/innovation/apps/pof_micro_site/_pof-fall-2013/_html_de/seestadt-aspern.html (20.2.2014)
- Peil, F. (2013): Computermodelle: So wollen Forscher Revolutionen vorhersagen. www.spiegel.de/netzwelt/web/big-data-extremereignisse-mittels-statistik-vorhersagen-a-912347.html (26.2.2014)
- Pélissié du Rausas, M., Manyika, J., Hazan, E., Bughin, J., Chui, M., Said, R. (2011): Internet matters: The Net's sweeping impact on growth, jobs, and prosperity. McKinsey & Company. www.nwoinnovation.ca/upload/documents/mgi-internet-matters-report.pdf (2.4.2014)
- Pentland, A. (2012a): Big Data's Biggest Obstacles. HBR Blog Network, <http://blogs.hbr.org/2012/10/big-datas-biggest-obstacles/> (7.4.2014)
- Pentland, A. (2012b): Reinventing Society in the Wake of Big Data. <http://edge.org/conversation/reinventing-society-in-the-wake-of-big-data> (31.3.2014)
- Pentland, A. (2014): Social physics. How good ideas spread-the lessons from a new science. New York
- Phelps, E.S., Tilman, L.M. (2010): Wanted: A First National Bank of Innovation. In: Harvard Business Review, S. 102–103
- Philip, J. (2013): 20 Most Promising Big Data Companies. CIO Review. Fremont www.cioreview.com/magazine/20-Most-Promising-Big-Data-Companies-PMTK750834712.html (20.2.2014)
- Philippon, T., Véron, N. (2008): Financing Europe's fast movers. In: Bruegel Policy Brief (1), S. 1–3
- Press, G. (2013a): A Very Short History Of Big Data. www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/ (20.2.2014)
- Press, G. (2013b): Top Ten Big Data Pure-Plays. www.forbes.com/sites/gilpress/2013/02/22/top-ten-big-data-pure-plays/ (20.2.2014)
- Proskawetz, K.-O. (2013): Key to Innovation. Cooperative Intelligent Transport Systems and Services (C-ITS). Smart Cities and Communities. Brüssel <http://eu-smartcities.eu/sites/all/files/Cooperative%20Intelligent%20Transport%20Systems%20and%20Services%20-%20Smart%20Cities%20Stakeholder%20Platform%20january.pdf> (8.4.2014)
- Pürner, H.A. (2013): NoSQL – die neue (alte) Datenbank-Generation? www.computerwoche.de/a/nosql-die-neue-alte-datenbank-generation,2497315 (20.2.2014)
- PwC (PricewaterhouseCoopers) (2012): Storing Entertainment Content in the Cloud. www.pwc.ch/user_content/editor/files/publ techno/pwc_storing_entertainment_content_in_the_cloud_e.pdf (21.2.2014)
- Quack, K. (2013): Wie Obama die Wahl gewann. www.computerwoche.de/a/wie-obama-die-wahl-gewann,2539909 (20.2.2014)
- Rajan, S.S. (2012): Big Data Analytics: Thinking Outside of Hadoop. UIMA and Hadoop in Big Data Analytics. Cloud Computing Journal. Woodcliff Lake, <http://cloudcomputing.sys-con.com/node/2349593> (20.2.2014)
- Rider, F. (1944): The Scholar and the Future of the Research Library. A Problem and Its Solution. New York



- Ried, S., Kisker, H., Matzke, P. (2010): The Evolution of Cloud Computing Markets. Forrester. Cambridge <http://fm.sap.com/data/upload/files/forrester%20-%20the%20evolution%20of%20cloud%20computing%20markets.pdf> (21.2.2014)
- Rodrigues, M., Oliveira, A., Costa, M., Fortunato, A.B., Zhang, Y. (2009): Sensitivity analysis of an ecological model applied to the Ria de Aveiro. In: *Journal of Coastal Research* (56), S. 448–452
- Roop, E.S. (2011): Can IBM's Watson Beat the Odds? In: *For The Record* 23(16), S. 14
- Ross, J.W., Beath, C.M., Quaadgras, A. (2013): You May Not Need Big Data After All. In: *Harvard Business Review* 91(12), S. 90–98
- Rotman, D. (2013): The Difference Between Makers and Manufacturers. www.technologyreview.com/review/508821/the-difference-between-makers-and-manufacturers (31.3.2014)
- Russom, P. (2011): Big Data Analytics. TDWI Best Practices Report. Fourth Quarter 2011. TDWI Research. Renton
- SAS (2012): Analytics as a Service. Customer Experiences. www.sas.com/offices/europe/uk/resources/brochure/aaas_research_brief.pdf (21.2.2014)
- SAS (2013): Data Visualization: Making Big Data Approachable and Valuable. www.sas.com/resources/asset/SASCIOMarketPulseDataVizWhitePaper.pdf (21.2.2014)
- Schäfer, A., Knapp, M., May, M., Voß, A. (2012): Big Data – Vorsprung durch Wissen. Innovationspotentialanalyse. Fraunhofer IAIS, www.iais.fraunhofer.de/fileadmin/user_upload/Abteilungen/KD/uploads_BDA/Innovationspotentialanalyse_BigData_FraunhoferIAIS_2012.pdf (21.2.2014)
- Schirmmayer, F. (2013): Wir wollen nicht. Im Zeitalter von Big Data. *Frankfurter Allgemeine Zeitung*. Frankfurt a.M. www.faz.net/aktuell/feuilleton/debatten/ueberwachung/im-zeitalter-von-big-data-wir-wollen-nicht-12545592.html (20.2.2014)
- Schleife, K., Reichhart, J., Laimach, T., Kulicke, M. (2012): Analyse von Wachstumshemmnissen kleiner und mittlerer Unternehmen am Beispiel der IT-Branche. Pierre Audoin Consultants (PAC); Fraunhofer ISI, www.bmwi.de/BMWi/Redaktion/PDF/Publikationen/Studien/analyse-von-wachstumshemmnissen-kleiner-u-mittlerer-unternehmen-it-branche-endbericht,property=pdf,bereich=bmwi2012,sprache=de,rwb=true.pdf (3.4.2014)
- Schmoch, U. (2007): Double-boom cycles and the comeback of science-push and market-pull. In: *Research Policy* 36(7), S. 1000–1015
- Schofield, J. (2012): Personal cloud to replace PC by 2014, says Gartner. *ZDNet*. München, www.zdnet.com/personal-cloud-to-replace-pc-by-2014-says-gartner-4010025617/ (20.2.2014)
- Schrader, C. (2011): Explosion des Cyberspace. *Süddeutsche Zeitung*. München www.sueddeutsche.de/digital/datenwachstum-der-digitalisierten-welt-explosion-des-cyberspace-1.1058394 (20.2.2014)
- Schrage, M. (2012): Was Manager an Big Data nicht verstehen. *Harvard Business Manager*, www.harvardbusinessmanager.de/meinungen/artikel/a-862658.html (20.2.2014)
- Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P. (2012): Analytics: Big Data in der Praxis. Wie innovative Unternehmen ihre Datenbestände effektiv nutzen. IBM, www-935.ibm.com/services/de/gbs/thoughtleadership/GBE03519-DE-00.pdf (21.2.2014)



LITERATUR

- Schubert, L., Jeffery, K. (2012): Advances in clouds. Research in future cloud computing. Expert group report, Europäische Kommission, Brüssel
- Schubert, L., Jeffery, K., Neidecker-Lutz, B. (2010): The future of cloud computing. Opportunities for European cloud computing beyond 2010. Expert group report, Europäische Kommission, Brüssel
- Schürmann, H. (2013): Frühwarnsysteme für Turbinen und Tomographen. In: Pictures of the Future, S. 80–82
- Sellars, S., Nguyen, P., Chu, W., Gao, X., Hsu, K.-l., Sorooshian, S. (2013): Computational Earth Science: Big Data Transformed Into Insight. In: Eos 94(32), S. 277–278
- Simonite, T. (2013): If Facebook Can Profit from Your Data, Why Can't You? www.technologyreview.com/news/517356/if-facebook-can-profit-from-your-data-why-cant-you (1.4.2014)
- Simonite, T. (2014): Sell Your Personal Data for \$8 a Month. MIT Technology Review. www.technologyreview.com/news/524621/sell-your-personal-data-for-8-a-month (1.4.2014)
- sim^{TD}-Konsortium (2013): Feldversuch im Rahmen des Forschungsprojektes sim^{TD} belegt: Car-to-X Kommunikation ist reif für den Alltagseinsatz. www.simtd.de/index.dhtml/object.media/deDE/8033/CS/-/news/Presse/simTD-Pressemitteilung_2013_DE.pdf (18.2.2014)
- Singh, J. (2011): NoSQL databases and MapReduce. Technical Skills Share Group. www.slideshare.net/j_singh/nosql-and-mapreduce
- Smith, J.E., Nair, R. (2005): The architecture of virtual machines. In: Computer 38(5), S. 32–38
- Solow, R.M. (1987): We'd better watch out. In: The New York Times Book Review, S. 36
- Spiceworks (2013): State of SMB IT 1H 2013. Semi-Annual Report on Small and Midsize Business Technology Plans & Purchase Intent. www.spiceworks.com/marketing/state-of-smb-it/ (21.2.2014)
- Spiegel Online (2014): Diabetes-Therapie: Google entwickelt smarte Kontaktlinse. Hamburg www.spiegel.de/netzwelt/gadgets/kontaktlinse-bald-klebt-google-direkt-auf-dem-auge-a-944009.html (20.2.2014)
- Stahl, L.-F. (2013): Gefahr im Kraftwerk. Industrieanlagen schutzlos im Internet. In: c't (11), S. 78–82
- Steadman, I. (2013): Watson is better at diagnosing cancer than human doctors. Wired. www.wired.co.uk/news/archive/2013-02/11/ibm-watson-medical-doctor (10.3.2014)
- Steinmueller, W.E. (2004): The European software sectoral system of innovation. In: Malerba, F. (ed.): Sectoral systems of innovation. Concepts, issues and analyses of six major sectors in Europe. New York, S. 193–241
- Stellermann, R., Haussmann, S., Schwalm, S. (2011): Verwaltungsmodernisierung auf Bundesebene. Studie zum Stand der Modernisierung der Bundesverwaltung. Steria Mummert Consulting; BearingPoint, www.bearingpoint.com/de-de/download/Reformstand_Bundesverwaltung_Verwaltungsmodernisierung.pdf (11.3.2014)

- Stobbe, L., Nissen, N.F., Proske, M., Middendorf, A., Schlomann, B., Friedewald, M. et al. (2009): Abschätzung des Energiebedarfs der weiteren Entwicklung der Informationsgesellschaft. Fraunhofer IZM, Fraunhofer ISI, www.bmwi.de/Dateien/BMWi/PDF/abschaetzung-des-energiebedarfs-der-weiteren-entwicklung-der-informationsgesellschaft,property=pdf,bereich=bmwi,sprache=de,rwb=true.pdf (1.4.2014)
- Stockburger, C. (2013): Car-to-X-Kommunikation: Wahrsager an Bord. www.spiegel.de/auto/aktuell/car-to-x-kommunikation-mit-simtd-a-906985.html (10.3.2014)
- Stonebraker, M., Madden, S., Abadi, D.J., Harizopoulos, S., Hachem, N., Helland, P. (2007): The end of an architectural era (It's time for a complete rewrite). In: Koch, C. (ed.): Proceedings of the 33rd International Conference on Very Large Data Bases. New York, S. 1150–1160
- Sun, L., Axhausen, K.W., Lee, D.-H., Cebrian, M. (2013): Efficient detection of contagious outbreaks in massive metropolitan encounter networks <http://arxiv.org/abs/1401.2815v1> (21.2.2014)
- TAB (Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag) (2012): Gesetzliche Regelungen für den Zugang zur Informationsgesellschaft (Beckert, B., Riehm, U.). TAB-Arbeitsbericht Nr. 149, Berlin
- TCS (TATA Consultancy Services) (2013): The Emerging Big Returns on Big Data. A TCS 2013 Global Trend Study. www.lesechos-events.fr/data/classes/produit_partenaire/fichier_5183_540.pdf (21.2.2014)
- TechNavio (2012): Global Big Data-as-a-service Market 2012–2016. www.technavio.com/report/global-big-data-service-market-2012-2016 (20.2.2014)
- Tene, O., Polonetsky, J. (2013): Big Data for All: Privacy and User Control in the Age of Analytics. In: Northwestern Journal of Technology and Intellectual Property 11(5), Paper 1
- The Economist (2010a): Agents of change. www.economist.com/node/16636121 (2.4.2014)
- The Economist (2010b): All too much. www.economist.com/node/15557421 (20.2.2014)
- The Economist (2010c): Data, data everywhere. www.economist.com/node/15557443 (20.2.2014)
- The Economist (2010d): New rules for big data. www.economist.com/node/15557487 (20.2.2014)
- The Economist (2013a): Don't even think about it. It is getting easier to foresee wrongdoing and spot likely wrongdoers. www.economist.com/news/briefing/21582042-it-getting-easier-foresee-wrongdoing-and-spot-likely-wrongdoers-dont-even-think-about-it (20.2.2014)
- The Economist (2013b): Trouble at the lab. Scientists like to think of science as self-correcting. To an alarming degree, it is not. www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble (7.4.2014)
- The Royal Society (2012): Science as an open enterprise. The Royal Society Science Policy Centre report 02/12. London
- Thomas, J.J., Cook, K.A. (2005): Illuminating the path. The Research and Development Agenda for Visual Analytics. Los Alamitos



LITERATUR

- Thomas, J.J., Cook, K.A. (2006): A visual analytics agenda. In: IEEE Computer Graphics and Applications 26(1), S. 10–13
- Tibco (2012): Volume, velocity, variety, value. Delivering The Elusive Fourth V Out Of Big Data. http://resources.idgenterprise.com/original/AST-0077653_TIBCO_Ebook_Final__Big_Data_print_versionv4.pdf (21.2.2014)
- Tilton, S. (2011): Harness the Power of Data Visualization to Transform Your Business. IT Business Edge, SAS, www.tentonmarketing.com/Portfolio/SAS-Data-Visualization_WP.pdf (21.2.2014)
- TNS Infratest (2012): Quo vadis Big Data. Herausforderungen – Erfahrungen – Lösungsansätze. München
- Transparency Market Research (2012): Big Data Market – Global Scenario, Trends, Industry Analysis, Size, Share And Forecast 2012–2018. Free Market Analysis. www.transparencymarketresearch.com/big-data-market.html (20.2.2014)
- T-Systems (2012): Wenn 52 Millionen Autos keinen Stau produzieren. In: Best Practice (3), S. 14
- Turlea, G., Nepelski, D., Prato, G. de, Lindmark, S., Panizza, A. de, Picci, L. et al. (2010): The 2010 report on R&D in ICT in the European Union. Europäische Kommission, <http://ftp.jrc.es/EURdoc/JRC57808.pdf> (3.4.2014)
- Turlea, G., Nepelski, D., Prato, G. de, Simon, J.-P., Sabadash, A., Stancik, J. et al. (2011): The 2011 report on R&D in ICT in the European Union. Europäische Kommission, <http://ftp.jrc.es/EURdoc/JRC65175.pdf> (3.4.2014)
- Ulbricht, C. (2012a): Big Data & Recht – Herausforderungen für den Datenschutz und die Causa O2. Recht 2.0. Karlsruhe www.rechtzweinnull.de/archives/554-Big-Data-Recht-Herausforderungen-fuer-den-Datenschutz-und-die-Causa-O2.html (20.2.2014)
- Ulbricht, C. (2012b): Big Data & Recht – Wem »gehören« eigentlich Daten oder wie Google nun gegen SEO-Tools vorgeht. Recht 2.0. Karlsruhe www.rechtzweinnull.de/archives/623-Big-Data-Recht-Wem-gehoren-eigentlich-Daten-oder-wie-Google-nun-gegen-SEO-Tools-vorgeht.html (20.2.2014)
- Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (2010): 10 Jahre Safe Harbor – viele Gründe zum Handeln, kein Grund zum Feiern. Kiel, <https://www.datenschutzzentrum.de/presse/20100723-safe-harbor.htm> (3.4.2014)
- Upbin, B. (2013): IBM's Watson Gets Its First Piece Of Business In Healthcare. www.forbes.com/sites/bruceupbin/2013/02/08/ibms-watson-gets-its-first-piece-of-business-in-healthcare/ (20.2.2014)
- Vande Moere, A., Tomitsch, M., Wimmer, C., Christoph, B., Grechenig, T. (2012): Evaluating the Effect of Style in Information Visualization. In: IEEE Transactions on Visualization and Computer Graphics 18(12), S. 2739–2748
- Vaquero, L.M., Rodero-Merino, L., Caceres, J., Lindner, M. (2008): A break in the clouds. In: ACM SIGCOMM Computer Communication Review 39(1), S. 50
- Velten, C., Janata, S. (2012a): Cloud Vendor Benchmark 2012. Cloud Computing Anbieter im Vergleich. Deutschland. Experton Group, www.it-strategie-tage.de/uploads/media/Experton_Cloud_Vendor_Benchmark_2012_Bericht_150512_delivery.pdf (13.11.2013)

- Velten, C., Janata, S. (2012b): Datenexplosion in der Unternehmens-IT. Wie Big Data das Business und die IT verändert. Experton Group, www.cio.de/filesserver/idg/wpcontent/495-datenexplosion-in-der-unternehmens-it.pdf (21.2.2014)
- Vesset, D., Morris, H.D., Little, G., Borovick, L., Feldman, S., Eastwood, M. et al. (2012): Worldwide Big Data Technology and Services 2012–2015 Forecast. Market Analysis. IDC, Framingham
- Veugelers, R., Cincera, M. (2010): Europe's missing yollies. Bruegel Policy Brief 2010/06
- Veugelers, R., Pottelsberghe, B. van, Véron, N. (2012): Lessons for ICT Innovative Industries: Three Experts' Positions on Financing, IPR and Industrial Ecosystems. Europäische Kommission, <http://ftp.jrc.es/EURdoc/JRC76458.pdf> (3.4.2014)
- Viola, G. (2013): Behörden können Stimmungslagen erkennen & Daten sicher analysieren. www.egovernment-computing.de/fachanwendungen/articles/406067 (26.2.2014)
- Violino, B. (2011): Preparing for the real costs of cloud computing. www.computerworld.com/s/article/359383/The_Real_Costs_of_Cloud_Computing (20.2.2014)
- Vries, M. de, Deijs, M., Canuti, M., Schaik, B.D. van, Faria, N.R., Garde, M.D. van de et al. (2011): A Sensitive Assay for Virus Discovery in Respiratory Clinical Samples. In: PLoS ONE 6(1), S. e16118
- Wakeman, N. (2012): Sandy makes the case for big data. <http://washingtontechnology.com/blogs/editors-notebook/2012/11/sandy-big-data-opportunity.aspx> (20.2.2014)
- Walker-Morgan, D.J. (2010): NoSQL im Überblick. heise online. <http://heise.de/1012483> (20.2.2014)
- Watson, H.J., Wixom, B.H. (2007): The Current State of Business Intelligence. In: Computer 40(9), S. 96–99
- Wei, Y., Blake, M.B. (2010): Service-oriented computing and cloud computing: Challenges and opportunities. In: IEEE Internet Computing 14(6), S. 72–75
- Weichert, T. (2013): Big Data und Datenschutz. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, <https://www.datenschutzzentrum.de/bigdata/20130318-bigdata-und-datenschutz.pdf>
- Wessling, C. (2013): Wurm im Werk. www.heise.de/tr/artikel/Wurm-im-Werk-1818812.html (3.4.2014)
- Wessner, C.W. (2009): An assessment of the SBIR program at the National Aeronautics and Space Administration. Washington
- Wieder, P., Butler, J.M., Theilmann, W., Yahyapour, R. (eds.) (2011): Service level agreements for cloud computing. New York
- Willumsen, H. (1999): Folkeregistreringen 75 år i 1999. Kopenhagen
- Wilson, D.D. (1993): Assessing the impact of information technology on organizational performance. In: Banker, R.D., Kauffman, R.J., Mahmood, M.A. (eds.): Strategic information technology management. Perspectives on organizational growth and competitive advantage. Harrisburg, S. 471–514
- Wolfgang Martin Team & IT Research (2012): Big Data. www.it-daily.net/downloads/Bulletin_BIGData_240712_final.pdf (21.2.2014)



LITERATUR

- Wolfgang Martin Team & IT Research (2013): Analytische Datenbanken. Trends in Data Warehousing und Analytik. www.it-daily.net/downloads/Bulletin_AnalytischeDB_final.pdf (21.2.2014)
- Wyld, D.C. (2010): The cloudy future of Government IT: Cloud Computing and public sector around the world. In: *International Journal of Web & Semantic Technology* 1(1), S. 1–20
- Yahyapour, R., Wieder, P. (2013): Cloud-Dienste für die Wissenschaft. Max-Planck-Gesellschaft, www.mpg.de/6957705/JB_2013 (21.2.2014)
- Youseff, L., Butrico, M., Da Silva, D. (2008): Toward a unified ontology of cloud computing. In: *Proceedings of the Grid Computing Environments Workshop*. Los Alamitos, S. 1–10
- Zhang, Q., Cheng, L., Boutaba, R. (2010): Cloud computing: State-of-the-art and research challenges. In: *Journal of Internet Services and Applications* 1(1), S. 7–18

ANHANG

TABELLENVERZEICHNIS 1.

Tab. II.1	Technische Grundlagen von Business Intelligence und Big Data	44
Tab. III.1	Erwartete Umsätze am Markt für Big Data	64
Tab. III.2	Erwartete Umsätze am Markt für Cloud Computing	72

ABBILDUNGSVERZEICHNIS 2.

Abb. II.1	Veranschaulichung des CAP-Theorems	46
Abb. II.2	Gängige Architektur von Cloudinfrastrukturen auf Anbieterseite	59
Abb. III.1	Umsätze mit Dienstleistungen, Soft-und Hardware am Markt für Big Data	65
Abb. III.2	Umsätze am Markt für Big Data nach Regionen	65
Abb. III.3	Umsätze mit SaaS, PaaS und IaaS am Markt für Cloud Computing	73
Abb. IV.1	Durchführung von Big-Data-Analysen nach Regionen	80
Abb. IV.2	Nutzung von Big-Data-Analysen nach Fachbereichen	81
Abb. IV.3	Bei Big-Data-Analysen verwendete Datentypen	82
Abb. IV.4	Bei Big-Data-Analysen verwendete Technologien	84
Abb. IV.5	Nutzung von Clouddiensten nach Branche	93
Abb. IV.6	Nutzung von SaaS-Diensten im Rahmen von Public-Cloud-Angeboten	93
Abb. IV.7	Nutzung von Cloud Computing in Unternehmen und der öffentlichen Verwaltung	95



**BÜRO FÜR TECHNIKFOLGEN-ABSCHÄTZUNG
BEIM DEUTSCHEN BUNDESTAG**

KARLSRUHER INSTITUT FÜR TECHNOLOGIE (KIT)

Neue Schönhauser Straße 10
10178 Berlin

Fon +49 30 28491-0
Fax +49 30 28491-119

buero@tab-beim-bundestag.de
www.tab-beim-bundestag.de