

Nutzerorientiertes Energiedatenmanagement

Zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften

(Dr.-Ing.)

von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Wi.-Ing. Fabian Rigoll

Tag der mündlichen Prüfung: 07. März 2017

Referent: Prof. Dr. Hartmut Schmeck
Korreferent: Prof. Dr. Achim Streit



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung –
Weitergabe unter gleichen Bedingungen 4.0 International Lizenz (CC BY-SA 4.0):
<https://creativecommons.org/licenses/by-sa/4.0/deed.de>

Kurzfassung

Die geschickte Erfassung und Nutzung von Energiedaten haben im Kontext der Energiewende das Potential, das zukünftige Energiesystem intelligenter und effizienter zu machen. Gleichzeitig muss jedoch eine Reihe technischer Herausforderungen beim Umgang mit Energiedaten bewältigt werden. Darüber hinaus sind Energiedaten oft personenbezogene Daten, welche die Privatsphäre der Dateneigner gefährden können.

In dieser Dissertation wird daher ein Konzept für ein Energiedatenmanagementsystem vorgestellt, welches sowohl technische als auch nutzergetriebene Anforderungen erfüllt. Das System unterstützt die Dateneigner beim Schutz ihrer Privatsphäre, ohne jedoch die Nutzbarkeit der Daten im Sinne der betroffenen Personen zu stark einzuschränken. Im Rahmen der vorliegenden Arbeit werden hierfür zunächst die Anforderungen erarbeitet, die ein Energiedatenmanagementsystem erfüllen sollte. Die Anforderungsanalyse gliedert sich in zwei Bereiche. Zum einen wird eine technische Datenlebenszyklusanalyse durchgeführt und zum anderen werden die Interessen der von der Energiedatenerfassung betroffenen Personen untersucht. Das Ergebnis ist ein Anforderungskatalog für Energiedatenmanagementsysteme. Darauf folgend wird ein modulares Energiedatenmanagementsystem entworfen, welches die Daten im Sinne der Dateneigner verwaltet und es diesen ermöglicht, feingranular über die Weitergabe ihrer Daten zu entscheiden. Nach der Spezifikation des Konzepts wird dieses im Rahmen eines Demonstrators prototypisch umgesetzt. An den Energiedaten interessierte Parteien können bei Verwendung eines solchen Systems nicht direkt auf die Daten zugreifen. Vielmehr wird ihnen die Möglichkeit geboten, Datenanfragen bei den Dateneignern einzureichen. Die Anfragen werden durch das Energiedatenmanagementsystem im Hinblick auf die potenzielle Gefährdung der Privatsphäre geprüft. Das System nimmt hierfür eine quantitative Bewertung des Gefährdungspotentials vor und präsentiert diese den betroffenen Personen sowohl in numerischer als

auch in farblich codierter Form. Die Dateneigner können daraufhin über die Weitergabe der Daten in der angefragten Aussagekraft entscheiden. Insbesondere erlaubt das System den Dateneignern, die Aussagekraft der Daten auf ein für sie akzeptables Maß zu reduzieren. Dieser Vorgang erfolgt interaktiv durch Verhandlungen zwischen Dateneigner und -interessent. Eine tatsächliche Datenweitergabe erfolgt schließlich erst nach einer ausdrücklichen Zustimmung durch die betroffenen Personen und zu den vereinbarten Konditionen. Das entworfene Konzept wird im Hinblick auf den Anforderungskatalog evaluiert. Anschließend wird geprüft, ob das System in der Lage ist, für typische Anwendungsfälle die Datenbasis zu liefern und gleichzeitig die Privatsphäre der betroffenen Personen zu schützen.

Danksagung

An dieser Stelle möchte ich all denjenigen meinen Dank aussprechen, die mich bei der Erstellung dieser Arbeit begleitet und unterstützt haben. Zunächst bedanke ich mich bei meinem Doktorvater, Prof. Dr. Hartmut Schmeck, für die Betreuung dieser Arbeit, das entgegengebrachte Vertrauen und die gewährte wissenschaftliche Freiheit. Darüber hinaus danke ich Prof. Dr. Achim Streit für die bereitwillige Übernahme des Korreferats.

Diese Arbeit entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB) des Karlsruher Instituts für Technologie (KIT). Dementsprechend gebührt auch meinen Kollegen der Forschungsgruppe „Effiziente Algorithmen“ großer Dank. Ich habe unseren intensiven und produktiven Austausch stets als große Bereicherung empfunden.

Schließlich möchte ich mich noch aus tiefstem Herzen bei meiner gesamten Familie bedanken. Ich schätze mich sehr glücklich, dass ich mich auf Euren Rückhalt und Eure bedingungslose Unterstützung immer verlassen kann.

Karlsruhe, März 2017

Fabian Rigoll

Inhaltsverzeichnis

Tabellenverzeichnis	xi
Abbildungsverzeichnis	xiii
1. Einleitung	1
1.1. Energiedatenmanagement für die Energiewende	3
1.2. Grundlegende Annahmen	5
1.3. Wesentliche Beiträge der Arbeit	7
1.3.1. Analyse der Anforderungen an Energiedatenmanagementsysteme	7
1.3.2. Konzeption eines Energiedatenmanagementsystems . .	8
1.4. Anmerkung zu eigenen Vorarbeiten	9
1.5. Aufbau der Arbeit	10
2. Grundlagen	11
2.1. Die Energiewende	11
2.1.1. Gründe für die Energiewende	13
2.1.2. Herausforderungen der Energiewende	18
2.1.3. Notwendigkeit der Energiewende	24
2.2. Energiedaten	25
2.2.1. Energieverbrauch und -produktion	25
2.2.2. Definition Energiedaten	26
2.2.3. Nutzen von Energiedaten	29
2.2.4. Gefahren bei der Verwendung von Energiedaten . . .	34
2.3. Energiedatenmanagement	35
2.3.1. Definition Energiedatenmanagement	35
2.3.2. Nutzer von Energiedatenmanagementsystemen	36

3. Verwandte Arbeiten	39
3.1. Umgang mit Energiedaten und deren Management	39
3.1.1. In der Forschung	40
3.1.2. In der Wirtschaft	42
3.2. Schutz von Energiedaten	49
3.2.1. Regulatorische Ansätze	49
3.2.2. Technologische Ansätze	54
3.2.3. Zusammenfassung	63
4. Anforderungsanalyse	65
4.1. Technische Datenlebenszyklusanalyse	65
4.1.1. Der Datenlebenszyklus als Werkzeug	66
4.1.2. Phase I: Erfassung	68
4.1.3. Phase II: Übertragung	80
4.1.4. Phase III: Speicherung	86
4.1.5. Phase IV: Analyse	92
4.1.6. Phase V: Verteilung	95
4.1.7. Phase VI: Löschung	98
4.1.8. Fazit zur technischen Datenlebenszyklusanalyse . . .	100
4.2. Privatsphäre und Energiedaten	101
4.2.1. Begriffsabgrenzung	102
4.2.2. Verfahren zur Extraktion von Informationen aus Ener-	
giedaten	102
4.2.3. Konkrete Auswirkungen auf die Privatsphäre	111
4.2.4. Resultierende Anforderungen	114
4.3. Weitere Nutzerinteressen	117
4.4. Zusammenfassung der Anforderungsanalyse	119
5. Konzept und Umsetzung	121
5.1. Architektur des Data Custodian Service	121
5.2. Datenmodell des Data Custodian Service	127
5.2.1. Beschreibung des Datenmodells	128
5.2.2. Anmerkungen zum Datenmodell	131
5.3. Umsetzung des Konzepts in einem Demonstrator	134
5.3.1. Zugrundeliegendes Softwaregerüst	134

5.3.2. Verarbeitung von Zeitreihen	144
5.3.3. Speicherung der Zeitreihen	148
5.3.4. Bewertung der Datenanfragen	151
5.3.5. Reduktion der Aussagekraft der Daten	168
5.3.6. Logbuchfunktion	174
5.3.7. Löschung historischer Daten	175
5.3.8. Zusammenfassung Demonstrator	176
5.4. Typische Arbeitsabläufe	178
5.4.1. Verfügbare Aktionen	179
5.4.2. Verwaltung der eigenen Energiedaten	179
5.4.3. Bereitstellung neuer Energiedaten	181
5.4.4. Erstellung neuer Datenanfragen	182
5.4.5. Verwaltung eingegangener Datenanfragen	185
5.4.6. Verwaltung eingereichter Anfragen	189
5.4.7. Überführung in ein Produkt für den Realwelteinsatz .	192
5.5. Fazit zum Konzept und zum Demonstrator	195
6. Evaluation	197
6.1. Prüfung der Anforderungen	197
6.1.1. Unterstützung unterschiedlicher Datenquellen	198
6.1.2. Robustheit gegenüber fehlenden und fehlerhaften Daten	200
6.1.3. Lokale Sammelstelle unter Kontrolle des Nutzers . . .	201
6.1.4. Transparenz der Datenübertragungswege	202
6.1.5. Unterstützung einer geeigneten Speicherform in Ab- hängigkeit der Datenart	203
6.1.6. Verschlüsselung bei externer Speicherung	204
6.1.7. Lokale Vorverarbeitung der Daten	205
6.1.8. Unterstützung unterschiedlicher Datenausgabeformate	207
6.1.9. Sicherheit der Datenübertragungswege	208
6.1.10. Löschung nicht mehr benötigter Daten	209
6.1.11. Einschränkung der Datenweitergabe	210
6.1.12. Reduktion der Aussagekraft auf akzeptables Mindest- maß vor Weitergabe	211
6.1.13. Aufklärung über mögliche Auswirkungen einer Daten- freigabe auf die Privatsphäre	212

6.1.14. Unterstützung zusätzlicher, externer Schutzmechanismen	213
6.1.15. Unterstützung verschiedener Endgeräte zur Darstellung der Benutzeroberfläche	214
6.1.16. Zusammenfassung zur Erfüllung der Anforderungen .	214
6.2. Datenbereitstellung für typische Anwendungsfälle	215
6.2.1. Anwendungsfall 1: Aufzeigen der größten Verbrauchsposten	216
6.2.2. Anwendungsfall 2: Bestimmung der Grundlast in einem Haushalt	217
6.2.3. Anwendungsfall 3: Auswahl des richtigen zeitvariablen Tarifs	219
6.2.4. Zusammenfassung Anwendungsfälle	221
6.3. Weitergehende Evaluationsmöglichkeiten	221
7. Zusammenfassung und Ausblick	223
7.1. Zusammenfassung der Arbeit	223
7.1.1. Zusammenfassung der Anforderungsanalyse	224
7.1.2. Zusammenfassung zum entwickelten Konzept und zu dessen Umsetzung	226
7.2. Einbettung in den Stand der Wissenschaft	230
7.3. Ausblick	233
A. Umsetzung des Datenmodells	235
Literaturverzeichnis	239

Tabellenverzeichnis

2.1. Interpretationen des Begriffs Energiedaten	28
4.1. Typische physikalische Größen und Einheiten im Kontext von Energiedaten	75
4.2. Übersicht über die resultierenden Anforderungen der Anforderungsanalyse	120
5.1. Gefährdungskennzahl in Abhängigkeit des Datenquellentyps	164
5.2. Gewichtung der einzelnen Gefährdungskennzahlen	166
5.3. Wählbare zeitliche Auflösungen für eine Datenanfrage	170
5.4. Wählbare Verrauschungsstärken	172
6.1. Übersicht über die resultierenden Anforderungen der Anforderungsanalyse (Wdh.)	198

Abbildungsverzeichnis

4.1. Datenlebenszyklusmodell wie in <i>LSDMA</i> verwendet	67
4.2. Mögliche Datenübertragungswege von Energiedaten	83
5.1. Architektur des <i>Data Custodian Service</i>	122
5.2. Das Entity-Relationship-Modell des <i>Data Custodian Service</i> . .	128
5.3. Gefährdungskennzahl in Abhängigkeit der zeitlichen Auflösung (maximal ein Jahr)	155
5.4. Gefährdungskennzahl für die zeitliche Auflösung (maximal 24 Stunden)	156
5.5. Gefährdungskennzahl für die zeitliche Auflösung (maximal 60 Minuten)	157
5.6. Gefährdungskennzahl für den angefragten Zeitraum (maximal zwei Jahre)	159
5.7. Gefährdungskennzahl für den angefragten Zeitraum (maximal 24 Stunden)	160
5.8. Gefährdungskennzahl für das Datenalter (maximal drei Jahre)	162
5.9. Gefährdungskennzahl für das Datenalter (maximal über vier Wochen)	163
5.10. Verwendete Bibliotheken im <i>Data Custodian Service</i>	177
5.11. Startseite für Nutzer	180
5.12. Startseite für Dateninteressenten	180
5.13. Übersicht zu den Energiedaten eines Nutzers	181
5.14. Detailansicht zu Energiedaten eines Nutzers	182
5.15. Bereitstellung neuer Energiedaten	183
5.16. Erstellung einer neuen Datenanfrage	184
5.17. Übersicht über eingegangene Datenanfragen	186
5.18. Detailansicht für eine eingegangene Datenanfrage (Nutzer) .	187
5.19. Detailansicht für eine eingegangene Datenanfrage 2 (Nutzer) .	188

Abbildungsverzeichnis

5.20. Übersicht über eingegangene Datenanfragen 2	189
5.21. Übersicht über eingereichte Datenanfragen	190
5.22. Detailansicht für eine eingereichte Datenanfrage (Dateninter- essent)	191
5.23. Detailansicht für eine eingereichte Datenanfrage 2 (Dateninter- essent)	191

Einleitung

Tag für Tag nutzen wir eine Vielzahl elektrischer Geräte, ohne bewusst darüber nachzudenken, wie viel Energie für deren Betrieb überhaupt aufgewendet werden muss. In Ländern mit zuverlässiger Stromversorgung, wie es in Deutschland der Fall ist, steht heute elektrische Energie für gewöhnlich immer auf Abruf zur Verfügung. Strom scheint einfach aus der Steckdose zu kommen und Stromausfälle sind die Ausnahme. Diese unbewusste Nutzung elektrischer Energie führt dazu, dass der damit verbundene energetische Aufwand oft massiv unterschätzt wird. Setzt man den Aufwand in Relation zur Arbeit, die ein Mensch mit Muskelkraft leisten kann, wird dies deutlich. Ein durchschnittlich trainierter Mensch kann dauerhaft etwa einhundert Watt leisten. Für den Betrieb eines modernen Fernsehgerätes müssten zwei Menschen zur Stromerzeugung beitragen. Das Zubereiten der Frühstückstoasts benötigt hingegen schon die gesamte Leistung eines Dutzends menschlicher Energieerzeuger. Eine Waschmaschine braucht während ihres Betriebs sogar die Leistung von rund zwanzig Personen. All diese Geräte können wir im Alltag jedoch mit einem simplen Knopfdruck einschalten und nutzen, wobei uns der tatsächlich anfallende Aufwand für gewöhnlich verborgen bleibt. Die meisten der heute lebenden Menschen nutzen auf diese Weise während ihres Lebens enorme Mengen von Energie. Erschwerend kommt hinzu, dass auch die Weltbevölkerung mit steigendem Tempo wächst. Schon jetzt liegt der jährliche Zuwachs an Erdenbewohnern pro Jahr bei über 80 Millionen Menschen, was in etwa der Einwohnerzahl Deutschlands entspricht [Sta16].

In der Zeit vor der industriellen Revolution waren sowohl die Bevölkerungszahl als auch der Pro-Kopf-Energiebedarf noch vergleichsweise gering. Fossile Energieträger wie Kohle und Erdöl wurden nicht oder nur in geringem

Kapitel 1. Einleitung

Ausmaß genutzt. Der älteste Brennstoff der Menschheit, das Holz, diente lange Zeit als Hauptwärmelieferant für die meisten Menschen. Im Gegensatz zu fossilen Energieträgern können sich die Holzbestände in relativ kurzer Zeit selbst erneuern und lassen sich daher nachhaltig nutzen. Sie weisen im Vergleich allerdings eine geringere Energiedichte auf, die dem Wachstum des Energieverbrauchs Grenzen setzte. Im Rahmen der industriellen Revolution im späten 18. beziehungsweise 19. Jahrhundert vervielfachte sich der Energiebedarf schlagartig. Neu entwickelte Maschinen bauten nicht mehr nur auf Muskelkraft von Menschen oder Arbeitstieren, sondern vor allem auf Dampf. Der entstehende Energiehunger konnte nicht mehr nur mit nachhaltigen Brennstoffen wie Holz gestillt werden. Stattdessen wurden zunehmend fossile Energieträger benötigt, um die neu gebauten Fabriken mit Energie versorgen zu können. Auch wenn weiterhin Wind- und Wasserkraft in mechanischen Anlagen verwendet wurden, sank deren relativer Anteil zu Gunsten der Nutzung von Dampf jedoch. Die damit einhergehenden Entwicklungen ermöglichen rasante Fortschritte und Wachstum. Der Schutz von Umwelt und Ressourcen spielte jedoch keine Rolle und selbst die Idee der Endlichkeit von fossilen Ressourcen kam erst später auf. Heute sind wir mehr denn je von Energie abhängig. In der Regel nicht mehr in Form von Dampf, sondern in Form von Elektrizität und flüssigen Brennstoffen.

Die einschneidenden Veränderungen der letzten Jahrhunderte beeinflussten nicht nur die Menschen, sondern auch unseren Planeten auf lange Zeit. Mit ihnen wurde ein Weg eingeschlagen, der unumkehrbare Entwicklungen mit sich bringt, die auch noch viele zukünftige Generationen beschäftigen werden. Durch den explodierten Verbrauch haben einige wenige Generationen bereits einen großen Teil der fossilen Energieträger verbrannt, die im Laufe von etlichen Jahrtausenden entstanden sind. In Folge dessen sind ganze Landstriche radikal verändert und zum Teil zerstört worden. Die Effekte der exzessiven Energienutzung sind nicht nur regional begrenzt, sondern haben globale Ausmaße angenommen. Der menschengemachte Klimawandel hat Auswirkungen auf beinahe alle Menschen dieser Welt, aber auch auf die Flora und Fauna unseres Planeten.

Politische Entscheidungen können geändert und widerrufen werden. Der ursprünglich geplante Ausstieg aus der Kernkraft in Deutschland, die anschließende Laufzeitverlängerung und der erneute Ausstieg in Folge der

1.1. Energiedatenmanagement für die Energiewende

Katastrophe von Fukushima sind ein Beispiel dafür. Zurzeit strebt Deutschland eine engagierte Umsetzung der Energiewende an und es besteht Grund zur Hoffnung, dass diese Ziele vorerst nicht abgeschwächt werden. Trotzdem verzichtet diese Arbeit bewusst auf die in verwandten Arbeiten oft übliche Motivation in Form verschiedener Gesetze und politischer Ziele bezüglich der Zukunft unserer Energieversorgung. Denn bereits aus rein rationaler Sicht führt kein Weg an einem Umdenken und einer Umwandlung unseres gesamten Energiesystems vorbei: Es ist lediglich eine Frage der Zeit, bis fossile Brennstoffe unseren Energiebedarf nicht mehr decken können. Die Menschheit ist von Energie abhängig und verlässt sich dennoch nach wie vor auf endliche Ressourcen, die bereits zu einem großen Teil aufgebraucht sind. All dies geschieht, obwohl die Grundvoraussetzungen vorhanden sind, um unseren gesamten Energiebedarf mittelfristig aus erneuerbaren Quellen zu decken und nachhaltig zu wirtschaften. Die Umsetzung der Energiewende sollte daher, vor allem auch im Hinblick auf zukünftige Generationen, ein vorrangiges Ziel der Menschheit sein. Es ist gleichzeitig aber eine Jahrhundertaufgabe, die große, gemeinsame Anstrengungen in vielen Bereichen erfordert [AB07].

1.1. Energiedatenmanagement für die Energiewende

Die Energiewende als Ganzes ist eine komplexe Aufgabe, welche Herausforderungen in den unterschiedlichsten Disziplinen aufwirft. Neben allen technischen Herausforderungen dürfen auch die sozialen und politischen Fragestellungen nicht vernachlässigt werden. So ermöglichen beispielsweise technische Fortschritte effizientere Kraftwerke, welche die erneuerbaren Energien besser nutzen können. Ein verändertes gesellschaftliches Verständnis für die Energienutzung kann aber gleichsam zu einer Senkung des Verbrauchs führen und damit ähnlich zielführend sein. Dementsprechend tragen viele verschiedene Teilespekte zum Erfolg der Energiewende bei. Nur wenn in allen Teilbereichen Fortschritte erzielt werden, lässt sich die Energiewende zügig und effizient umsetzen.

Diese Arbeit leistet einen Beitrag in einem weiteren der vielen Teilgebiete, die die Energiewende vorantreiben: der Energieinformatik. Diese Disziplin

beschäftigt sich unter anderem mit der effizienten Nutzung von Informations- und Kommunikationstechnologien, um ein nachhaltiges Energiesystem zu schaffen. Im Vordergrund stehen dabei zwei Hauptaufgaben. Dies ist zum einen die Steigerung der Energieeffizienz über das hinaus, was alleine mit ingenieurtechnischen Mitteln möglich ist. Zum anderen ist es die effiziente Integration erneuerbarer Energien in unser Energiesystem, indem dieses intelligenter gemacht wird [Goe+14].

Eine systematische Erfassung und eine geschickte Verwendung von Energiedaten können dazu beitragen, das zukünftige Energiesystem intelligenter zu machen. In diesem Kontext umfasst der Begriff Energiedaten unter anderem Daten in Zeitreihenform zur Nutzung von Energie in Haushalten oder auch in Unternehmen. Solche Daten können dabei helfen, ein effizientes Energiesystem aufzubauen und zu betreiben: Sie sind die Basis für einen optimalen Betrieb und einen anforderungsgerechten Ausbau des Energiesystems. Darüber hinaus können sie verwendet werden, um nützliche Mehrwertdienste anzubieten. Sie können Menschen über ihren Energieverbrauch aufklären und sie diesbezüglich positiv beeinflussen.

Energiedaten sind jedoch häufig auch personenbezogene Daten, mit deren Hilfe umfangreiche Rückschlüsse auf das Leben und Verhalten der betroffenen Personen möglich sind. Geraten diese Daten in die Hände unbefugter Dritter, ist eine missbräuchliche Verwendung meist nicht zu verhindern. Die Betroffenen sind sich dieses Dilemmas jedoch in der Regel nicht bewusst. Sie erkennen den Informationsgehalt der aus ihrer Sicht abstrakten Daten nicht und haben dementsprechend weder ein Gespür für die damit verbundene Gefährdung ihrer Privatsphäre, noch für den potenziellen Nutzen der Daten. Gegenwärtig erfahren betroffene Personen diesbezüglich kaum Unterstützung. Erfasste Daten von sogenannten intelligenten Zählern oder intelligenten Zwischensteckdosen werden oft ohne bewusste Entscheidung der Nutzer weitergegeben. Über mögliche Konsequenzen wird nicht aufgeklärt und dem Schutz der Privatsphäre wird nur wenig Gewicht beigemessen. Ein geeignetes Energiedatenmanagementsystem, in welchem die Energiedaten aufbereitet und verwaltet werden, könnte den betroffenen Personen – im Folgenden *Nutzer* genannt – hingegen helfen, die Kontrolle über ihre Daten zu behalten und ihre Privatsphäre zu schützen. Den Nutzern eines Energiedatenmanagementsystems stehen externe Parteien – im Folgenden *Dateninteressenten* genannt –

gegenüber, welche möglichst umfangreichen Zugriff auf die Energiedaten erhalten möchten. Ein geeignetes Energiedatenmanagementsystem kann dabei helfen den Interessenkonflikt zwischen den Nutzern und den Dateninteressenten zu entschärfen.

1.2. Grundlegende Annahmen

Im Folgenden werden einige Annahmen erläutert, auf welchen diese Arbeit beruht.

Annahme (1). *Energiedaten werden hauptsächlich dafür eingesetzt, einen signifikanten Beitrag bei der effizienten Umsetzung der Energiewende zu leisten.*

Der Ausbau und die Nutzung moderner Informations- und Kommunikationstechnologien eröffnen einen Einblick in das Energiesystem, der zuvor nicht möglich war. Energiedaten sollten aber nicht nur um ihrer selbst Willen oder gar entgegen den Nutzerinteressen erfasst werden, sondern dazu dienen, detaillierte Informationen über ein komplexes Energiesystem zu erhalten, um dieses effizient zu betreiben. Feingranulare Informationen über Energiesystemzustände sowie Energieverbrauch und -produktion ermöglichen fundierte Entscheidungen und bilden unter anderem die Basis für verschiedene Optimierungsansätze oder Mehrwertdienste. Mit ihnen wird einer der Grundsteine gelegt, um das gesamte System effizienter zu gestalten und erneuerbare Energien besser zu integrieren. Eine solche Verwendung dürfte daher auch im Sinne der von der Energiedatenerfassung betroffenen Personen sein. Dementsprechend sollte dies auch den Hauptzweck der Energiedaten Nutzung darstellen, sodass betroffene Personen überhaupt ein Interesse haben, ihre Daten zu teilen. Allerdings muss je nach Anwendungsfall abgewogen werden, welche Daten in welcher Aussagekraft tatsächlich für die Erfüllung der jeweiligen Aufgabe erforderlich sind. Insbesondere sollte auch das Prinzip der Datensparsamkeit angemessen berücksichtigt werden.

Annahme (2). *Energiedaten enthalten oft personenbezogene Daten. In der Regel erkennen die betroffenen Personen aber weder den Informationsgehalt der Daten noch die damit verbundenen Gefahren.*

Mit unterschiedlichen Verfahren lassen sich umfangreiche Informationen aus Energiedaten extrahieren. In der jüngeren Vergangenheit entwickelte Verfahren zur Verarbeitung von Energiedaten sind in der Lage, detaillierte Profile zu erstellen, welche das Nutzerverhalten widerspiegeln. Diese abgeleiteten Informationen können für Zwecke verwendet werden, die den Nutzerinteressen entsprechen. Ebenso ist aber ein Missbrauch der Daten möglich. In jedem Fall handelt es sich bei diesen personenbezogenen Daten um ein schützenswertes Gut, über welches die betroffenen Nutzer selbst entscheiden sollten. Insbesondere technisch wenig versierte Personen sind jedoch meist nicht in der Lage, auch nur abzuschätzen, welche Informationen sich aus den sie betreffenden Energiedaten extrahieren ließen. Dementsprechend können sie auch nicht erkennen, welche Gefahren aber auch welchen Nutzen eine Weitergabe der Daten mit sich bringen kann.

Annahme (3). *Ein geeignetes System kann Energiedaten so aufbereiten, dass Nutzer abschätzen können, welchen Informationsgehalt die sie betreffenden Energiedaten aufweisen. Darüber hinaus kann es Nutzer dabei unterstützen, bewusste Entscheidungen über den Umgang mit ihren Energiedaten zu treffen.*

Die Nutzer sind meist nicht in der Lage, ohne weitere Hilfe zu erkennen, welche Informationen sich aus den sie betreffenden Energiedaten ableiten ließen. Ein geeignetes Energiedatenmanagementsystem könnte seine Nutzer diesbezüglich aber unterstützen, indem es die Energiedaten analysiert und abschätzt, welchen Informationsgehalt diese aufweisen. Darüber hinaus kann eine Einschätzung der Auswirkungen auf die Privatsphäre vorgenommen werden. Wenn diese Ergebnisse den Nutzern auf geeignete Art und Weise präsentiert werden, können diese wiederum bewusste Entscheidungen über den Umgang mit ihren Energiedaten treffen. Das Energiedatenmanagementsystem gibt den Nutzern damit die Möglichkeit, potenziellen Nutzen und mögliche Risiken einer Datenweitergabe an Dritte zu beurteilen.

Diese Annahmen bilden den Rahmen für die vorliegende Arbeit. Gleichzeitig stellen sie auch die Motivation für den Entwurf eines nutzerorientierten Energiedatenmanagementsystems dar. Von ihnen ausgehend wurden die wesentlichen Beiträge erarbeitet, welche im Folgenden vorgestellt werden. Darüber hinaus wird im Laufe dieser Arbeit auf die zuvor getroffenen Annah-

men zurückgegriffen und gezeigt, dass sie plausibel sind, beziehungsweise erfüllt werden können.

1.3. Wesentliche Beiträge der Arbeit

Die wesentlichen Beiträge dieser Arbeit gliedern sich in zwei aufeinander aufbauende Teilbereiche: eine eingehende Analyse der Anforderungen an Energiedatenmanagementsysteme und ein darauf basierendes Konzept für ein solches System.

1.3.1. Analyse der Anforderungen an Energiedatenmanagementsysteme

Im Rahmen der Anforderungsanalyse werden Energiedaten, ihre Eigenheiten und die Handhabung dieser Art von Daten eingehend untersucht. Ziel der Untersuchung ist die Ausarbeitung eines Anforderungskatalogs, welcher bei der Konzeption eines Energiedatenmanagementsystems als Grundlage dienen kann.

Einerseits erfolgt eine Datenlebenszyklusanalyse, welche im Speziellen die technischen, energiedatenspezifischen Anforderungen untersucht. Diese Analyse deckt den gesamten Datenlebenszyklus von der Erfassung über die Übertragung, Speicherung, Analyse und Verteilung bis hin zur Löschung ab. Auf diese Weise können aus den Eigenschaften von Energiedaten technische Anforderungen abgeleitet werden, welche ein Energiedatenmanagementsystem erfüllen sollte.

Andererseits wird eine rein technische Analyse den Anforderungen der Nutzer – den von der Energiedatenerfassung betroffenen Personen – nicht gerecht. Deshalb schließt sich eine Untersuchung der Nutzerinteressen an. Im Vordergrund steht hier vor allem der Schutz der Privatsphäre der Nutzer eines Energiedatenmanagementsystems: Während Dateninteressenten möglichst umfangreichen Zugriff auf die Daten wünschen, sollte aus Sicht der von der Datenerfassung Betroffenen der Wahrung von Datenschutz und Privatsphäre besondere Bedeutung beigemessen werden.

Aus dieser ganzheitlichen Analyse ergibt sich eine Reihe von technischen und nutzergetriebenen Anforderungen, welche ein nutzerorientiertes Energiedatenmanagementsystem erfüllen sollte. Dieser Katalog dient als Grundlage für den Entwurf eines entsprechenden Konzepts.

1.3.2. Konzeption eines Energiedatenmanagementsystems

Mithilfe der Ergebnisse aus der Anforderungsanalyse wird ein Konzept für ein geeignetes Energiedatenmanagementsystem erarbeitet. Der entworfene *Data Custodian Service* erfüllt damit nicht nur die technischen Anforderungen, sondern trägt auch gleichsam den Nutzerinteressen Rechnung.

Aus technischer Sicht ist insbesondere die Vielzahl unterschiedlicher Arten von potenziell fehlerbehafteten Energiedaten eine Herausforderung. Das Konzept sieht nicht zuletzt deshalb einen modularen Aufbau des Systems vor, sodass eine flexible Anpassung an veränderte oder neue Anforderungen möglich ist. Das Konzept wurde jedoch nicht nur unter dem Gesichtspunkt technischer Herausforderungen entworfen. Vielmehr stehen die von der Energiedatenerfassung betroffenen Personen – die Nutzer – und deren Interessen im Vordergrund. Aus diesem Grund ist das vorrangige Ziel des *Data Custodian Service* die Unterstützung der Nutzer bei Entscheidungen über die Freigabe ihrer Energiedaten. Das System dient in diesem Kontext als Mittler zwischen Dateninteressenten und betroffenen Personen. Es nimmt von externen Parteien eingehende Datenanfragen entgegen, beurteilt diese und erklärt den Nutzern die potenziellen Auswirkungen einer solchen Datenfreigabe. Die Nutzer können Anfragen zu den vereinbarten Konditionen annehmen oder, gegebenenfalls mit einem Hinweis bezüglich des Grundes, ablehnen, woraufhin die Dateninteressenten wiederum eine veränderte Datenanfrage einreichen können. Auf diese Weise wird ein Verhandlungsprozess ermöglicht, im Rahmen dessen die Aussagekraft der Daten auf ein für die Nutzer akzeptables Maß reduziert wird. Diese Vorgehensweise dient einerseits dem Schutz der Privatsphäre der betroffenen Personen und schränkt andererseits die Nutzbarkeit der Daten so wenig wie möglich ein. Der *Data Custodian Service* stellt damit eine Möglichkeit dar, den Konflikt zwischen Dateninteressenten auf der einen Seite und von der Energiedatenerfassung betroffenen Personen auf der anderen Seite zu entschärfen. Auf diese Weise können Energiedaten

und ihre Verwendung einen Beitrag zur Energiewende leisten, ohne dass die Privatsphäre und der Datenschutz in nicht akzeptablem Maße eingeschränkt werden. Im Verlauf der Arbeit wird insbesondere auch die Umsetzung des entworfenen Konzepts in Form einer *Proof-of-Concept*-Implementierung vorgestellt.

1.4. Anmerkung zu eigenen Vorarbeiten

Diese Arbeit ist zum größten Teil im Rahmen des Helmholtz-Projekts *Large Scale Data Management and Analysis* (LSDMA) [Wez+12; Jun+14; Jun+15; Jun+16] entstanden. Das in LSDMA angesiedelte *Data Life Cycle Lab Energy* beschäftigt sich hauptsächlich mit den Herausforderungen bei der Verwaltung und Verarbeitung von Energiedaten entlang des gesamten Datenlebenszyklus sowie der Entwicklung von domänenspezifischen Lösungen.

Die Untersuchungen wurden auch in Kooperation mit weiteren Projekten aus dem Bereich der Energieinformatik durchgeführt, unter anderem im Rahmen einer Zusammenarbeit mit dem *EIT-ICT-Labs*-Projekt *HEGRID*. Die Ergebnisse dieser gemeinsamen Forschung wurden beispielsweise in [RGS14] und [Git+16] publiziert.

Die wesentlichen Beiträge dieser Arbeit sind zum einen die Erarbeitung und Analyse der Anforderungen an ein Energiedatenmanagementsystem und zum anderen die Konzeption eines Energiedatenmanagementsystems, welches diese Anforderungen erfüllt. Die Arbeiten im *Data Life Cycle Lab Energy* beeinflussten beide Aspekte. Eine erste Version des in dieser Arbeit vorgestellten Konzepts wurde in [RS14] veröffentlicht. Darüber hinaus wurde in [Rig+14] ein hierarchisch aufgebautes Energiemanagementsystem skizziert, in welchem das vorgestellte Konzept eines Energiedatenmanagementsystems zum Einsatz kommen kann. Die genannten Publikationen sind folglich die Vorarbeiten, auf welchen das in dieser eigenständigen Arbeit beschriebene Konzept beruht. Die Architektur wurde seither weiterentwickelt, das Konzept basiert jedoch nach wie vor auf der in [RS14] beschriebenen Grundidee.

1.5. Aufbau der Arbeit

Die vorliegende Arbeit gliedert sich in sieben Kapitel, wovon diese Einleitung das erste darstellt. Das zweite Kapitel beschreibt die für diese Arbeit nötigen Grundlagen. Es zeigt, warum die Energiewende zügig umgesetzt werden sollte und was dabei die größten Herausforderungen sind. Darüber hinaus wird dargelegt, was Energiedaten sind und welchen Beitrag sie und ihre Verwendung im Rahmen der Energiewende leisten können. Im dritten Kapitel werden verwandte Arbeiten vorgestellt. Diese umfassen einerseits den Umgang mit und das Management von Energiedaten in Forschung und Wissenschaft. Andererseits werden auch Verfahren zum Schutz der Privatsphäre im Kontext von Energiedaten betrachtet. Im vierten Kapitel erfolgt eine detaillierte Anforderungsanalyse. Diese ist in mehrere Abschnitte aufgeteilt: Auf eine technische Datenlebenszyklusanalyse folgt eine Untersuchung der Nutzerinteressen mit einem Fokus auf Datenschutz und Privatsphäre. Aus den Analysen ergibt sich eine Reihe von Anforderungen, welche ein Energiedatenmanagementsystem erfüllen sollte. Diese Anforderungsliste ist die Grundlage für die Ausarbeitung eines nutzerorientierten Energiedatenmanagementsystems im fünften Kapitel. Zunächst wird ein modulares Konzept für ein solches Datenmanagementsystem vorgestellt und ein passendes Datenmodell entworfen. Anschließend wird die Umsetzung im Rahmen eines Demonstrators inklusive der typischen Arbeitsabläufe beschrieben. Im sechsten Kapitel erfolgt die Evaluation des zuvor erarbeiteten Konzepts. Die Arbeit schließt mit einer Zusammenfassung und dem Ausblick in Kapitel sieben.

Grundlagen

Die Energiewende ist eine Jahrhundertaufgabe, welche mit gemeinsamen Anstrengungen bewältigt werden muss. Allen Widrigkeiten zum Trotz gibt es jedoch zwingende Gründe, die Umsetzung nicht weiter zu verzögern, sondern sogar zu beschleunigen. Da die Energiewende gleichzeitig der wichtigste Treiber bei der Erfassung und Verwertung von Energiedaten ist, beschäftigt sich der erste Teil dieses Kapitels mit ihr. Neben den Gründen für die Energiewende werden auch die damit verbundenen Herausforderungen näher beleuchtet. Der zweite Teil des Kapitels widmet sich den Energiedaten sowie deren Nutzen und den verbundenen Gefahren. Schließlich wird der Begriff des Energiedatenmanagements eingeführt und diesbezügliche Grundlagen erläutert.

2.1. Die Energiewende

Die Abkehr von fossilen Brennstoffen und die dementsprechend verstärkte Nutzung regenerativer Energien wird unter dem Begriff Energiewende zusammengefasst. Über die Zugehörigkeit einzelner Energieträger zu den erneuerbaren Energien herrscht keineswegs Einigkeit. Im Folgenden soll daher die Definition von Quaschning [Qua15] für erneuerbare Energien Anwendung finden. Gemäß dieser versteht man unter dem Begriff der erneuerbaren oder regenerativen Energien diejenigen, welche unter menschlichen Zeithorizonten unerschöpflich sind. Insbesondere ist so eine Einteilung in die drei Bereiche Sonnenenergie, Planetenenergie und geothermische Energie möglich. Während geothermische Energie vor allem in Form von Erdwärme und Planetenenergie in Form von Gezeiten in Erscheinung tritt, wird Sonnenenergie in

vielfältiger Art und Weise sichtbar. Neben der direkten Solarstrahlung treten andere Effekte wie beispielsweise Niederschlag, Wind und Meereströmungen auf, die ihren Ursprung in der Sonnenenergie haben. Auch Biomasse wie Holz oder weiter verarbeitete Produkte wie Biogas können als umgewandelte Sonnenenergie angesehen werden. Diese Energieformen müssen meist mit technischen Anlagen erneut umgewandelt werden, um für den Menschen von Nutzen zu sein. Dies geschieht unter anderem mittels Photovoltaikanlagen, Windkraftanlagen oder Gezeitenkraftwerken. Zu beachten ist, dass im Rahmen dieser Definition die Kernenergie nicht zu den erneuerbaren Energien zählt [Qua15].

Auch heute ist die Entwicklung technischer Anlagen zur Nutzung von erneuerbaren Energien keineswegs abgeschlossen, sondern vielmehr Gegenstand aktueller Forschungsanstrengungen. Allerdings wurden die Wind- und die Wasserkraft schon vor Jahrhunderten in (rein mechanischen) Anlagen genutzt, um beispielsweise Korn zu mahlen oder Wasser zu pumpen. Doch erst die im Vergleich zu heute noch recht einfache Verfügbarkeit und die umfangreiche Nutzung fossiler Brennstoffe haben im 20. Jahrhundert rasanten technischen Weiterentwicklungen ermöglicht. Diese führten vor allem in westlichen Ländern zu gesteigertem Wohlstand und Fortschritt, aber auch zu rasant wachsendem Energieverbrauch. Während im vergangenen Jahrhundert der steigende Energiebedarf noch alleine durch fossile Brennstoffe gedeckt werden konnte, wird dies in absehbarer Zeit nicht mehr der Fall sein. Nicht nur die Zahl der Menschen auf der Erde, sondern auch deren durchschnittlicher Energieverbrauch steigt. Bis zu einem gewissen Grad kann angenommen werden, dass Wohlstand und Energieverbrauch aneinandergekoppelt sind. Nicht zuletzt deshalb wird die Energieversorgung als eine der größten Herausforderungen des 21. Jahrhunderts angesehen [AB07]. Aus heutiger Sicht böte die Kernfusion möglicherweise einen langfristigen Ausweg. Allerdings ist mit Fusionskraftwerken mit positiver Energiebilanz vermutlich frühestens in 50 Jahren zu rechnen. Die verfügbaren erneuerbaren Energien hingegen übersteigen den weltweit nötigen Bedarf um ein Vielfaches. Des Weiteren ist die Deckung des Gesamtenergiebedarfs durch erneuerbare Energien technisch schon heute grundsätzlich möglich, so dass die Nutzung regenerativer Energien eine potenzielle Lösung für die Energieprobleme des 21. Jahrhunderts darstellt [Qua15].

2.1.1. Gründe für die Energiewende

Es gibt eine Vielzahl von Gründen, welche für eine zügige Umsetzung der Energiewende sprechen. Im Folgenden soll ein nicht erschöpfender Überblick hierzu gegeben werden.

2.1.1.1. Ökologische Gründe für die Energiewende

Bereits der – aus erdgeschichtlicher Sicht – extrem kurze Zeitraum der exzessiven Nutzung fossiler Energieträger hat einschneidende Folgen gezeigt. Annähernd alle Veröffentlichungen über den Klimawandel der heutigen Zeit stimmen darin überein, dass dieser existiert und anthropogen ist [Coo+13]. In der Literatur überwiegt darüber hinaus die Ansicht, dass der Klimawandel sich nicht mehr aufhalten lässt, sondern nur noch dessen Ausmaß begrenzt werden kann [Coo+13; Qua15].

In den vergangenen Jahrzehnten konnte ein verstärktes Abschmelzen der Gletscher und Pole beobachtet werden. Auch der Anstieg des Meeresspiegels ist nicht mehr zu leugnen. Darüber hinaus ist zu befürchten, dass die Intensität und die Häufigkeit von Wetterextremen in Zukunft weiter zunehmen werden. Eine präzise Prognose aller Auswirkungen ist durch das komplexe Ökosystem der Erde aber nicht möglich. Nicht zuletzt ist die zukünftige Entwicklung auch stark davon abhängig, wie schnell und wie weit die durchschnittliche Temperatur auf der Erde steigen wird. Aktuell beträgt die Rate der durch den Menschen ausgelösten Erderwärmung etwa $0,2\text{ }^{\circ}\text{C}$ pro Jahrzehnt. Um die potenziell katastrophalen Folgen einzuschränken, wird auf politischer Ebene eine Begrenzung der Erderwärmung auf $2\text{ }^{\circ}\text{C}$ angestrebt [Han+06; Qua15].

Maßgeblich verursacht wird die Erderwärmung durch anthropogene Treibhausgase wie beispielsweise Methan oder Kohlenstoffdioxid. Der Treibhauseffekt ist dafür verantwortlich, dass ein Teil der Energie der auf der Erde auftreffenden Strahlung nicht ins Weltall reflektiert, sondern in der Atmosphäre gespeichert wird. Ohne den natürlichen Treibhauseffekt wäre die Erdoberfläche zu kalt für uns heute bekanntes Leben. Jedoch verursachen von Menschen emittierte Gase eine Verschiebung des Gleichgewichts und dadurch eine Verstärkung des natürlichen Treibhauseffekts. So ist heute beispielsweise der Gehalt an Kohlenstoffdioxid in der Atmosphäre auf dem Maximum der ge-

samtens letzten 650 000 Jahren. Um den Klimawandel abzubremsen, ist wieder eine massive Reduktion des Ausstoßes solcher Gase nötig. Der durch Verbrennungsprozesse verursachte Ausstoß von Kohlenstoffdioxid könnte durch die Energiewende signifikant gesenkt werden, wenn eine großflächige Ersetzung fossiler Brennstoffe durch erneuerbare Energien erfolgen würde [Qua15].

Neben diesen globalen Auswirkungen der Verbrennung fossiler Energieträger gibt es auch zahlreiche, teilweise irreversible, Eingriffe in die Natur bei der Gewinnung der nötigen Rohstoffe. Umweltkatastrophen wie der Unfall auf der Ölbohrplattform *Deepwater Horizon* im Jahr 2010 führen vor Augen, welche drastischen Folgen die Erdölförderung haben kann [Cam+10]. Doch auch bei planmäßigem Abbau von fossilen Brennstoffen kann der Einfluss auf gesamte Landstriche große Ausmaße annehmen. Durch den Braunkohletagebau werden quadratkilometergroße Flächen zerstört und für lange Zeit verändert [Bun16]. Ähnliches gilt für den Abbau von uranhaltigen Erzen als Ausgangsstoff für die Brennstäbe von Kernkraftwerken. Auch hier müssen viele Tonnen Gestein bewegt werden, um die in geringer Menge vorkommenden Uranerze fördern zu können [Qua15].

2.1.1.2. Ökonomische Gründe für die Energiewende

Aus ökonomischer Sicht sprechen ebenfalls viele Argumente für eine rasche Umsetzung der Energiewende. Wie der Name der fossilen Brennstoffe schon erahnen lässt, handelt es sich bei ihnen nicht um erneuerbare Energiequellen. Erdöl, Erdgas und Kohle sind über Jahrtausende aus organischem Material entstanden. Dementsprechend sind diese Ressourcen endlich und nicht in unbegrenztem Maße verfügbar. Aufgrund vieler Unsicherheiten sind exakte Prognosen über die Reichweite der vorhandenen Ressourcen kaum möglich. Es ist jedoch davon auszugehen, dass bei konstantem Verbrauch die verfügbaren Öl- und Gasvorräte noch deutlich vor Ende dieses Jahrhunderts aufgebraucht sein werden [Cra12]. Die Steinkohlevorräte werden vermutlich in weniger als 150 Jahren erschöpft sein. Auch wenn vermutlich noch weiter Vorkommen entdeckt werden können, ändert dies nichts an der Endlichkeit der fossilen Energiequellen. Vielmehr wird lediglich die Grenze etwas weiter in die Zukunft verschoben. Uran als Spaltstoff für Kernkraftwerke steht ebenfalls nicht unbegrenzt zur Verfügung. Die Lagerstätten wären – sollte der

Weltbedarf allein aus Kernenergie gedeckt werden – noch innerhalb dieses Jahrhunderts ausgeschöpft [Cra12; Qua15].

Auch wenn die Reichweite der fossilen Ressourcen noch mehrere Jahrzehnte beträgt, sind die ökonomischen Folgen der drohenden Knappheit schon heute spürbar. Der Abbau und die Verwertung fossiler Energieträger wird immer aufwendiger sowie gefährlicher und damit teurer. Es wird damit gerechnet, dass innerhalb der nächsten zwanzig Jahre das globale Ölfördermaximum erreicht sein wird. In wenigen Jahren Abstand wird dann vermutlich das globale Fördermaximum für Erdgas folgen. Dementsprechend darf damit gerechnet werden, dass langfristig die Preise für Erdöl und Erdgas steigen werden. Dies ist insbesondere dann der Fall, wenn die Nachfrage nach fossilen Energieträgern nicht sinkt oder im Gegenteil sogar steigt. Eine Substitution fossiler Brennstoffe durch Kernenergie ist aus ökonomischer Sicht ebenfalls nicht möglich. Würde der gesamte Energiebedarf aus Kernenergie gedeckt, wäre die Reichweite ähnlich kurz wie bei Öl und Gas und damit dieselbe Problematik gegeben [Cra12; Qua15].

Die Energiewende und erneuerbare Energien im Speziellen werden häufig dafür kritisiert, dass sie hohe Kosten verursachen. Insbesondere wird argumentiert, dass die Energieerzeugung aus konventionellen Quellen deutlich preiswerter sei. Bei dieser Argumentation werden aber meist die externen Kosten vernachlässigt [Wes13]. Unter externen Kosten versteht man solche Kosten, die nicht von den Verursachern, sondern von anderen Individuen oder der Gesellschaft als Ganze getragen werden. Dies umfasst Kosten für schleichend auftretende Schäden an Umwelt und Menschen ebenso wie Kosten für Unfälle. Ein Beispiel hierfür sind von saurem Regen verursachte Schäden an Wäldern aber auch Gebäuden [Umw12; Wes13].

Eine genaue monetäre Bewertung externer Kosten ist überaus schwierig und nicht unstrittig. So ist beispielsweise in [Umw12] eine Spanne von 0,2 ct/kWh bis hin zu 200 ct/kWh für Kernenergie aufgeführt. Je nach Bewertung des Risikos und der Eintrittswahrscheinlichkeit entstehen massive Unterschiede. Weitere konkrete Kosten sind in der aktuellen Fassung jedoch nicht mehr aufgeführt. In der älteren Auflage aus dem Jahr 2007 [Umw07] sind hingegen noch konkrete Zahlen abgedruckt. Die dort für fossile Energien angegebenen externen Kosten betragen ein Vielfaches der externen Kosten für erneuerbare Energien. So wird beispielsweise bei Braunkohle mit 8,7 ct/kWh

gerechnet, während die externen Kosten für Strom aus Photovoltaik lediglich mit 0,8 ct/kWh bepreist werden. Die durchschnittlichen externen Kosten für Strom (ohne Kernkraft) betragen im Jahr 2005 geschätzte 5,8 ct/kWh. Auch wenn diese konkreten Zahlen heute vermutlich überholt sind, kann doch geschlossen werden, dass die externen Kosten konventioneller Energieträger die der Erneuerbaren deutlich übersteigen. Würde eine gesamtwirtschaftliche Betrachtung durchgeführt, wäre damit der reale Preisunterschied deutlich geringer als meist angegeben. Erschwerend kommt hinzu, dass im Bereich der konventionellen Energieerzeugung nicht mehr mit großen technischen Fortschritten zu rechnen ist, während sich viele Bereiche der erneuerbaren Energieerzeugung technisch noch am Anfang befinden und Effizienzsteigerungen erwartet werden dürfen.

2.1.1.3. Ethische und soziale Gründe für die Energiewende

Armaroli *et al.* [AB07] vergleichen die Erde mit einem Raumschiff, welches mit einem Energieproblem durch die Unendlichkeit des Weltalls fliegt. Alternative Planeten oder andere Ausweichpläne sind nicht in Sicht. Trotzdem findet immer noch ein Überstrapazieren der vorhandenen Ressourcen statt. Aus diesem Blickwinkel können Fragen bezüglich Nachhaltigkeit und Verantwortung für unser Tun nicht ignoriert werden. Insbesondere ist es schwierig zu rechtfertigen, dass über Jahrtausende entstandene Ressourcen in Form von fossilen Energieträgern durch einige wenige Generationen beinahe aufgebraucht wurden. Der Vorwurf, Raubbau zu betreiben, wiegt umso schwerer, da zumindest technisch ein Umstieg auf eine nachhaltige und ressourcenschonende Energiewirtschaft möglich wäre [AB07; Qua15].

Doch nicht nur die Erschöpfung der endlichen Ressourcen ist ein ethisches Problem, auch die mit der Förderung der Rohstoffe verbundenen Gefahren und Probleme werfen moralische Fragen auf. Immer wieder vorkommende Unfälle in Kohlegruben oder die Zerstörung kompletter Regionen durch Tagebau haben wenig mit Nachhaltigkeit und Rücksichtnahme auf zukünftige Generationen gemein. Im Gegensatz zu Unfällen und zerstörten Landschaften erfahren gesundheitliche Probleme in Folge von Luftverschmutzung meist weniger Aufmerksamkeit. Die *World Health Organization* (WHO) [Sco15] schätzt aber, dass pro Jahr etwa 3,7 Millionen Menschen in Folge von Luftverschmut-

zung sterben. Im Vergleich dazu werden jährlich etwa 1,3 Millionen Menschen im Straßenverkehr getötet. Ein großer Teil der Luftverschmutzung wird durch die Verbrennung fossiler Energieträger verursacht. Im Gegensatz zu Kohlenstoffdioxid haben viele der schädlichen Partikel aber nur eine relativ kurze Verweildauer in der Atmosphäre, so dass eine Abkehr von fossilen Brennstoffen bereits kurzfristig eine positive Wirkung hätte.

Wie zuvor bereits beschrieben ist die langfristige Weiternutzung von Kernenergie ökonomisch nicht sinnvoll. In noch viel größerem Maße widerspricht sie aber dem Gedanken der Nachhaltigkeit und vorausschauendem Handeln. Die bei der Kernspaltung entstehenden Spaltprodukte haben zum Teil Halbwertszeiten von vielen Jahrtausenden. Aus menschlicher Sicht sind solche Zeiträume nicht planbar und es ist nicht abzuschätzen, welche Folgen zukünftige Generationen erwarten. Auch die Frage nach geeigneten Endlagerstätten ist bei Weitem noch nicht abschließend geklärt. Somit ist es auch aus ethischer Sicht fragwürdig, ob länger als unbedingt nötig auf die Kernenergie gesetzt werden darf.

2.1.1.4. Politische Entscheidungen zur Energiewende

In Deutschland herrscht weitgehend Einigkeit darüber, dass erneuerbare Energien im Energiesystem der Zukunft eine wichtige Rolle spielen werden. Die konkrete Ausgestaltung dieses Systems ist allerdings Gegenstand zahlreicher Debatten. Es ist durchaus machbar, dass beinahe die gesamte Energieversorgung in Deutschland durch Erneuerbare übernommen wird [HP14]. Auch wenn die politischen Ziele Deutschlands nicht ganz so weit gehen, sind sie dennoch ambitioniert. So sollen bis zum Jahr 2050 die Treibhausgasemissionen um 80 % gesenkt werden. Des Weiteren soll der Anteil an erneuerbaren Energien in der Stromversorgung bis zum Jahr 2050 auf 80 % gesteigert werden. Auch der Ausstieg aus der Kernenergie bis zum Jahr 2022 ist nach der Katastrophe in Fukushima beschlossen [Bun15c]. Die dadurch wegfallenden Kapazitäten müssen durch neue Kraftwerke ersetzt werden, die zum Teil auch konventioneller Natur sein werden, wobei insbesondere auf eine gute Regelbarkeit geachtet werden muss [Bun12]. Darüber hinaus wird angestrebt, die Energieeffizienz zu steigern, beispielsweise in Form einer Reduktion des Wärmebedarfs in Gebäuden [Bun11].

Dänemark hat eine Vorreiterrolle im Bereich der erneuerbaren Energien und strebt an, was Henning *et al.* [HP14] auch für Deutschland für möglich erachten. Bis zum Jahr 2050 soll der gesamte Energiebedarf Dänemarks komplett durch erneuerbare Energien gedeckt werden. Demgegenüber gibt es immer noch zahlreiche Staaten, die weiterhin auf Kernkraft und fossile Energieträger setzen. Die ursprünglich geplante Laufzeitverlängerung der Kernkraftwerke in Deutschland und der – wahrscheinlich – endgültige Ausstieg aus der Kernenergie haben gezeigt, dass sich politische Ausrichtungen ändern können. Es bleibt zu hoffen, dass die gesteckten Ziele ambitioniert und mit Nachdruck verfolgt werden, da die überwiegende Zahl an Argumenten für die Energiewende spricht.

2.1.2. Herausforderungen der Energiewende

Auch wenn die Energiewende alternativlos und überfällig scheint, gibt es dennoch Hürden, die bewältigt werden müssen. Im Folgenden wird auf einige der Herausforderungen, die mit der Umsetzung der Energiewende verbunden sind, eingegangen.

2.1.2.1. Abkehr von einer zentralen Energieversorgung

Im klassischen Energiesystem versorgen wenige zentrale Erzeuger viele dezentrale Verbraucher. Dabei werden Kraftwerke häufig strategisch so platziert, dass sich in ihrer Nähe ein Großteil der entsprechenden Verbraucher befindet und die Kraftwerke effizient mit den nötigen Brennstoffen versorgt werden können. Kraftwerke, die aus erneuerbaren Energieträgern Energie gewinnen, sind aber häufig standortgebunden. Dies ist beispielsweise bei Off-Shore-Windparks oder der Nutzung von Wasserkraft der Fall. Bei diesen Kraftwerkstypen geben die verfügbaren und nutzbaren erneuerbaren Energien den Standort vor [Cra12].

Durch eine Einspeisevergütung wurden in Deutschland Anreize für den Ausbau der Photovoltaikanlagen gesetzt. Auch wenn es diverse vergleichsweise große und zentrale Anlagen gibt, existieren daneben viele kleine, private Photovoltaikanlagen und weitere kleine Erzeuger wie Blockheizkraftwerke. Des Weiteren wurden in den letzten Jahren zahlreiche neue Windkraftwerke

errichtet. Diese Entwicklungen verursachen eine weitere Verlagerung der Produktionskapazitäten weg von wenigen zentralen Standorten hin zu vielen dezentralen. Das klassische, für diese Art der Energiegewinnung ursprünglich nicht ausgelegte, Stromnetz kann dadurch unter Umständen an seine Grenzen geführt werden, da die Orte der Energiegewinnung bzw. -nutzung zum Teil weit voneinander entfernt sind und große Energiemengen über weite Strecken transportiert werden müssen. Die Situation wird verschärft durch Schwankungen bei der Einspeisung und damit einhergehenden lokalen Spannungsunregelmäßigkeiten sowie Engpässen. Die Stabilität der Netze kann bei solchen Einflüssen nicht mehr nur durch Maßnahmen der Übertragungsnetzbetreiber gewährleistet werden, sondern muss auf der Verteilnetzebene angegangen werden.

Die gesteigerte Zahl dezentraler Erzeuger erschwert darüber hinaus die Steuerung der Energieerzeugung. Während im klassischen Energiesystem noch vergleichsweise wenige, große Kraftwerke geregelt werden mussten, sind nun viele, kleine, heterogene Erzeuger zu integrieren. Eine Regelbarkeit aller dezentralen Einspeiser ist dabei nicht immer sichergestellt. Doch selbst wenn eine Eingriffsmöglichkeit gegeben ist, gestaltet sich die Regelung des gesamten Systems aufgrund der deutlich höheren Zahl an Akteuren ungleich komplexer. Ohne entsprechende intelligente Regelung ist nur ein reaktives Verhalten nötig, was häufig bedeutet, dass konventionelle Kraftwerke verwendet werden müssen, um Regelenergie bereitzustellen, mit deren Hilfe auftretende Differenzen ausgeglichen werden [Cra12].

2.1.2.2. Fluktuiierende Energieproduktion

Kraftwerke, die mit fossilen Brennstoffen arbeiten, sind ebenso wie Kernkraftwerke meist sehr gut regelbar. Im Voraus festgelegte Kraftwerksfahrpläne können mit einer hohen Zuverlässigkeit eingehalten werden. Dabei gibt es selbstverständlich Unterschiede in der Geschwindigkeit, mit der die Leistung verändert werden kann. So reagieren beispielsweise Kernkraftwerke aber auch Braunkohlekraftwerke recht träge, während Gaskraftwerke in sehr kurzen Zeiträumen ihre Leistung ändern und damit Regelenergie bereitstellen können. Dennoch ist den konventionellen Kraftwerken gemein, dass ihre Leistungen in gewissen Grenzen einfach geregelt werden können [Wes13].

Demgegenüber sind die meisten Kraftwerkstypen, die erneuerbare Energien nutzen, nicht oder nur schlecht regelbar. Wenn keine ausreichende Solarstrahlung zur Verfügung steht, können Photovoltaikanlagen keinen Strom produzieren. Analog können Windkraftanlagen bei Flaute keine Energie gewinnen. Laufwasserkraftwerke können bei Niedrigwasser ihren Ertrag ebenfalls nicht steigern. Pumpspeicherkraftwerke weisen sich zwar durch eine sehr gute Regelbarkeit aus, die es erlaubt, die Leistung in kürzester Zeit anzupassen, jedoch sind in Deutschland die Installationskapazitäten bereits weitgehend ausgeschöpft. Deren Anteil an der Gesamtstromproduktion wird folglich gering bleiben. Folglich kann bei einem hohen Anteil an erneuerbaren Energien die Produktion nicht immer der Nachfrage folgen. Dementsprechend ist ein Wandel weg von der verbrauchsorientierten Produktion hin zu einem produktionsorientierten Verbrauch nötig [Fan+12; IA09; Bun15c]. Dies gilt umso mehr, da ungenutzte erneuerbare Energie nicht nachträglich verbraucht werden kann. Es wird somit nötig sein, Speicher für elektrische Energie bereitzustellen, die es erlauben, nicht genutzte Energie für zukünftige Versorgungsengpässe vorzuhalten [Cra12; Wes13].

2.1.2.3. Erntefaktor erneuerbarer Energien

Im Vergleich zur Reichweite fossiler Brennstoffe scheint die verbleibende Brenndauer der Sonne nach menschlichen Maßstäben unvorstellbar lang zu sein. Während die fossilen Brennstoffe bei konstantem Verbrauch innerhalb der nächsten Jahrzehnte bis Jahrhunderte aufgebraucht sein werden, wird die Sonne unsere Erde noch ungefähr sieben Milliarden Jahre mit Energie versorgen [Cra12]. Die Reichweite der erneuerbaren Energien übertrifft damit diejenige der fossilen Brennstoffe um Größenordnungen. Dennoch haben die erneuerbaren Energien einen nicht zu vernachlässigenden Nachteil gegenüber konventionellen Kraftwerken.

Der Begriff Erntefaktor beschreibt das Verhältnis von tatsächlich genutzter Energie zu eingesetzter Energie. In der englischsprachigen Literatur ist die Bezeichnung *energy returned on (energy) invested* (EROI) üblich. Vereinfacht ausgedrückt gibt der Erntefaktor an, wie viele Einheiten Energie verfügbar gemacht werden können, wenn eine Einheit investiert wird. Bei der Abschätzung der investierten Energie sollte der gesamte Lebenszyklus des zu untersuchen-

den Kraftwerks betrachtet und einbezogen werden. Somit müssen nicht nur der Bau und der Betrieb eines Kraftwerks energetisch bewertet werden. Vielmehr muss auch der Energieaufwand für die Brennstoffversorgung und den Rückbau des Kraftwerks abgeschätzt werden.

Weißbach *et al.* [Wei+13] haben für verschiedene Kraftwerkstypen den Erntefaktor abgeschätzt. Die für konventionelle Kraftwerke angegebenen Erntefaktoren sind um eine Größenordnung höher als die der Kraftwerke, welche erneuerbare Energien nutzen. Bei ausschließlicher Betrachtung des Erntefaktors schneiden Kernkraftwerke mit großem Abstand am besten ab. Schätzungsweise 75 Einheiten Energie können mit einer Einheit produziert werden. Dies erklärt sich vor allem durch die enorme Energiedichte des verwendeten Spaltmaterials. Dabei sind Energieaufwendungen des gesamten Lebenszyklus inklusive Abbau, Lagerung und Rückbau eingerechnet. Um dieselbe Energie mit einem Kohlekraftwerk gewinnen zu können, muss im Vergleich ein Vielfaches der Masse an Brennstoff gefördert und transportiert werden. Mit einem Erntefaktor von 50 schneiden Wasserkraftwerke auch hervorragend ab. Steinkohle- und Erdgaskraftwerke mit einem Erntefaktor von etwa 30 liegen im Mittelfeld. Windkraftwerke, welche einen großen Teil der Stromproduktion in Deutschland erbringen sollen, weisen nur noch einen Erntefaktor von 16 auf. Photovoltaikanlagen haben mit 3,9 den schlechtesten Erntefaktor der verglichenen Kraftwerkstypen. Es gibt vor allem zwei Gründe für die schlechte Einschätzung der erneuerbaren Energien. Zum einen ist das die relativ kurze angenommene Laufzeit. Bei Photovoltaik- oder Windkraftanlagen wird mit einer Lebensdauer von etwa 20 Jahren gerechnet, während bei Kohle- und Kernkraftwerken 50 bis 60 Jahre erwartet werden. Zum anderen können Photovoltaik- und Windkraftanlagen nicht dauerhaft Strom produzieren, sondern sind abhängig von der Verfügbarkeit erneuerbarer Energien. Konventionelle Kraftwerke können – von Stillständen aufgrund von Wartung oder Reparaturen abgesehen – theoretisch durchgehend betrieben werden. Um den aus erneuerbaren Energien gewonnenen Strom ebenfalls zeitlich unabhängig von der Produktion nutzen zu können, wurde auch eine Integration von Speichern für elektrische Energie betrachtet. Aus energetischer Sicht verschlechtern sie den Erntefaktor weiter, da Bau, Betrieb und Rückbau solcher Speicher große Mengen Energie benötigen. Dieser Energieeinsatz wird aus energetischer Sicht in der gesamten Lebensdauer nicht mehr

aufgewogen [Wei+13]. Diese Einschätzungen beruhen jedoch naturgemäß nicht auf dem aktuellen technologischen Stand. So stammen die in [Wei+13] zitierten Untersuchung zur Photovoltaik beispielsweise aus dem Jahr 2007. Dementsprechend ist nicht zuletzt aufgrund der intensivierten Forschung und Entwicklung mit einer signifikanten Verbesserung im Bereich der Nutzung von erneuerbaren Energien zu rechnen.

Es muss zusätzlich betont werden, dass bei der Abschätzung des Erntefaktors lediglich die Energiebilanz betrachtet wird und externe Kosten nicht mit einbezogen werden. Darüber hinaus hängt der berechnete Erntefaktor so stark von den angenommenen Randbedingungen ab, dass diese das Ergebnis maßgeblich beeinflussen können. Aus diesem Grund sind in der Literatur zum Teil widersprüchliche Angaben zu finden [GS12; Wei+13]. Zu kritisieren ist an diesem Ansatz außerdem, dass der energetische Aufwand für eine Kohlenstoffdioxidabscheidung ignoriert wird. Sollen mit konventionellen Kraftwerken ähnlich gute Kohlenstoffdioxidbilanzen erzielt werden wie mit erneuerbaren Energien, so ist zwingend eine Reduktion des Treibhausgasausstoßes nötig. Dies bedeutet aber, dass eine große Menge an Energie aufgewendet werden muss, um Kohlenstoffdioxid aus den Verbrennungsgasen abzuscheiden. Dadurch würde der Erntefaktor der Kohle- und Erdgaskraftwerke deutlich sinken. Der Abbau und Transport von Uran verursacht ebenfalls Kohlenstoffdioxidemissionen, so dass bei Beachtung dieser ebenfalls ein schlechterer Erntefaktor für Kernenergie erwartet werden muss [Bun13b]. Nicht zuletzt kommt hinzu, dass die verbliebenden Vorkommen fossiler Brennstoffe mit immer aufwendigeren Methoden nutzbar gemacht werden müssen. Auch dies wird zu einer weiteren Verschlechterung der entsprechenden Erntefaktoren führen. Dem entgegen stehen die zu erwartenden technischen Fortschritte im Bereich der Nutzung erneuerbarer Energien, die eine weitere Verbesserung der Erntefaktoren mit sich bringen werden.

Trotzdem ist festzustellen, dass mit dem heutigen Stand der Technik die Erntefaktoren der meisten Kraftwerke für erneuerbare Energien schlechter sind als die der konventionellen Kraftwerke [Wei+13]. Aus diesem Grund ist es umso wichtiger, dass angestrebt wird, die gesamte aus Erneuerbaren produzierte Energie zu nutzen und möglichst selten die Kraftwerke wegen Überproduktion abzuschalten.

2.1.2.4. Gesellschaftliche Vorbehalte gegen die Energiewende

Allgemein erfahren Kraftwerke, die durch erneuerbare Energieträger gespeist werden, eine hohe Akzeptanz in Deutschland. Dies gilt selbst dann, wenn die Kraftwerke im eigenen Wohnumfeld betrieben werden. Mit 72 % Zustimmung erfahren Solarparks die größte Beliebtheit bei Kraftwerken. Auch die Windkraft ist mit 61 % weitgehend akzeptiert. Im Vergleich dazu erfahren Kohle- und Kernkraftwerke mit 11 % beziehungsweise nur 5 % eine starke Ablehnung [Str14].

Im Unterschied zu großen konventionellen Kraftwerken haben kleinere Photovoltaik- beziehungsweise Windkraftanlagen aber eine um Größenordnungen geringere Leistung. Dementsprechend ist eine vielfach größere Zahl solcher Anlagen vonnöten. Während es noch recht einfach möglich ist, außer Sichtweite jeglicher konventionellen Kraftwerke zu wohnen, fällt dies bei den Erneuerbaren deutlich schwerer. Der Wunsch, solche möglichen Unannehmlichkeiten anderen zuzuschieben, um sie nicht selbst tragen zu müssen, wird im Englischen *Nimby* genannt. Das *Cambridge Advanced Learner's Dictionary & Thesaurus* listet *Nimby* als Abkürzung für „*Not in my back yard*“. Da Kraftwerke für erneuerbare Energien potenziell in mehr „Hinterhöfen“ stehen als konventionelle Kraftwerke, ist selbst bei allgemeiner, breiter Zustimmung mit vereinzeltem Widerstand aus der Bevölkerung zu rechnen, da die Zahl betroffener Menschen größer ist.

Eine gewisse Mitschuld an einer möglichen ablehnenden Haltung gegenüber der Nutzung erneuerbarer Energien trägt zum Teil auch die öffentliche Berichterstattung. So untersuchten Zhou *et al.* den Einfluss von Windkraftanlagen auf die Bodentemperatur in umliegenden Gebieten und veröffentlichten die Ergebnisse in *Nature* [Zho+12]. Die Messungen ergaben eine Steigerung der Durchschnittstemperaturen in Bodennähe im Vergleich zu Gebieten ohne Windkraftanlagen. Diese Ergebnisse wurden auf die Umwälzung der verschiedenen Luftschichten zurückgeführt. Die Veröffentlichung wurde in Folge in der internationalen Presse diskutiert und dabei häufig stark verfälscht wiedergegeben. Dies gipfelte in der Behauptung, die Veröffentlichung zeige, dass Windkraftanlagen zur globalen Erderwärmung beitragen könnten beziehungsweise diese mitverschuldeten. Dabei handelt es sich aber um eine gravierende Fehlinterpretation und -darstellung der Ergebnisse [Plu12]. Selbstverständlich

müssen auch die Nachteile und potenziellen Gefahren der Nutzung erneuerbarer Energien erforscht und veröffentlicht werden. Es ist aber auch unsere Aufgabe als Wissenschaftler, dafür Sorge zu tragen, dass gewonnene Erkenntnisse korrekt und unverfälscht wiedergegeben und weiterverwendet werden. Übertragen auf die Energiewende und erneuerbare Energien bedeutet dies auch, dass wir Aufklärungsarbeit leisten und unbegründete Vorurteile ausräumen sollten. Ohne starken gesellschaftlichen Rückhalt wird die Energiewende eine noch schwierigere Aufgabe.

2.1.3. Notwendigkeit der Energiewende

Die Energiewende wird die Menschheit vermutlich noch einige Jahrzehnte vor immer neue Herausforderungen stellen. Es ist nicht damit zu rechnen, dass in naher Zukunft die Kernfusion unsere Energieversorgung übernehmen kann. Die zuvor genannten gravierenden Nachteile und die Endlichkeit der konventionellen Energieerzeugung lassen aus heutiger Sicht nur die Nutzung erneuerbarer Energien als Ausweg. Die Energiewende ist folglich eine notwendige Entwicklung, zu der keine echten Alternativen existieren. Die damit verbundenen Veränderungen und Herausforderungen lassen jedoch Raum für Verbesserungen und Optimierungen. So ist beispielsweise zu erwarten, dass durch technische Fortschritte die Energieausbeute weiter gesteigert und die Effizienz verbessert werden können. Darüber hinaus werden neue Methoden Anwendung finden, um die Gesamtheit der zahlreichen Akteure eines komplexen zukünftigen Netzes effizient steuern zu können. Es existieren viele weitere Bereiche, welche ihren Teil zur Energiewende beitragen können.

Diese Einführung in die Thematik der Energiewende soll aufzeigen, dass es eine Vielzahl zu lösender Herausforderungen gibt, welche nicht einfach durch massiven Ressourceneinsatz gelöst werden können, sondern intelligente Lösung erfordern. Im nächsten Abschnitt werden Energiedaten vorgestellt, welche solche intelligenten Ansätze ermöglichen und deshalb im Rahmen der Energiewende genutzt werden sollten.

2.2. Energiedaten

Obwohl wir Energie in vielen ihrer zahlreichen Formen tagtäglich nutzen, sind Energie und die damit verwandten Begriffe für viele Menschen etwas Abstraktes. Darüber hinaus besteht nicht immer Einvernehmen darüber, was genau unter den jeweiligen Begriffen zu verstehen ist. Daher werden im Folgenden unter anderem sowohl die Begriffe Energieverbrauch beziehungsweise -produktion erklärt, als auch der Begriff der Energiedaten definiert.

2.2.1. Energieverbrauch und -produktion

Der Begriff Energieverbrauch hat sich im alltäglichen Sprachgebrauch etabliert. Beim Kauf von Haushaltsgeräten, aber auch von Fernsehern oder Lampen klären Energielabel potenzielle Käufer über den Energieverbrauch der Geräte auf. Auch bei den Abrechnungen der Energieversorger wird der Kunde mit seinem Energieverbrauch konfrontiert. Dennoch bleibt diese Größe für viele oft nur ein abstrakter Wert, dessen Höhe sich lediglich in den entstehenden Kosten widerspiegelt [BN08]. Die Definition des Duden ist ebenfalls wenig aufschlussreich. Energieverbrauch wird schlicht als „Verbrauch an Energie“ [Sch13] definiert. Auch wenn dieser Begriff ein fester Bestandteil unserer Sprache geworden ist, ist er streng genommen aus physikalischer Sicht dennoch nicht korrekt.

Wenn umgangssprachlich von Energieverbrauch die Rede ist, ist darunter eigentlich Exnergieverbrauch zu verstehen. Bei Exergie handelt es sich um denjenigen Anteil an Energie in einem System, der Arbeit verrichten kann. Im Gegensatz zu Energie kann Exergie dementsprechend durchaus verbraucht werden. Wird eine bestimmte Menge an Exergie verwendet, um Arbeit zu verrichten, so steht sie anschließend innerhalb des Systems nicht mehr zur Verfügung. Die aufgewendete Energiemenge wird in eine andere Energieform umgewandelt, aber nicht vernichtet [BK12]. Dies ist beispielsweise dann der Fall, wenn elektrische Energie verwendet wird, um Arbeit zu verrichten und im Zuge dessen in Wärmeenergie umgewandelt wird. Die entstandene Wärmeenergie lässt sich nicht mehr komplett zu elektrischer Energie umwandeln. Dennoch ist keine Energie verbraucht oder vernichtet worden, sie wurde lediglich von einer in eine andere Form überführt.

Analog zum Verbrauch ist der Begriff Energieproduktion zu behandeln. Genau wie Energie nicht vernichtet werden kann, kann sie auch nicht produziert werden. Vielmehr kann Energie von einer – für technische Zwecke – weniger nützlichen Energieform in eine andere umgewandelt werden. Anwendung findet dieses Prinzip beispielsweise in konventionellen Kraftwerken, in welchen die chemisch gespeicherte Energie fossiler Brennstoffe durch Verbrennen in Wärmeenergie und schließlich in elektrische Energie umgewandelt wird. Elektrische und mechanische Energie bestehen zu 100 % aus Exergie und können theoretisch auch komplett in andere Energieformen überführt werden. Thermische Energie kann hingegen unterschiedliche Exergieniveaus aufweisen und es ist nicht möglich, sie in ihrer Gesamtheit in eine andere Energieform zu überführen. Daraus ergibt sich wiederum, dass es unterschiedliche Wertigkeiten von Energie gibt. Diese Unterscheidungen sind alleine mit dem Begriff der Energie nicht möglich [VDI14].

Das Dilemma, zwischen wissenschaftlich korrekter und allgemeinsprachlich üblicher Bezeichnung für die Nutzung von Energie unterscheiden zu müssen, hat auch der Verein Deutscher Ingenieure (VDI) erkannt. In der Richtlinie 4661 [VDI14] gestattet der VDI die Verwendung der Begriffe Energieproduktion beziehungsweise Energieverbrauch, obwohl diese aus wissenschaftlicher Sicht inkorrekt sind. Begründet wird dies mit der weiten Verbreitung der Begriffe und der zu erwartenden Verwirrung, würde man versuchen, die Bezeichnungen im allgemeinen Sprachgebrauch zu ändern. Der Unterschied von Kraft und Energie wird analog behandelt. Dementsprechend wird der Begriff Kraftwerk akzeptiert, obwohl eine solche Anlage eigentlich keine Kraft, sondern Energie abgibt.

Diese Arbeit verzichtet der Einfachheit halber auf die Unterscheidung von Energie beziehungsweise Exergie und folgt dementsprechend auch den Empfehlungen des VDI. Im weiteren Verlauf werden wie bisher auch schlicht die Begriffe Energieverbrauch und -produktion verwendet, auch wenn eigentlich der Verbrauch beziehungsweise die Produktion von Exergie gemeint sind.

2.2.2. Definition Energiedaten

Für den Begriff Energiedaten existiert keine allgemeingültige Definition. Vielmehr kann er je nach Kontext unterschiedliche Bedeutungen haben, die zum

Teil stark voneinander abweichen. So veröffentlicht das Bundesministerium für Wirtschaft und Energie (BMWi) beispielsweise eine *Gesamtausgabe der Energiedaten* [Bun15d] und beschreibt diese wie folgt.

„Die ‚Energiedaten‘ sind eine Sammlung des Wirtschaftsministeriums mit aktuellen Daten zur Energieversorgung aus zuverlässigen heimischen und internationalen Quellen. [...]“ [Bun15d]

Bei dieser Veröffentlichung handelt es sich um eine umfangreiche Sammlung verschiedener Daten mit Energiebezug, welche durch das BMWi veröffentlicht werden. Die Daten umfassen ein weites Spektrum an Informationen, welche die Energiegewinnung, den -verbrauch und auch -handel betreffen. Die Spanne reicht von Heizwerten verschiedener Energieträger über nach Sektoren aufgeschlüsselten Energieverbräuchen bis hin zu internationalen Handelsbeziehungen. Die meist jährlich aktualisierten Daten werden aus volkswirtschaftlicher Sicht und hochaggregiert dargestellt. Daten zu einzelnen Individuen sind nicht enthalten. Vielmehr stellt die Veröffentlichung ein Gesamtbild der Energiesituation innerhalb Deutschlands sowie die damit zusammenhängenden Beziehungen zu anderen Ländern dar. Im Folgenden wird jedoch eine andere, für diese Arbeit passendere, Definition verwendet werden.

In der verwandten Literatur (vgl. Kapitel 3) erfährt der Begriff der Energiedaten oftmals eine von der oben genannten Definition abweichende Bedeutung. Es handelt sich dabei nämlich nicht um Daten mit volkswirtschaftlichem Charakter, sondern meist um Zeitreihen von Messwerten mit einem irgendwie gearteten Energiebezug. Dies umfasst Messwerte von intelligenten Stromzählern ebenso wie auf Geräteebene erfasst Verbrauch. Aber nicht nur Werte zur Nutzung elektrischer Energie zählen zu den Energiedaten, sondern beispielsweise auch Gasverbrauchsdaten. Eine exemplarische Darstellung verschiedener Interpretationen des Begriffs Energiedaten findet sich in Tabelle 2.1. Bereits dieser kleine Anteil an Veröffentlichungen zu diesem Thema bildet ein breites Spektrum an verschiedenen Arten von Energiedaten ab, so dass der Begriff der Energiedaten relativ weit gefasst werden sollte.

Kapitel 2. Grundlagen

Tabelle 2.1.: Der Begriff Energiedaten ist nicht eindeutig definiert. Je nach Kontext kann er unterschiedliche Bedeutungen haben, wie in dieser Tabelle für verschiedene Interpretationen exemplarisch dargelegt.

Interpretation des Begriffs Energiedaten	Quelle
Am Hausanschlusspunkt erfasste Messwerte intelligenter Stromzähler	[Mol+10; KJ11; GJL12]
Auf Geräteebene erfasster Stromverbrauch	[KJ11; Mar+11; BCR12]
Messwerte intelligenter Gaszähler	[Coh+10]
Zu Präsentationszwecken aufbereitete Daten über den Energieverbrauch	[FSS10; VWD12; BCR12]
In Energiemanagementsystemen verwendete Daten	[LMW10]

Für den weiteren Verlauf dieser Arbeit soll daher folgende Definition für Energiedaten Anwendung finden.

Definition 1 (Energiedaten). *Energiedaten sind Daten, welche die folgenden Eigenschaften erfüllen:*

- *Die Daten enthalten Informationen zur Nutzung von Energie in Haushalten beziehungsweise Unternehmen oder zum Netzzustand beziehungsweise zum Zustand von Netzkomponenten.*
- *Die Daten liegen in Form von Zeitreihen vor.*

Die beiden vorausgesetzten Eigenschaften fassen zwar eine große Menge unterschiedlicher Daten zusammen, können aber bei der Abgrenzung zu anderen Arten von Daten, die einen Energiebezug aufweisen, hilfreich sein. Energiedaten beziehen sich demnach ausdrücklich nicht nur auf eine bestimmte Form von Energie wie beispielsweise Strom. Daten zur Nutzung thermischer Energie, wie im Bereich der Fernwärme oder Daten zum Verbrauch chemischer Energie wie bei Erdgas, sind entsprechend ebenfalls Energiedaten. Auch Daten zum Netzzustand wie beispielsweise die Netzfrequenz gehören dazu.

Auch wenn ihr Name suggeriert, dass Datenpunkte von Energiedaten über eine Energieeinheit wie beispielsweise Kilowattstunden verfügen sollten, werden sie hier nicht darauf beschränkt. So wären beispielsweise auch die durchschnittliche Leistung in einem angegebenen Zeitraum oder der Gasverbrauch in Kubikmetern während eines Zeitraums angemessene Einheiten. Die Art der Datenentstehung wird ebenfalls nicht genauer definiert. Dies erlaubt es, nicht nur tatsächlich gemessene Werte, sondern auch Ergebnisse aus Simulationen den Energiedaten zuzurechnen. Im Unterschied zu den Ergebnissen der Simulationen in Form von Zeitreihen gehören aber die Simulationsmodelle, aus denen die Daten entstanden sind, nicht zu Energiedaten, wenngleich sie unbestreitbar einen Energiebezug aufweisen.

Die in dieser Arbeit verwendete Definition von Energiedaten fordert, dass sie in Form von zeitlich geordneten Datenpunkten vorliegen müssen. Die zu den Datenpunkten gehörenden Zeitpunkte oder -räume müssen aber nicht zwangsläufig regelmäßig verteilt sein. Des Weiteren müssen die angegebenen Zeiten nicht absolut definiert sein, sondern können auch relativ angegeben werden. Somit sind beispielsweise neben zeitlich exakt eingeordneten Verbrauchsprofilen auch Geräteprofile abbildbar, bei welchen lediglich ein typischer Verlauf über die Zeit angegeben wird und bei welchen der Anfangszeitpunkt keine Rolle spielt.

Die obige Definition setzt ein Vorhandensein von Metadaten nicht voraus. Dennoch sei an dieser Stelle betont, dass Energiedaten mit Informationen über die enthaltenen Daten angereichert werden sollten. Insbesondere weil eine Vielzahl verschiedener Arten von Zeitreihen in unterschiedlichen Einheiten in die Kategorie der Energiedaten fällt, ist es umso wichtiger, dass die Art der Daten eindeutig gekennzeichnet ist.

2.2.3. Nutzen von Energiedaten

Energiedaten nehmen eine entscheidende Rolle beim Vorantreiben der Energiewende ein. Nicht nur Energieversorgungsunternehmen, sondern auch Verbraucher können von detaillierten, feingranularen Daten über den Energieverbrauch beziehungsweise die Energieproduktion profitieren. Energiedaten können auf verschiedenen Ebenen einen effizienteren und bewussten Umgang mit Energie ermöglichen [SK12].

2.2.3.1. Transparenz des Energieverbrauchs steigern

Für die meisten Verbraucher ist die Nutzung von Energie ein intransparenter Prozess. Energie ist ein abstrakter Begriff, dessen Bedeutung sich einem Großteil der Bevölkerung nicht erschließt [Dar06]. Insbesondere werden Einheiten wie Watt, Kilowattstunden oder auch Ausstoß von Kohlenstoffdioxid pro Kilowattstunde im Alltag kaum verwendet. Verbraucher orientieren sich viel eher an der Höhe der Stromrechnung am Ende einer Abrechnungsperiode [HNB10; Kar11].

In der heutigen hochtechnisierten Welt benutzen wir ganz selbstverständlich verschiedene Formen von Energie, ohne darüber bewusst nachdenken zu müssen. Das Umlegen eines Schalters sorgt für Beleuchtung im Haus, über einen einfachen Drehregler lässt sich die Temperatur des Backofens einstellen und mit einer drahtlosen Ladestation kann das Smartphone mit Energie versorgt werden. Dass Energie nicht „einfach so“ verfügbar ist, wird uns oft erst wieder durch einen Stromausfall ins Bewusstsein gerufen. Die unbewusste Nutzung ist zwar äußerst komfortabel, führt aber auch zur Verschwendungen von Energie. Darby *et al.* [Dar06] vermuten, dass alleine direkte Rückmeldungen über den Energieverbrauch ohne weitere Optimierungen Verhaltensänderungen bewirken können, die eine Energieersparnis von bis zu 15 % bedeuten. Zahlreiche weitere Autoren untersuchen die Einflussmöglichkeiten auf den Energieverbrauch von Haushalten, wobei meist Ersparnisse in ähnlicher Größenordnung angenommen werden [vgl. u.a. BL99; FSS10; Vas+12]. Hillemacher [Hil14] analysiert die Auswirkungen dynamischer Strompreissignale in Haushalten und betrachtet im Zuge dessen auch die Transparenzeffekte. Die aus dem Transparenzeffekt entstehende Einsparung liegt mit durchschnittlich 2,6 % jedoch niedriger als beispielsweise bei [Dar06] angegeben. Somit ist zu vermuten, dass die tatsächlich realisierbaren Einsparungen von zahlreichen weiteren Faktoren abhängen.

Häufig scheitern Versuche, Energie zu sparen, nicht an mangelnder Motivation, sondern an Unwissenheit oder falschen Einschätzungen bezüglich des eigenen Energieverbrauchs [CTG08; Att+10]. Die Möglichkeit, die Höhe der Stromrechnung zu reduzieren, bietet einen Anreiz, jedoch fehlt oftmals eine konkrete Rückmeldung, was Verhaltensänderungen tatsächlich bewirken können. Dementsprechend ist Verbrauchern der Zusammenhang zwischen ihrem

Verhalten und den Auswirkungen auf den Energieverbrauch unklar [Ste08; BN08; Kar11].

Faruqui *et al.* [FSS10] kritisieren diese Situation als nicht hinnehmbar und vergleichen den Stromverbrauch mit dem Benzinverbrauch von Fahrzeugen. Man stelle sich vor, eine Familie verfüge über zwei sehr unterschiedliche Fahrzeuge, von denen eines deutlich mehr Kraftstoff verbraucht als das andere. Keines der Fahrzeuge verfügt über eine Verbrauchsanzeige und die Autos können ohne direkt bezahlen zu müssen beliebig oft aufgetankt werden. Am Ende des Jahres erhält die Familie dann eine Gesamtrechnung über den Benzinverbrauch beider Fahrzeuge ohne jegliche Aufschlüsselung. Dieses Beispiel erscheint vollkommen absurd, beschreibt die Situation bezüglich Stromverbrauch aber sehr treffend.

In vielen weiteren Veröffentlichungen [vgl. u.a. Abr+05; VWD12; Vas+12; BCR12; WN03; WN07] wird die Bedeutung einer verbraucherorientierten Aufbereitung und Präsentation von Energiedaten hervorgehoben. Insbesondere wird betont, dass Verbraucher durchaus bereit sind, ihr Verhalten zu ändern, um Energie zu sparen. Ihnen ist für gewöhnlich aber nicht klar, was sie konkret tun können, um wirklich einen Unterschied zu bewirken. Direkte Rückmeldungen an den Verbraucher ermöglichen es ihm, einen Zusammenhang zwischen Verhalten und Energieverbrauch herzustellen. Energiedaten können, wenn sie denn richtig genutzt werden, unseren Energieverbrauch transparenter und für die breite Masse verständlich machen. Dies würde vermutlich zu einer Reduktion des Energieverbrauchs und zu einem bewussteren Umgang mit den noch vorhandenen Ressourcen führen. Somit sind Energiedaten in der Lage, ein wichtiger Treiber für die Umsetzung der Energiewende zu sein.

Doch nicht nur beim Sparen von Energie können Energiedaten helfen. Auch die effiziente Nutzung zum „richtigen“ Zeitpunkt wird durch Energiedaten unterstützt.

2.2.3.2. Energiedaten im Energiemanagement

Es ist nicht nur entscheidend, wie viel Energie verbraucht wird, sondern auch, zu welcher Zeit dies geschieht. Es ist zwar nur in begrenztem Maße möglich, unseren Energieverbrauch einzuschränken und zu reduzieren, aber

dafür kann ein gewisser Anteil zeitlich verschoben werden. Dies spielt im klassischen Stromnetz nur eine untergeordnete Rolle. Die Brennstoffe sind durchgehend verfügbar und werden früher oder später verbraucht, da die Energieproduktion der -nachfrage folgt. Für die Gesamtbilanz macht dies keinen Unterschied. In einem Energiesystem mit einem großen Anteil volatiler Erzeuger, die von erneuerbaren Energien abhängen, spielt der Zeitpunkt des Verbrauchs hingegen eine große Rolle [Qui09]: Energie aus Wind und Solarstrahlung ist beispielsweise nicht permanent verfügbar und kann dementsprechend auch nicht durchgehend genutzt werden. Im Umkehrschluss sollte diese Art von Energie dann genutzt werden, wenn sie verfügbar ist, um den Verbrauch fossiler Brennstoffe so weit wie möglich zu reduzieren (vgl. Abschnitt 2.1.2.2).

Energiemanagementsysteme können diese Aufgabe der Optimierung des Energieverbrauchs übernehmen. Durch geschicktes Planen können Lasten zeitlich so verschoben werden, dass verfügbare erneuerbare Energien möglichst effizient ausgenutzt werden und in Zeiten von Energiemangel möglichst wenig Energie benötigt wird, so dass die Reserven fossiler Brennstoffe geschont werden. Im Gegensatz zum klassischen Energiesystem gäbe es keine verbrauchsorientierte Produktion mehr, sondern einen produktionsorientierten Verbrauch. Bei umfangreichem Einsatz wäre dies darüber hinaus eine Möglichkeit, den kostenintensiven Ausbau von Energiespeichern zu reduzieren.

Die Güte der Verbrauchsanpassung ist allerdings in hohem Maße von den verfügbaren Informationen abhängig. Um den Verbrauch möglichst gut an die volatilen erneuerbaren Energien anpassen zu können, sind feingranulare, detaillierte Energiedaten nötig. Dies kann gegebenenfalls sogar die Anpassung der gesamten Informationsinfrastruktur im Energienetz erforderlich machen [Kim+10]. Allerdings kann nicht nur eine höhere zeitliche Auflösung die Qualität der Optimierung verbessern. Insbesondere kann auch eine Aufschlüsselung des Verbrauchs nach einzelnen Geräten einen signifikant besseren Einblick ermöglichen und damit die Basis für verbesserte Optimierungen bilden. Mit diesem Themenkomplex beschäftigt sich eine Vielzahl verschiedener Veröffentlichungen [vgl. u.a. KS14; Fan+12; Fan+13; MSW10; MR03; Qui09; SK12]. Die Anwendungsfälle müssen sich jedoch nicht nur auf Haushalte oder produzierende Industrie beschränken. In [Lin+11] stellen Lin

et al. beispielsweise ein Verfahren zur Betriebsoptimierung von Rechenzentren vor. Ziel der Optimierung ist es, die Computer entsprechend des Bedarfs an Rechenleistung an- beziehungsweise abzuschalten, um so den Energiebedarf zu reduzieren, ohne jedoch die Nutzbarkeit des Rechenzentrums einzuschränken.

Neben der reinen Optimierung können Energiedaten auch damit verbundene Anreizsysteme unterstützen oder sogar erst ermöglichen. Verbraucher haben nicht zwingend ein eigenes Interesse an der Durchführung einer Lastverschiebung. Aus diesem Grund sehen diverse Modelle Anreize vor, welche beispielsweise monetärer Natur sind. Da diese Anreize nur gewährt werden sollten, wenn eine Lastverschiebung tatsächlich durchgeführt wurde, bieten sich Energiedaten zur Überprüfung an [SK12; Qui09; Ber+11].

2.2.3.3. Verbesserte Prognosen durch Energiedaten

In klassischen Stromnetzen liegen den Energieversorgungsunternehmen in der Regel keine detaillierten Informationen zum aktuellen Zustand des Niederspannungsnetzes vor. Eine großflächige Erfassung und Übertragung von Messwerten findet nicht statt. Stattdessen wird der Energieverbrauch mithilfe verschiedener Standardlastprofile prognostiziert und die Stromproduktion dementsprechend geplant. Im Betrieb auftretende Differenzen werden auf höheren Netzebenen erfasst und ausgeglichen. Die Prognosegüte schlägt sich somit direkt in den Kosten für die nötige Regelenergie nieder. Dementsprechend ist es im Interesse der Energieversorgungsunternehmen, die Prognosen kontinuierlich zu verbessern.

Detaillierte Energiedaten, die großflächig erfasst werden, können einen Beitrag zur Verbesserung der Prognosen leisten. Mit repräsentativen Verbrauchsdaten von Verbrauchern verschiedener Kategorien können die existierenden Standardlastprofile aktualisiert werden. Dadurch könnte die generelle Prognosegüte verbessert werden. Trotzdem werden nach wie vor Abweichungen von diesen Prognosen zu beobachten sein. Kontinuierlich erfasste Energiedaten bieten aber eine Möglichkeit der direkten Rückmeldung bezüglich solcher Abweichungen. Die Energieversorger werden damit kurzfristig auch über lokale Effekte im Niederspannungsnetz informiert und können, falls erforderlich, gezielt gegensteuern [vgl. u.a. Mül10; KD12].

2.2.4. Gefahren bei der Verwendung von Energiedaten

Energiedaten weisen, wie im vorherigen Abschnitt gezeigt, einen unbestreitbaren Nutzen auf. Aber gerade die wertvollen in den Energiedaten enthaltenen Informationen machen diese zu einer Gefahr für die Privatsphäre:

„[...] the major benefit provided by the SG [smart grid], i.e. the ability to get richer data to and from customer meters and other electric devices, is also its Achilles' heel from a privacy viewpoint.“ [Nat14]¹

Die Nutzung von Energiedaten als „Guckloch in Haushalte“ [KS14] stellt eine ernste Gefahr für die Privatsphäre aller von einer Energiedatenerfassung betroffenen Personen dar. Jedoch erscheint die Situation auf den ersten Blick nicht sehr gravierend. Der Vergleich mit Smartphones und Kreditkarten, deren Nutzung für viele alltäglich ist, liegt nahe. Dass bei der Verwendung dieser und aller damit verbundenen Produkte in großem Stil personenbezogene Daten gesammelt und daraus detaillierte Profile erstellt werden, sollte den meisten Nutzern bekannt sein. Den entstehenden Komfort bezahlt ein Verwen-der nämlich nicht nur mit Geld, sondern auch mit seinen Daten [Sta10]. Im Ge-gensatz zur Nutzung von Energie kann man aber durchaus auf Mobiltelefone, Kreditkarten und Ähnliches verzichten. Selbstverständlich kann dies, je nach individueller Einstellung, mit großen Einschränkungen und Komforteinbußen verbunden sei. Auf die Nutzung von Energie in Form von Strom und Wärme zu verzichten, ist hingegen in der heutigen Welt fast ausgeschlossen [SK12]. Dementsprechend wäre von einer flächendeckenden Energiedatenerfassung nahezu jeder betroffen.

Rund um die Extraktion von Informationen aus Energiedaten hat sich ein eigener Forschungszweig entwickelt [vgl. u.a. ZR11]. Die vielfältigen Verfah-ren arbeiten auf unterschiedlichste Art und Weise. Dementsprechend ist es schwierig, Energiedaten gegen alle erdenklichen Angriffe zu schützen. Umso wichtiger ist jedoch eine eingehende Untersuchung potenzieller Gefahren, wie sie detailliert in der Bedarfsanalyse erfolgt (vgl. Abschnitt 4.2). Aus diesem

¹Diese Aussage wird unter anderem in [Fan+12] und [Kal+10] zitiert. In der aktuellen, dritten Revision des ursprünglich zitierten Artikels ist die entsprechende Passage jedoch ohne Angabe von Gründen nicht mehr vorhanden.

Grund sei für weitere Ausführungen zu den konkreten Auswirkungen auf die Privatsphäre insbesondere auf Abschnitt 4.2.3 verwiesen.

2.3. Energiedatenmanagement

In den letzten Jahren ist die Menge der erfassten Energiedaten stark gewachsen und es ist zu vermuten, dass sich dieser Anstieg in Zukunft weiter beschleunigt. Nicht zuletzt deshalb sind Systeme nötig, in welchen Energiedaten verwaltet werden können. Abhängig vom Kontext kann der Begriff Energiedatenmanagement jedoch unterschiedlich verstanden werden. In diesem Abschnitt werden daher die Begriffe Energiedatenmanagement beziehungsweise Energiedatenmanagementsystem definiert und potenzielle Nutzer solcher Systeme aufgezeigt.

2.3.1. Definition Energiedatenmanagement

Der Begriff Energiedatenmanagement kann, je nach Umfeld, eine unterschiedliche Bedeutung haben. An dieser Stelle wird er deshalb im Kontext dieser Arbeit eingeordnet und abgegrenzt. Die VDI-Richtlinie 4602 [vgl. VDI07; VDI13] zum Thema Energiemanagement enthält beispielsweise folgende Definition.

„Im Zusammenhang mit dem Energiehandel wird unter Energiedatenmanagement insbesondere die Bereitstellung, Übermittlung, Verarbeitung und Speicherung von Bilanzkreis-, Fahrplan- und Zählerdaten verstanden.“ [VDI07]

Diese Definition bezieht sich explizit auf den Energiehandel. Obwohl sich diese Arbeit vor allem mit Energiedaten, wie sie in Haushalten und Unternehmen anfallen, beschäftigt, lässt sich obige Definition als Grundlage für eine neue Definition im Rahmen dieser Arbeit heranziehen.

Definition 2 (Energiedatenmanagement). *Energiedatenmanagement umfasst die Bereitstellung, Übermittlung, Verarbeitung und Speicherung von Energiedaten gemäß Definition 1 aus Abschnitt 2.2.2 dieser Arbeit.*

Definition 3 (Energiedatenmanagementsystem). *Ein Energiedatenmanagementsystem ist ein Software- oder Workflowsystem, welches zum Zwecke des Energiedatenmanagements gemäß Definition 2 eingesetzt werden kann.*

An dieser Stelle sei betont, dass insbesondere im Kontext der Energieversorgungsunternehmen ein anderes Verständnis der Begriffe verbreitet ist. Aus dieser Perspektive interagiert ein Energiedatenmanagementsystem mit einem Zählerdatenmanagementsystem. Letzteres ist für die Aufbereitung und Speicherung der fernausgelesenen Energiedaten zuständig, während das Energiedatenmanagementsystem auf dieser Datengrundlage Analysen und Prognosen ermöglicht [Ern13]. Bei einem noch weiter gefassten Verständnis können in einem solchen System auch andere Kundendaten vorgehalten und gepflegt werden. Des Weiteren kann es bei der Marktkommunikation und der Interaktion mit anderen Marktpartnern unterstützen. Das eigentliche Verwalten der Daten von intelligenten Energiezählern erfolgt aber in den genannten Meter-Data-Managementsystemen, welche die eigentlichen Energiedatenmanagementsysteme entlasten sollen. Auf diesen Bereich spezialisierte Unternehmen bieten Softwareprodukte und Dienste an, die der Prozessunterstützung in dieser Sparte der Energiewirtschaft dienen. Exemplarisch seien hier [Hei16] und [Süd16] genannt. Im Gegensatz zum Energieversorgungssektor bezieht sich obige Definition explizit nicht auf die dort verwendeten ganzheitlichen Systeme. In dieser Arbeit sind mit Energiedatenmanagementsystemen solche gemeint, welche speziell auf die Verarbeitung von Energiedaten gemäß der Definition in Abschnitt 2.2.2 zugeschnitten sind. Die Anwenderzielgruppe ist ebenfalls abweichend. Typische Verwender solcher Systeme wären nicht Energieversorgungsunternehmen, sondern einerseits Privatpersonen oder Unternehmen, deren Energiedaten verwaltet und geschützt werden sollen, und andererseits externe Parteien, welche Interesse an den Daten haben und auf diese zugreifen wollen. Tiefergehende Ausführungen zu potenziellen Nutzern folgen im nächsten Abschnitt dieser Arbeit.

2.3.2. Nutzer von Energiedatenmanagementsystemen

Bei Energiedaten handelt es sich in aller Regel um sensible und schützenswerte Daten (vgl. Abschnitte 2.2.2 und 2.2.4). Von diesen Daten betroffen sind in

der Regel Privatpersonen oder Unternehmen. Auf der anderen Seite stehen Interessenten, die die Daten erhalten und zu einem gewissen Zweck auswerten möchten. Somit lassen sich die Nutzer von Energiedatenmanagementsystemen in zwei Lager einteilen. Auf der einen Seite diejenigen, deren Daten erfasst und verwaltet werden und auf der anderen Seite Dateninteressenten, die Zugriff auf die Energiedaten erhalten möchten.

2.3.2.1. Betroffene der Energiedatenerfassung

Personen, deren Energiedaten erfasst werden, haben ein besonderes Interesse daran, dass ihre Daten geschützt werden. Aus diesem Grund sollte ein Energiedatenmanagementsystem von Anfang an so konzipiert werden, dass die Interessen der Betroffenen gewahrt werden [MR03]. Dies gilt insbesondere für technisch weniger versierte Privatpersonen, in deren Haushalten Energiedaten erfasst werden. In der Literatur ist dies auch die am stärksten vertretene Nutzergruppe [vgl. u.a. EE15; GJL12; KD11]. Diese Arbeit konzentriert sich ebenfalls auf diesen Personenkreis, beschränkt sich aber nicht ausschließlich darauf.

Einige der in dieser Arbeit vorgestellten Konzepte lassen sich auch auf andere Bereiche übertragen. Falls Energiedaten beispielsweise auch am Arbeitsplatz erfasst werden, sind dort ebenfalls personenbezogene Daten zu erwarten. Detaillierte Aufzeichnungen, wie viel Strom in Büroräumen verbraucht wird, lassen umfangreiche Schlüsse über die Angestellten zu. Anhand des Stromverbrauchs lassen sich die Arbeits- und Pausenzeiten ebenso kontrollieren wie die Nutzung von Kopierern, Kaffeemaschinen und Ähnlichem. Daher können auch Mitarbeiter oder der Betriebsrat eines Unternehmens, welches Energiedaten erfasst, zu potenziellen Nutzern eines Energiedatenmanagementsystems gezählt werden.

Nicht nur die Mitarbeiter, sondern auch die Unternehmen selbst können ein Interesse am Schutz der Energiedaten haben. Falls die erfassten Daten Rückschlüsse auf Geschäfts- oder Produktionsprozesse erlauben, muss ihr Schutz gegenüber Außenstehenden gewährleistet werden. Dies kann vor allem bei energieintensiven Prozessen zum Problem werden, da dort ein Stillstand der Produktion sehr einfach am Energieverbrauch erkennbar ist. Bei einer

entsprechend fein aufgelösten Messung ist aber durchaus auch denkbar, dass Prozessparameter aus den Energiedaten abgeleitet werden könnten.

Die beschriebenen Gruppen (Privatpersonen in Haushalten, Mitarbeiter eines Unternehmens, Unternehmen selbst) werden in dieser Arbeit als *Nutzer* eines Energiedatenmanagementsystems bezeichnet. Sie können ein solches System verwenden, um sie betreffende Energiedaten zu verwalten und zu schützen. Die Interessen dieser Nutzer sind in Abschnitten 4.2 und 4.3 Gegenstand detaillierter Untersuchung.

2.3.2.2. Energiedateninteressenten

Die gespeicherten Energiedaten können für verschiedene Zwecke analysiert und weiterverwendet werden. Sie weisen, wie in Abschnitt 2.2.3 gezeigt, einen vielfachen Nutzen auf. Dementsprechend können nicht nur die Personen, welche von der Erfassung der Energiedaten betroffen sind, sondern auch dritte Parteien ein Interesse daran haben. In Abgrenzung zum erläuterten Nutzerbegriff werden Angehörige dieser dritten Parteien im Folgenden als *Dateninteressenten* bezeichnet.

Der Wunsch der Dateninteressenten, Zugriff auf die gespeicherten Energiedaten zu erhalten, kann begründet oder unbegründet sein und dem Wunsch der betroffenen Personen (der Nutzer) entsprechen oder auch nicht. Deshalb sollte ein nutzerorientiertes Energiedatenmanagementsystem es ermöglichen, die gespeicherten Energiedaten den Nutzerwünschen entsprechend zu schützen und zu verwalten. Der Interessenkonflikt von Nutzern und Dateninteressenten sowie die daraus resultierenden Anforderungen werden in Abschnitt 4.2 beleuchtet.

Verwandte Arbeiten

Große Mengen von Energiedaten fallen nicht nur in Unternehmen, sondern auch in Privathaushalten an. Obwohl mit einer steigenden Menge erfasster Daten deren Verwaltung aufwendiger wird, werden die Daten nicht immer in eigenen Systemen vorgehalten. Oftmals handelt es sich nicht um spezialisierte Energiedatenmanagementsysteme, sondern um Energiemanagementsysteme, in welchen Energiedaten als Nebenprodukt anfallen und behandelt werden. Dennoch existieren auch Ansätze und Systeme, welche sich im Speziellen dem Umgang mit Energiedaten widmen. Typische Vertreter aus Forschung und Wirtschaft werden im ersten Teil dieses Kapitels vorgestellt. Im zweiten Teil werden Arbeiten zum Schutz von Energiedaten besprochen. Diese umfassen regulatorische Maßnahmen ebenso wie technologische Ansätze.

3.1. Umgang mit Energiedaten und deren Management

Energiedaten werden je nach Anwendungsfall in sehr unterschiedlichen Systemen vorgehalten. Die Spanne reicht von einfachen textbasierten Ansätzen bis hin zu verteilten Data-Warehouse-Systemen. Nicht zuletzt ist dies auch abhängig vom geplanten Verwendungszweck und von der Art und Größe der Datenbasis. Im weiteren Verlauf werden die unterschiedlichen Ansätze in zwei Gruppen eingeteilt. Zunächst werden Ansätze aus der Forschung betrachtet und anschließend werden Systeme und Produkte aus der Wirtschaft vorgestellt.

3.1.1. In der Forschung

Energiedatensätze von kleinerem Umfang lassen sich durchaus auch in Textdateien oder einfachen relationalen Datenbanken speichern. Spezialisierte Ansätze bieten sich hingegen bei großen Datensätzen oder für aufwendige Analysen an. Obwohl zahlreiche Veröffentlichungen zur Extraktion von Informationen aus Energiedaten beziehungsweise zum Schutz davor existieren, wurde bisher recht wenig zum Management von Energiedaten veröffentlicht. Eine mögliche Erklärung wäre, dass Energiedaten häufig nicht in dedizierten Systemen erfasst und verwaltet werden. Vielmehr werden sie oft als notwendiges Nebenprodukt der eigentlichen Vorgänge betrachtet und in übergeordneten Systemen – beispielsweise Energiemanagementsystemen – gespeichert. Darüber hinaus sind viele Datensätze mit üblichen Standardmethoden noch gut handhabbar, so dass in manchen Anwendungsfällen dedizierte Systeme nicht nötig zu sein scheinen. Trotzdem sollen im Folgenden Ansätze mit jeweils unterschiedlichen Zielen vorgestellt werden.

3.1.1.1. Der Referenzdatensatz für Energiedaten: REDD

Der Forschungsbereich des Energiemonitorings (vgl. Abschnitt 4.2) hatte lange Zeit keinen einheitlichen Datensatz, den alle Beteiligten verwenden konnten, um ihre Ergebnisse zu vergleichen. Mit der Veröffentlichung des *Reference Energy Disaggregation Data Set (REDD)* [KJ11] verfolgen Kolter *et al.* das Ziel, diese Lücke zu füllen.

Der Datensatz enthält real gemessene (nicht simulierte) Energieverbrauchsdaten aus mehreren Haushalten. In diesen wurde nicht nur der Gesamtverbrauch am Hausanschlusspunkt gemessen, sondern auch detailliertere Messungen im Haus vorgenommen. Diese Daten liegen mit einer Auflösung von etwa 1 Hz vor. Darüber hinaus werden in *REDD* weitere Messdaten mit sehr hoher zeitlicher Auflösung zur Verfügung gestellt. Diese Hochfrequenzdaten erlauben eine detaillierte Analyse der eigentlich sinusförmigen Spannung beziehungsweise der gemessenen Abweichungen von der angenommenen Idealform.

Der Datensatz enthält damit exemplarische Daten, wie sie in verschiedenen Algorithmen verwendet werden, die die Informationsextraktion aus Ener-

3.1. Umgang mit Energiedaten und deren Management

giedaten zum Ziel haben (vgl. Abschnitt 4.2.2). Die Güte der zu untersuchenden Algorithmen kann mithilfe der Daten überprüft werden, indem die Ergebnisse der errechneten Lastgangzerlegung mit den real gemessenen Einzelverbräuchen verglichen werden. Dies ist nicht nur für Messdaten vom Hausanschlusspunkt mit niedriger Messfrequenz möglich, sondern auch für Hochfrequenzdaten, welche in einigen Verfahren benötigt werden.

Die umfangreichen Messdaten werden durch Kolter *et al.* im Internet angeboten und können kostenfrei heruntergeladen werden (vgl. [KJ11]). Um eine einfache Verwendung für eine Vielzahl von Anwendungen zu ermöglichen, liegen die Datensätze in einer ordnerbasierten Struktur vor. Metadaten sind über Ordner- beziehungsweise Dateinamen codiert. Die einzelnen Messpunkte der niederfrequenten Messungen bestehen damit lediglich aus einem Zeitstempel und einer Gleitkommazahl. Die Messpunkte der hochfrequenten Messungen bestehen wiederum aus einem Zeitstempel und mehreren Gleitkommazahlen je Zeitschritt. Obwohl die Daten in dieser Form den Nachteil aufweisen, nicht direkt weiterverarbeitet werden zu können, ist eine Umwandlung in geeignete Formate einfach und schnell möglich. Eine Anwendung in unterschiedlichen Systemen ist damit mit relativ wenig Aufwand möglich. Nicht zuletzt deshalb bietet sich dieser Realdatensatz als Referenz für bestehende und neue Algorithmen an.

3.1.1.2. Ein Data Warehouse für Energiedaten: MIRABEL DW

Im EU-Projekt *MIRABEL* wird ein datengetriebener Ansatz für den Lastausgleich in elektrischen Netzen verfolgt. Die Daten umfassen unter anderem Informationen zu Energieverbrauch und -produktion zu einem bestimmten Zeitpunkt, sowie Angebote der Teilnehmer diesbezüglich. Diese Daten werden verwendet, um über einen marktähnlichen Mechanismus einen Ausgleich im Netz planen und herstellen zu können. Technisch gesehen werden die im Projekt anfallenden Daten im Rahmen eines Data-Warehouse, *MIRABEL DW* genannt, vorgehalten und verwaltet [STP12; Boe+12].

Für das Data-Warehouse wurde ein Datenbankschema ausgearbeitet. Dieses sieht Zeitreihen für den Verbrauch und die Produktion vor. Auch die Angebote diesbezüglich werden im Data-Warehouse gespeichert. Die Datensätze werden jeweils Nutzerrollen beziehungsweise Aktoren zugeordnet. Anfänglich

werden die Daten lediglich auf einem einzelnen Rechner vorgehalten, auch wenn konzeptuell geplant ist, sie über Millionen von Rechenknoten verteilen zu können. Eine dem Data-Warehouse angepasste Abfragesprache ermöglicht es, Gebote und Zeitreihen abzufragen, wobei auch diverse Filtermöglichkeiten vorgesehen sind. Die Beantwortung aufwendiger Abfragen benötigt mehrere Minuten, sobald die Zahl der betroffenen Datenpunkte in die Millionen geht. Datenschutz und Privatsphäre werden in *MIRABEL DW* nicht behandelt. Dies ist als kritisch anzusehen, da umfangreiche personenbezogene Daten in einem Datenbanksystem vorgehalten werden, welches komplexe Anfragen ermöglicht und damit gegebenenfalls Rückschlüsse auf individuelle Personen zulässt.

Bei *MIRABEL DW* handelt es sich um ein Data-Warehouse, welches auf wenige, spezielle Anwendungen zugeschnitten ist. Im Gegensatz zu *REDD*, welches einen möglichst flexiblen, dabei aber gegebenenfalls ineffizienten Ansatz darstellt, ist *MIRABEL DW* hochspezialisiert und für den vorgesehenen Zweck effizient, dabei aber gleichzeitig komplex und wenig flexibel. Die beiden Ansätze repräsentieren damit die Bandbreite verschiedener Ansätze der Energiedatenhaltung in der Forschung. Gleichzeitig wird aber auch deutlich, dass insbesondere im Bereich der Energiedaten spezialisierte Ansätze erst dann verfolgt werden, wenn die entsprechenden Forschungsfragen es erfordern.

3.1.2. In der Wirtschaft

Energiedaten stellen für Unternehmen oft ein wirtschaftliches Gut dar, da deren Analyse umfangreiche Erkenntnisse bezüglich der betroffenen Haushalte und Personen zulässt. Es ist das Geschäftsmodell einiger Unternehmen, die Daten auszuwerten und die Erkenntnisse profitbringend zu veräußern. Dementsprechend ist der Umgang mit Energiedaten in der Wirtschaft auch ein anderer als in der Forschung.

3.1.2.1. Zugriff auf Energiedaten für alle: Green Button

Die *Green Button initiative* wurde im September des Jahres 2011 in den USA ins Leben gerufen [vgl. Nat15; Dep16; Gre16]. Mit der Initiative folgen meh-

3.1. Umgang mit Energiedaten und deren Management

rere Dutzend Energieversorger dem Aufruf, ihren Endkunden detaillierte Energieverbrauchsdaten zur Verfügung zu stellen. Die Aufforderung ging maßgeblich vom Weißen Haus aus und wurde innerhalb weniger Monate umgesetzt, sodass bereits im Januar des Jahres 2012 der Betrieb offiziell aufgenommen werden konnte. Nach einem äußerst schnellen Wachstum nehmen mittlerweile über 60 Millionen Haushalte und Unternehmen sowie mehr als 50 Energieversorger in den USA an dem Programm teil [SP13; Gre16].

Erklärtes Ziel der Initiative ist es, den Endkunden einfach und sicher Energiedaten zur Verfügung zu stellen, sowie diese auf Wunsch durch Dritte analysieren zu lassen. Die Daten sind sowohl in menschenlesbarer Form als auch in maschinenlesbaren Formaten verfügbar. Nach einer Authentifizierung ist es den Kunden möglich, über einen im wörtlichen Sinne „grünen Knopf“ auf einer Webseite die zu dem jeweiligen Haushalt gehörigen Daten zu betrachten und herunterzuladen. Darüber hinaus können Nutzer von *Green Button* die Daten nicht nur online betrachten und herunterladen, sondern auch diversen Diensten erlauben, auf diese zuzugreifen. Als Dienstleistungen beworben werden unter anderem die Überprüfung des Energieverbrauchs auf Auffälligkeiten, Programme zur Effizienzsteigerung oder die Bestätigung gesteckter Effizienzziele [Gre16].

Der *Green Button standard* schreibt vor, dass in den Daten keine Informationen enthalten sein dürfen, welche es ermöglichen, die betroffenen Personen zu identifizieren. Lediglich gemessene Verbrauchsdaten dürfen auf Wunsch an Dritte übertragen werden [Gre16]. Diese Aussage darf allerdings unter keinen Umständen als Garantie für absoluten Datenschutz und Privatsphäre angesehen werden. Sobald ein Nutzer sich authentifiziert, kann er die zu seinem Haushalt gehörigen Daten betrachten. An dieser Stelle findet demzufolge offensichtlich eine Zuordnung von Haushalten und Datensätzen statt. Selbst wenn die Stelle, an welcher die Datensätze zugeordnet werden, vor dem Zugriff unbefugter Dritter geschützt werden kann, ist nicht ausgeschlossen, dass der Betreiber der Plattform nach wie vor auf diese Datenbank zugreifen kann. Sogar für den Fall, dass eine direkte Zuordnung der Energiedaten zu persönlichen Informationen technisch ausgeschlossen wäre, ist dies nicht zwangsläufig ein ausreichender Schutz. Bei entsprechenden Datenmengen und der Zuhilfenahme weiterer Informationsquellen ist eine Zuordnung von Energiedaten zu Personen durchaus möglich [JJR11]. Nutzer können dies im

Fall von *Green Button* technisch nicht verhindern. Sie haben zwar die Möglichkeit, die Daten an keine dritte Partei weiterzuleiten, dem Betreiber der Plattform kann der Zugriff jedoch effektiv nicht verboten werden. Den Nutzern wird damit eine Sicherheit suggeriert, welche zwar gegenüber Dritten gegeben sein mag, gegenüber den Betreibern aber nicht.

Eine derart große Datensammlung ist eine Verlockung nicht nur für Geheimdienste und Polizeibehörden, sondern auch für Kriminelle. Selbst wenn die Betreiber mit den besten Absichten handeln, kann nie ausgeschlossen werden, dass die Daten nicht doch missbraucht werden [vgl. u.a. LW08; LMW10; SK12]. Die zahlreichen Skandale rund um oft illegal bei Hackerangriffen erlangte und anschließend veröffentlichte Daten zeigen, dass selbst bei großem Interesse der Betreiber an einem Schutz der Daten nicht ausgeschlossen werden kann, dass Unbefugte Zugriff erhalten. Sobald die Daten aber die Kontrolle der Nutzer verlassen haben, ist Missbrauch Tür und Tor geöffnet [Sta10].

Zweifelsfrei ist zu begrüßen, dass Haushalte Zugriff auf ihre Energiedaten erhalten sollen, um sie ihren Wünschen entsprechend nutzen zu können. Mit *Green Button* ist damit ein erster Schritt in die richtige Richtung getan worden. Allerdings sollte ein solches System dem Nutzer die Entscheidung ermöglichen, ob die ihn betreffenden Daten überhaupt an eine zentrale Sammelstelle weitergegeben werden dürfen. Da dies nicht geschieht, ist die Initiative insbesondere aus Datenschutzgesichtspunkten nicht nur positiv zu bewerten.

3.1.2.2. Energiedaten von intelligenten Stromzählern auf Bestellung: *Discovery*

Das Unternehmen *Discovery* GmbH ist ein Messstellenbetreiber, welcher intelligente Stromzähler und damit verbundene Dienste anbietet [Dis16]. Die von dem deutschen Unternehmen *EasyMeter* GmbH hergestellten Stromzähler werden dem Endkunden gegen eine Einmalzahlung und eine anschließende monatliche Gebühr zur Verfügung gestellt.

Die im Sekundenbereich erfassten Verbrauchsdaten werden nicht nur zu Abrechnungszwecken gespeichert und übertragen, sondern auch dem Endkunden über ein Onlineportal präsentiert. Ziel des Unternehmens ist es nach eigenen Angaben, dem Kunden den eigenen Stromverbrauch näherzubringen. Zu diesem Zweck hat der Kunde Einblick in aktuelle und historische Verbräu-

che, erhält Hinweise, welche Geräte für den Stromverbrauch verantwortlich sein können und ähnliches. Darauf hinaus wirbt *Discovergy* damit, Kunden Handlungsempfehlungen bei einem anstehenden Tarifwechsel zu geben, so dass der Kunde Geld sparen kann. In Summe soll der Kunde so nicht nur genauer über seinen Stromverbrauch Bescheid wissen, sondern auch Energie und Geld sparen können.

Trotz der vermeintlich guten Absichten bezüglich Energiesparen wurde *Discovergy* in der Literatur massiv kritisiert. Greveler *et al.* [GJL12] haben den Netzwerkverkehr des intelligenten Zählers von *Discovergy* näher untersucht. Das Gerät wird an einen im Haushalt vorhandenen Router angeschlossen und tauscht mit den vom Unternehmen betriebenen Servern Daten über das Internet aus. Zum Zeitpunkt der Auswertungen durch Greveler *et al.* im Jahr 2011 erfolgte die gesamte Übertragung unverschlüsselt und unsigniert. Die Daten wurden in einem Textformat ohne jeglichen Schutz an die *Discovergy*-Server übertragen. Insbesondere wäre es möglich gewesen, den Netzwerkverkehr abzufangen und eigene, gefälschte Daten an die Server zu schicken. Dadurch hätte dem Messstellenbetreiber auf einfache Art und Weise ein gefälschter Verbrauch untergeschoben werden können. Falls die Identifikationsnummern anderer Zähler bekannt gewesen wären oder hätten geraten werden können, wären auch deren Verbrauchsdaten angreifbar gewesen.

Zusätzlich zu diesem Problem zeigen Greveler *et al.*, dass die gemessenen Daten ausreichend detailliert sind, um die in einem Haushalt konsumierten Filme zu identifizieren. Der Stromverbrauch eines modernen Fernsehers schwankt mit der Bildhelligkeit, so dass bei bekannten Verbrauchsprofilen von Filmen eine Zuordnung möglich wird. Dadurch stellt eine derartig feingranulare Erfassung von Energiedaten auch schon am Hausanschlusspunkt eine Gefahr für die Privatsphäre dar. Selbst falls die Daten bei ihrer Übertragung und Speicherung geschützt werden, muss der Nutzer dem Messstellenbetreiber vertrauen, dass dieser die Daten nicht missbraucht.

Auch wenn die Datenübertragung bei *Discovergy*-Zählern zwischenzeitlich vermutlich dem Stand der Technik entspricht und entsprechend gesichert ist, zeigt der Vorfall doch gut das Problem diverser intelligenter Stromzähler. Sollte durch neu entdeckte Sicherheitslücken oder auch durch Nachlässigkeit des Betreibers die Sicherheit der Datenübertragung nicht mehr gewährleistet werden können, führt dies direkt zu einer massiven Gefährdung der Privatsphäre.

Insbesondere dann, wenn Daten nicht nur mit einer geringen zeitlichen Auflösung im Stunden- oder Minutenbereich, sondern eventuell gar im Bereich weniger Sekunden erfasst werden.

Neben *Discovergy* gibt es in Deutschland weitere Unternehmen, welche intelligente Stromzähler und damit verbundene Dienstleistungen in Form von Datenaufbereitung und -präsentation anbieten. Die Vorteile für den Kunden und die damit verbundenen Gefahren und Nachteile unterscheiden sich bei den meisten Produkten nur wenig. Es ist zu vermuten, dass dieses Geschäftsfeld in Zukunft verstärkt von Energieversorgern genutzt werden wird, um sich im Markt abzuheben und Kunden zu binden. Dementsprechend dürfte die Zahl an Anbietern in Zukunft weiter steigen.

3.1.2.3. Energiedaten von Zwischensteckdosen: **Plugwise**

Das niederländische Unternehmen *Plugwise* entwickelt und vertreibt seit dem Jahr 2006 drahtlose Energiemanagement- und Kontrollsysteme. Diese Systeme sollen einen besseren Einblick in den Energieverbrauch ermöglichen und damit eine Energieersparnis zwischen 10 % und 40 % mit sich bringen [Plu16].

Plugwise bietet unterschiedliche Gerätekästen zur Messung des Energieverbrauchs an, darunter unter anderem intelligente Zwischensteckdosen, sogenannte *Circles*. Diese erfassen den Stromverbrauch der angeschlossenen Geräte und können die gemessenen Daten drahtlos übertragen. Je nach Paket handelt es sich bei dem Empfänger um ein eigenständiges Gerät, *Stretch* genannt, oder einen Empfänger mit USB-Anschluss, der nicht eigenständig funktioniert und an einem Computer betrieben werden muss. Die Steckdosen können über entsprechende Programme nicht nur ausgelesen werden, sondern ermöglichen es auch, angeschlossene Geräte an- beziehungsweise auszuschalten.

Das *Stretch*-Gerät erfasst die gemessenen Daten der einzelnen Zwischensteckdosen und überträgt diese auch an einen Server der Firma *Plugwise*. Die Daten können dann über unterschiedliche Anwendungen vom Nutzer angesehen und ausgewertet werden. Auch wenn *Plugwise* angibt, dass die Daten Dritten nicht zugänglich gemacht werden, verlassen sie dennoch die Kontrolle des Nutzers. Darüber hinaus ist der mobile Zugriff auf die Daten lediglich für ein halbes Jahr kostenfrei möglich. Im Anschluss an diese Testphase muss der

3.1. Umgang mit Energiedaten und deren Management

Kunde eine geringe monatliche Gebühr entrichten, um weiter mobil auf die Daten zugreifen zu können.

Im Gegensatz zum *Stretch*-Gerät funktioniert der USB-Empfänger nur an einem Computer und nicht eigenständig. Das von *Plugwise* angebotene Programm *Source* erfasst über den USB-Empfänger die gemessenen Daten und erlaubt es, die Zwischensteckdosen zu schalten. Die empfangenen Daten werden auf dem verwendeten Rechner in einer lokalen Datenbank gespeichert und können ausgewertet werden. Ebenso ist es möglich, die Steckdosen spontan oder nach Zeitplänen zu schalten. Dies soll in Summe eine Reduktion des Energieverbrauchs ermöglichen. Problematisch ist, dass ein Nutzer zur Annahme verleitet werden könnte, dass bei Nutzung des USB-Empfängers keine Daten das Haus verlassen. Im Lizenzvertrag des Programms *Plugwise Source* ist unter anderem der Abschnitt *Privacy and use of personal and other data* mit folgendem Inhalt zu finden.

„Plugwise uses the data processed by the Software for scientific, statistic and historic purposes and in order to expand and improve its products and services. Unless it has received separate permission, Plugwise will compile only technical data (including usage data) that cannot be traced back to natural persons in the User's household and/or business.“ [Plu14]

Demzufolge werden technische Daten inklusive „*usage data*“ an *Plugwise* gesendet, wenn der USB-Empfänger verwendet wird. Zwar wird im Lizenzvertrag versichert, dass die Daten nicht wieder Haushalten zugeordnet werden können. Es wird jedoch in der Literatur gezeigt, dass eine Zuordnung durchaus noch möglich sein kann, auch wenn keine direkt identifizierenden Merkmale mehr vorhanden sind [JJR11]. Auch wenn – wie bereits zuvor beschrieben – mit bester Absicht gehandelt wird, besteht darüber hinaus stets die Gefahr, dass die Daten in die Hände von unbefugten Dritten gelangen, die die Daten missbrauchen könnten. Diese und weitere Gründe könnten Nutzer folglich dazu bewegen, verhindern zu wollen, dass Daten das Haus und damit die Kontrolle der Nutzer verlassen. Auch diesbezüglich findet sich ein Abschnitt in den Nutzungsbedingungen.

„Part of the right of use is for Plugwise to automatically receive the technical anonymous usage data from the User in exchange for

(an automatic notification on) free Software updates, if and when they become available. The User will be excluded from the right to receive free Software updates in the event the User, for whatever reason, prevents Plugwise to receive the data [...].“ [Plu14]

Verhindert ein Nutzer die Weitergabe der Daten, verliert er somit das Recht Aktualisierungen des Programms nutzen zu können. Falls der Nutzer seine Daten nicht weitergeben, aber trotzdem Aktualisierungen beziehen möchte, ist eine jährliche Gebühr zu entrichten.

„[...] the User may purchase Software updates for a yearly fee of 7,5 % of the cost of the Software and the Hardware (firmware).“ [Plu14]

Bei einem Preis von etwa 300 Euro für ein Starterpaket mit neun *Circles* (Stand 2016) handelt es sich somit um eine jährliche Gebühr von circa 20 Euro, die ein Nutzer entrichten müsste, um einerseits Aktualisierungen zu erhalten und andererseits seine Daten nicht weitergeben zu müssen. Aus Sicht des Unternehmens ist es durchaus zu verstehen, dass der Verkauf der Hardware als Geschäftsmodell möglicherweise nicht ausreicht. Dementsprechend ist es auch nachvollziehbar, dass die erfassten Daten ausgewertet werden sollen. Aus Nutzersicht hingegen ist dies kritisch zu bewerten, da es einen Eingriff in die Privatsphäre darstellen kann, auch wenn die Daten in anonymisierter Form erfasst werden. Dementsprechend ist es positiv, dass überhaupt die Möglichkeit existiert, die Weitergabe der Daten zu unterbinden, auch wenn dies mit einer jährlichen Gebühr belegt ist.

Alternativ besteht die Möglichkeit, auf die von *Plugwise* zur Verfügung gestellte Software komplett zu verzichten. Es existiert eine Reihe von Bibliotheken für unterschiedliche Programmiersprachen, welche einen Zugriff auf die *Circles* ermöglichen¹. Mit diesen können Eigenentwicklungen die Zwischensteckdosen nutzen, so dass es in der Hand des Programmierers liegt, was mit den Daten geschieht.

¹vgl. bspw. <https://github.com/SevenW/Plugwise-2-py>

3.2. Schutz von Energiedaten

Der Schutz von Energiedaten kann auf vielerlei Art und Weise erfolgen, wobei für gewöhnlich eine Einteilung in zwei Kategorien vorgenommen wird: Einerseits gibt es regulatorische Maßnahmen, die den Umgang mit Energiedaten einschränken und andererseits existieren technologische Vorkehrungen, die Energiedaten aktiv schützen sollen [Eng13; Fan+13; JKD12]. Im Folgenden werden neben den wichtigsten regulatorischen Ansätzen auch typische Vertreter der technologischen Schutzmethoden besprochen.

3.2.1. Regulatorische Ansätze

Unter regulatorischen Ansätzen werden an dieser Stelle Gesetze und Richtlinien sowie Rahmenwerke verstanden, welche den Umgang mit Daten im Allgemeinen oder Energiedaten im Speziellen regeln.

3.2.1.1. Das deutsche Bundesdatenschutzgesetz

Bezüglich Datenschutz ist in Deutschland zunächst das Bundesdatenschutzgesetz zu nennen, welches sich nicht explizit auf Energiedaten bezieht, sondern allgemein für personenbezogene Daten gilt. Von besonderer Relevanz ist zunächst Paragraph 3a.

„Die Erhebung, Verarbeitung und Nutzung personenbezogener Daten und die Auswahl und Gestaltung von Datenverarbeitungssystemen sind an dem Ziel auszurichten, so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen. Insbesondere sind personenbezogene Daten zu anonymisieren oder zu pseudonymisieren, soweit dies nach dem Verwendungszweck möglich ist und keinen im Verhältnis zu dem angestrebten Schutzzweck unverhältnismäßigen Aufwand erfordert.“

Der erste Teil des Paragraphen wird für gewöhnlich unter den Begriffen der Datenvermeidung oder Datensparsamkeit zusammengefasst. Dem Ziel, die Menge an personenbezogenen Daten so weit wie möglich einzuschränken, folgen im Bereich der intelligenten Stromnetze aber nur sehr wenige. Es ist im

Gegenteil sogar ein Trend zu einer größeren Durchdringung mit Messgeräten und einer verstärkten Datensammlung abzusehen. Dies geht so weit, dass das im intelligenten Stromnetz entstehende Netzwerk in Bezug auf Größe und Mächtigkeit mit dem heutigen Internet verglichen wird [Fan+13]. Eine solche Entwicklung dürfte es gemäß Bundesdatenschutzgesetz zumindest in Deutschland eigentlich nicht geben.

Insbesondere wird in diesem Gesetz neben der Datensparsamkeit auch eine Zweckgebundenheit der erfassten Daten gefordert. Paragraf 28 enthält dazu unter anderem folgenden Passus.

„Das Erheben, Speichern, Verändern oder Übermitteln personenbezogener Daten oder ihre Nutzung als Mittel für die Erfüllung eigener Geschäftszwecke ist zulässig, [...]. Bei der Erhebung personenbezogener Daten sind die Zwecke, für die die Daten verarbeitet oder genutzt werden sollen, konkret festzulegen.“

Dementsprechend ist die Verwendung personenbezogener Informationen wie beispielsweise Energiedaten für Geschäftszwecke unter gewissen Bedingungen durchaus zulässig. Allerdings sind die konkreten Zwecke bereits bei der Erhebung der Daten festzulegen. Bei den aktuellen Entwicklungen im intelligenten Stromnetz darf bezweifelt werden, dass der Zweck im Vorfeld in allen Fällen konkret definiert wurde. Vielmehr scheint es einen Trend zum Sammeln von Daten zu geben, ohne genau zu wissen, auf welche Weise und zu welchem konkreten Zweck die Daten verwendet werden sollen [vgl. u.a. Fan+13]. Problematisch ist weiterhin, dass die betroffenen Personen nicht sicher sein können, auf welche Art die Daten wirklich genutzt werden, selbst wenn ein konkreter Zweck im Vorfeld definiert wurde. Sobald die Daten die Kontrolle der Personen einmal verlassen haben, kann unter normalen Umständen ein Missbrauch nicht mehr verhindert werden [Sta10].

3.2.1.2. Das BSI-Schutzprofil

Das Bundesministerium für Wirtschaft und Energie hat das Bundesamt für Sicherheit in der Informationstechnik (BSI) mit der Ausarbeitung von Anforderungen an ein geeignetes intelligentes Messsystem beauftragt. Die Entwicklung umfasst Anforderungen an die folgenden Aspekte [Bun15b].

- Vertrauenswürdige Produktkomponenten (Smart Meter Gateway mit integriertem Sicherheitsmodul)
- Deren sicherer IT-Betrieb (Administration)
- Vertrauenswürdige Kommunikationsinfrastruktur (Smart Metering PKI)

Das sogenannte Smart-Meter-Gateway-Schutzprofil beschreibt Bedrohungsszenarien, die ein Messsystem beziehungsweise ein Smart-Meter-Gateway betreffen können. Dementsprechend werden Mindestanforderungen an Sicherheitsmaßnahmen definiert [Bun15b]. Diese Anforderungen sind im zum Schutzprofil gehörigen Dokument „Schutzprofil für die Kommunikationseinheit eines intelligenten Messsystems für Stoff- und Energiemengen“ (BSI-CC-PP-0073) [Bun14] ausgearbeitet. Neben einer Trennung der verschiedenen Kommunikationsnetze werden auch Verschlüsselung und Integritätssicherung gefordert. Der Fokus liegt dabei auf der Kommunikation zu den angeschlossenen Zählern.

Die „Technischen Richtlinien“ (TR-03109) [Bun15a] ergänzen das Schutzprofil. Sie erweitern die im Schutzprofil aufgeführten Mindestsicherheitsanforderungen um eine durch ein Messsystem zu erbringende Mindestfunktionalität. Insbesondere werden Testspezifikationen definiert und kryptografische Vorgaben gemacht.

Die Gesamtheit der Anforderungen dient als Maßstab für die Zertifizierung von Messsystemen. Von Unternehmen entwickelte Produkte können nach einer positiven Prüfung ein Zertifikat erhalten, welches bescheinigt, dass die Schutzziele erfüllt werden. Die konkrete technische Ausgestaltung obliegt aber weiterhin den Herstellern [Bun15b]. Beispielhaft sei die Arbeit von Peuker *et al.* [PE14] genannt, in der sie einen Prototypen mit offenem Quellcode vorstellen, welcher die Anforderungen des BSI-Schutzprofils beachtet.

Das Schutzprofil wird in der Literatur zum Teil stark kritisiert [KS14]. Insbesondere wird der schwache Schutz der Privatsphäre bemängelt. Es ist gemäß Schutzprofil möglich, Energiedaten, die nicht nur Abrechnungszwecken dienen, unter Pseudonymen zu verschicken. Eine solche Pseudonymisierung ist aber kein ausreichender Schutz [JJR11]. Dementsprechend sind weitergehende Schutzmaßnahmen erforderlich, um die Privatsphäre nicht zu gefähr-

den [KS14]. Somit ist das Schutzprofil zwar ein Schritt in die richtige Richtung, aber möglicherweise noch nicht ausreichend.

3.2.1.3. Gesetzentwurf zur Digitalisierung der Energiewende

Der Bundestag hat im Juni 2016 das „Gesetz über den Messstellenbetrieb und die Datenkommunikation in intelligenten Energienetzen (Messstellenbetriebsgesetz – MsbG)“ beschlossen [Deu16a; Deu16b]. Mit dem Gesetz ergeben sich weitreichende Änderungen im intelligenten Stromnetz Deutschlands. Das Gesetz nimmt unter anderem Bezug auf den Messstellenbetrieb sowie die damit verbundene Datenerhebung und -kommunikation. Insbesondere wird damit beschlossen, dass schrittweise bis zum Jahr 2032 alle Stromzähler in Deutschland gegen intelligente Stromzähler ausgetauscht werden. Die intelligenten Zähler müssen laut Messstellenbetriebsgesetz jedoch nicht in allen Haushalten über Kommunikationsmöglichkeiten verfügen, sondern können – unter anderem abhängig vom Jahresverbrauch – auch als einfache elektronische Zähler ausgeführt werden. Im Gesetz werden diese Geräte als „moderne Messeinrichtung“ bezeichnet. Die sogenannten „intelligenten Messsysteme“ müssen hingegen über eine Kommunikationsmöglichkeit verfügen.

Auch wenn die Entscheidung des Bunderats und eine praktische Umsetzung noch ausstehen, sollen an dieser Stelle knapp die für diese Arbeit wichtigsten Aspekte angesprochen werden. Paragraf 49 regelt die „Erhebung, Verarbeitung und Nutzung personenbezogener Daten“.

„(1) Personenbezogene Daten dürfen ausschließlich von den in Absatz 2 genannten Stellen erhoben, verarbeitet und genutzt werden (berechtigte Stellen). [...]

(2) Zum Umgang mit diesen Daten sind berechtigt:

[...] jede Stelle, die über eine schriftliche Einwilligung des Anschlussnutzers verfügt, die den Anforderungen des § 4a des Bundesdatenschutzgesetzes genügt.“

Paragraf 50 präzisiert „Zulässigkeit und Umfang der Erhebung, Verarbeitung und Nutzung von Daten“.

„Die Erhebung, Verarbeitung und Nutzung von Daten [...] darf nur mit Einwilligung des Anschlussnutzers erfolgen oder soweit dies erforderlich ist [...]“

Aus Nutzersicht ist zunächst zu begrüßen, dass die Erhebung und Nutzung personenbezogener Daten eingeschränkt werden soll. Darüber hinaus wird von Fall zu Fall entschieden werden müssen, ob eine Datennutzung „erforderlich ist“ beziehungsweise welche konkreten Auswirkungen diese Passage auf den Datenschutz haben wird. Ohne konkrete technologische Maßnahmen bleibt das Gesetz auch nur ein regulatorischer Schutz (vgl. Abschnitt 3.2.1.5). Erst die praktische Umsetzung des Gesetzes wird den tatsächlichen Nutzen für die betroffenen Personen zeigen.

3.2.1.4. Regulierungsanstrengungen auf europäischer Ebene

Die auf europäischer Ebene laufenden Standardisierungsanstrengungen befassen sich mit der Privatsphäre auch nur am Rande. Mit dem *Smart Grid Architecture Model* [Sma12b; Sma12a] wurde ein standardisiertes Modell eines intelligenten Stromnetzes im Auftrag der EU-Kommission geschaffen. Zwar ist eine der Hauptanforderungen an eine solche Architektur der Schutz personenbezogener Daten und der Privatsphäre durch rollenbasierte Zugriffsmodelle [Vei+13]. Konkrete Vorgaben werden allerdings nicht ausgearbeitet. Mit dem *Smart Grid Architecture Model* können vielmehr zu erarbeitende Anwendungsfälle auf das Modell abgebildet und daraus die zu erfassenden und kommunizierenden Daten abgeleitet werden. Zu diesem Zweck wurde die *Smart Grid Information Security Toolbox* entwickelt, welche die entsprechenden Schritte eines solchen Prozesses vorgibt [Sma12a]. Es handelt sich somit bei dem *Smart Grid Architecture Model* um ein Rahmenwerk, welches bei der Ausarbeitung konkreter Vorgaben hilft, diese aber nicht vorschreibt. Dementsprechend werden in der aktuellen Form keine expliziten Vorgaben über den Umgang mit Energiedaten gemacht. Allerdings existieren bereits Arbeiten, die sich mit der konkreten Ausgestaltung der unterschiedlichen Komponenten im Rahmenwerk beschäftigen [vgl. u.a. Vei+13].

3.2.1.5. Die eingeschränkte Wirksamkeit regulatorischer Ansätze

Regulatorische Ansätze zum Schutz von Daten sind in Bezug auf ihre Wirksamkeit vergleichbar mit Ampeln im Straßenverkehr (vgl. u.a. [PHB06]). Sie sollten für alle Teilnehmer streng bindend sein und erfüllen damit in der Regel ihren Zweck bis zu einer gewissen Grenze. Der größte Teil der (Verkehrs-)Teilnehmer hält sich an die Vorgaben und verzichtet auf Übertretungen. Es ist jedoch technisch nicht ausgeschlossen, dass einzelne sich über die Vorgaben hinwegsetzen und sie ignorieren. Die regulatorischen Ansätze können ebenso wie Verkehrsampeln nicht verhindern, dass sie missachtet werden. Lediglich im Nachgang ist eine Sanktionierung möglich. Ein Schaden ist dann aber zu meist schon angerichtet. Dementsprechend sind regulatorische Maßnahmen ein guter und wichtiger Schritt, aber für sich selbst genommen nicht ausreichend [SK12]. Im Folgenden werden daher technologische Ansätze vorgestellt, welche gegebenenfalls einen Verlust von Privatsphäre oder einen Missbrauch von Daten schon im Vorfeld verhindern können.

3.2.2. Technologische Ansätze

In diesem Abschnitt werden unterschiedliche technologische Ansätze zum Schutz von Energiedaten vorgestellt. Sowohl Jawurek *et al.* [JKD12] als auch Finster *et al.* [FB14a] haben jeweils Übersichten diesbezüglich erstellt. Dort sind weitere Informationen zu einigen der hier vorgestellten Methoden zu finden.

3.2.2.1. Kryptografische Ansätze

Homomorphe Verschlüsselung ist ein häufig verwendetes Ansatz, wenn Berechnungen auf Energiedaten durchgeführt werden sollen, ohne dass Einschränkungen der Privatsphäre hingenommen werden können [Erk+13]. Diese Art von kryptografischen Verfahren erlaubt es, einige mathematische Operationen auf verschlüsselten Daten auszuführen, wobei diese den Operationen auf unverschlüsselten Werten entsprechen. Auf Energiedaten angewendet bedeutet dies, dass eine vom Haushalt unabhängige Partei bestimmte Berechnungen auf den Energiedaten durchführen kann, ohne direkten Einblick in diese zu bekommen. Somit kann beispielsweise durch Aufsummieren eine Abrechnung

für den Stromverbrauch erstellt werden, ohne dass der Abrechnungsdienstleister Zugang zu Details des Energieverbrauchs bekommt, während gleichzeitig eine Verfälschung des Verbrauchs ausgeschlossen wird. Exemplarisch für diese Methode seien an dieser Stelle die Arbeiten von Rial *et al.* [RD11] und Garcia *et al.* [GJ11] genannt. Auch wenn die aktuell veröffentlichten Arbeiten zum Teil in der Praxis einsetzbar wären, scheint eine flächendeckende Umsetzung unwahrscheinlich. Die Methoden funktionieren meist nur mit starken Einschränkungen und unter bestimmten Annahmen. Insbesondere sind die mathematischen Operationen in der Regel auf Addition und Multiplikation beschränkt. Dies reduziert die Zahl möglicher Anwendungsszenarien, so dass die Verfahren in der Regel lediglich für Abrechnungszwecke vorgesehen sind. Darüber hinaus sind die meisten Verfahren überaus rechenintensiv, was insbesondere bei wenig leistungsfähigen intelligenten Zählern ein Problem sein kann.

Das Verfahren von Shi *et al.* [Shi+11] weist Parallelen zu den zuvor genannten Verfahren auf, nutzt jedoch nicht nur homomorphe Verschlüsselung. Das Prinzip ist jedoch ähnlich. Eine Reihe kryptografischer Methoden erlaubt es, dass einzelne Teilnehmer ihren Energieverbrauch verschlüsselt an einen Aggregatoren senden. Dieser muss nicht vertrauenswürdig sein, da er die einzelnen Nachrichten nicht entschlüsseln kann. Er ist jedoch in der Lage, aus den Mitteilungen die Summe aller Verbräuche zu errechnen. Die oben angesprochenen Einschränkungen gelten hier in ähnlicher Weise.

An dieser Stelle sei betont, dass die vorgestellten Arbeiten die kryptografische Verfahren nutzen, um die Energiedaten vor unerwünschten Einblicken auf Seiten des Empfängers zu schützen. Es handelt sich ausdrücklich nicht nur um eine Verschlüsselung der Daten bei ihrer Übertragung. Eine solche Transportverschlüsselung wird als übliche Praxis gemeinhin vorausgesetzt, auch wenn es in der Realität Ausnahmen geben mag (vgl. Abschnitt 3.1.2.2).

3.2.2.2. Verschleierung durch Rauschen

In den Arbeiten von Ács *et al.* [ÁC11] werden Energiedaten durch künstliches Rauschen so verschleiert, dass sich das Rauschen aller Teilnehmer in Summe wieder aufhebt. Dazu wird eine Menge von Haushalten mit intelligenten Stromzählern zu einer Gruppe zusammengefasst. Jeder Zähler übermittelt

periodisch die gemessenen Daten, wobei diese zuerst künstlich verrauscht und anschließend verschlüsselt werden. Der Empfänger der Daten kann die Summe aller empfangenen Werte einer Periode bilden und erhält damit den aktuellen Energieverbrauch der Gruppe, da das künstlich hinzugefügte Rauschen sich gegenseitig auslöscht. Ihm bleibt jedoch ein detaillierter Einblick in die Verbräuche der einzelnen Haushalte verwehrt, da diese jeweils – wie oben beschrieben – lediglich verrauschte Werte liefern. Die Ausprägung des Rauschens erlaubt es, zwischen Privatsphäre und Nützlichkeit der Daten abzuwagen. Auch wenn dieser Ansatz für einen speziellen Fall – der aggregierte Verbrauch einer Gruppe von Verbrauchern soll festgestellt werden – gut geeignet ist, lässt er sich nicht ohne Weiteres auf andere Anwendungsfälle übertragen. Das Rauschen wird von jedem einzelnen Teilnehmer individuell erzeugt. Damit gibt es keine Garantie, dass sich das Rauschen tatsächlich komplett aufhebt. Somit ist der Ansatz für eine geringe Zahl oder gar einzelne Haushalte gänzlich ungeeignet, da die Daten bei zu kleinem Rauschen nicht ausreichend verschleiert werden können oder bei stärkerem Rauschen die aggregierten Daten nicht die wahren Gegebenheiten widerspiegeln [BSU10].

Das Verfahren *SMART-ER* von Finster *et al.* [FB14b] arbeitet nach einem ähnlichen, wenn auch erweiterten Prinzip. Die erfassten Verbrauchsdaten jedes intelligenten Zählers werden durch diese in mehrere Fragmente aufgeteilt. Die Zähler senden jeweils ein Fragment an andere Knoten ihrer Gruppe und tauschen so Teile ihrer gemessenen Lastgänge aus, wobei die Gruppen kontinuierlich neu gebildet werden. Durch folgende Aggregationsschritte und das Weiterleiten der Daten an den zentralen Empfänger erhält dieser Informationen über den Gesamtverbrauch einer Gruppe. Einzelne Lastgänge sind durch den Austausch der Daten und die Summenbildung allerdings verschleiert. Auch bei diesem Verfahren ist eine Mindestzahl von Haushalten nötig, damit eine ausreichende Verschleierung gewährleistet werden kann. Einen vergleichbaren Ansatz stellen Kim *et al.* in [KNS11] vor.

3.2.2.3. Vertrauenswürdige Dritte

Unter anderem Bohli *et al.* [BSU10] schlagen eine unabhängige und vertrauenswürdige dritte Partei vor, welche die Daten der intelligenten Stromzähler empfängt, bevor diese zum eigentlichen Empfänger, dem Messstellenbetreiber

gelangen. Im Gegensatz zu diesem verfügt die unabhängige Partei über keine Informationen bezüglich der Identität der erfassten Haushalte. Der Messstellenbetreiber hingegen könnte, wenn er alle Daten der Zähler direkt erhielte, umfangreiche Profile anfertigen und diese wiederum Personen zuordnen. Um dies zu verhindern, schlagen Bohli *et al.* vor, die Daten aller Zähler einer Gruppe zuerst durch die dritte Partei aggregieren zu lassen, bevor sie weitergeleitet werden. Der Messstellenbetreiber kennt so stets den aktuellen Gesamtverbrauch in einem Netzabschnitt beziehungsweise in einer Gruppe von Zählern. Einzelne Lastgänge kann er jedoch nicht mehr identifizieren. Um jedoch am Ende einer Abrechnungsperiode eine korrekte Einzelabrechnung pro Haushalt erstellen zu können, erfragt er von der dritten Partei den aggregierten Verbrauch über die betreffende Periode. Mit diesem Konzept kann der Energieversorger einerseits das Netz permanent überwachen und andererseits den Verbrauch korrekt abrechnen, obwohl ihm ein direkter Einblick in die Verbräuche der Haushalte verwehrt bleibt. Vergleichbare, wenn auch im Detail abweichende, Ansätze werden unter anderem in [DK13], [EK10] und [KDK11] vorgestellt.

Eine solche Vorgehensweise verlagert das Problem des Datenschutzes meist nur vom Messstellenbetreiber zu einer dritten Partei, welcher die Beteiligten vertrauen müssen. Es ist technisch nicht auszuschließen, dass der Aggregatator der Energiedaten seine Stellung missbraucht und die Daten unerlaubterweise auswertet und für eigene Zwecke nutzt. Dies kann selbst dann nicht ausgeschlossen werden, wenn die Identität der einzelnen Datenquellen nicht bekannt ist, sondern mit Pseudonymen gearbeitet wird [JJR11].

3.2.2.4. Reduktion der zeitlichen Auflösung

Je höher die zeitliche Auflösung der erfassten Energiedaten ist, desto größer kann deren theoretischer Informationsgehalt sein. Elektrische Geräte, welche nur für sehr kurze Zeit betrieben werden, können hingegen bei einer zeitlich gering aufgelösten Messung nicht erfasst werden. Dementsprechend liegt es nahe, die zeitliche Auflösung zu reduzieren, um den Verlust von Privatsphäre zu minimieren. Engel [Eng13] schlägt vor, die Energiedaten direkt in verschiedenen Auflösungen zu erfassen und zu speichern. Jeder dieser unterschiedlich aufgelösten Datensätze wird mit einem individuellen Schlüssel verschlüsselt

und die Datensätze im Anschluss wieder gebündelt. Datenempfänger können nun jeweils nur die Daten einsehen, für welche Sie einen Schlüssel erhalten haben. Je nach Berechtigung beziehungsweise Rolle erhalten die Empfänger so Einblick in niedrig oder höher aufgelöste Daten. In [EE13] stellen Engel *et al.* eine Erweiterung ihres Konzepts vor. Sie zeigen, dass ihr Ansatz mit den üblichen homomorphen Verschlüsselungsverfahren kompatibel ist und diese ergänzen kann. Somit kann einerseits ein besserer Schutz gewährleistet werden und andererseits feiner gesteuert werden, welche Partei welche Daten lesen darf.

Eibl *et al.* untersuchen in [EE14; EE15] systematisch den Einfluss der zeitlichen Auflösung auf Verfahren zur Informationsextraktion aus Energiedaten (vgl. Abschnitt 4.2). Erwartungsgemäß kann festgestellt werden, dass eine verringerte zeitliche Auflösung eine Verschlechterung der Erkennungsrate zur Folge hat. Dies lässt sich hauptsächlich damit begründen, dass die meisten Algorithmen auf irgendeiner Form von Kantenerkennung beruhen. Durch die Reduktion der Auflösung treten die charakteristischen Kanten weniger deutlich oder gar nicht mehr auf. Dementsprechend können die Algorithmen diese auch nicht mehr identifizieren. Allerdings geht damit nicht zwangsläufig eine signifikante Verbesserung der Privatsphäre einher. Selbst wenn konkrete Vorgänge oder spezielle Gerätetypen nicht mehr erkannt werden können, ist es noch möglich, das allgemeine Verhalten wie beispielsweise An- oder Abwesenheit zu beobachten.

Schließlich quantifizieren Eibl *et al.* [EE15] den Einfluss der zeitlichen Auflösung auf die Erkennungsrate mithilfe des Referenzdatensatzes *REDD* (vgl. 3.1.1.1). Sie zeigen, dass die Erkennungsrate deutlich abnimmt, sobald der Abstand zweier Messpunkte mindestens halb so groß ist wie der durchschnittliche Zeitraum, in welchem ein Gerät am Stück in Betrieb ist. Für einen normalen Haushalt bedeutet dies, dass im Stundentakt erfasste Energiedaten kaum noch Rückschlüsse auf die Verwendung konkreter Geräte zulassen. Andererseits können durchaus auch bei solchen Auflösungen noch Abwesenheiten oder besondere Gewohnheiten erkannt werden. Zusammenfassend ist die vergleichsweise einfach durchführbare Auflösungsreduktion aber eine effektive Vorgehensweise, um den Verlust der Privatsphäre zumindest einzuschränken.

3.2.2.5. Verwendung von elektrischen Speichern zur Verschleierung des Energieverbrauchs

Varoyadan *et al.* [VK11] sowie Kalogridis *et al.* [Kal+10; KFB11] schlagen in ihren Arbeiten die Verwendung von elektrischen Speichern zur Verschleierung des Energieverbrauchs in Wohngebäuden vor. Der von den Geräten benötigte Strom wird zwar prinzipiell durch das Stromnetz zur Verfügung gestellt, Energieverbrauch und Gerätenutzung müssen aber nicht zeitlich korrelieren. Da die Geräte nicht direkt, sondern über ein Gerät zur Steuerung des Stromflusses und die Batterie mit elektrischer Energie versorgt werden, lässt sich das Lastprofil aus Netzsicht verändern und verschieben.

Eine Steuerungseinheit regelt in Kombination mit der Batterie den Stromfluss im Haushalt und damit den vermeintlichen Stromverbrauch aus Stromnetzsicht. Zum einen kann der gesamte Strom oder auch nur ein Teil durch die Batterie zur Verfügung gestellt werden. Die dafür nötige Energie muss dementsprechend vor der Nutzung beziehungsweise danach aus dem Netz bezogen werden, um die Batterie wieder aufzuladen. Zum Zeitpunkt der Gerätenutzung wird dadurch aus Sicht des Netzes weniger oder kein Strom im Haushalt verbraucht. Die Verwendung eines Gerätes kann somit verschleiert oder scheinbar zeitlich verschoben werden.

Alternativ kann die Batterie als zusätzlicher Verbraucher dienen und während des Betriebs eines Gerätes zusätzlich Strom aus dem Netz beziehen. Auch diese Betriebsweise lässt eine Verschleierung des Lastprofils zu, da der gesamte Stromverbrauch auf einen konstanten Wert eingestellt werden kann, der oberhalb des eigentlichen Lastprofils liegt. Aus Netzsicht ist dann zwar zu erkennen, dass Strom verbraucht wird, nicht aber, welches Gerät diesen Verbrauch verursacht.

Darüber hinaus ist es aus Sicht der Autoren durchaus denkbar, ein definiertes Lastprofil mit der Batterie abzufahren, um die Verwendung eines bestimmten Gerätes vorzutäuschen. Bei konsequenterem Einsatz eines solchen Systems kann aus Netzsicht kaum noch beurteilt werden, welche Qualität die Aussagen der Informationsextraktionsverfahren haben. Selbst wenn die Verfahren noch Geräte erkennen, kann nicht ausgeschlossen werden, dass es sich um emulierte Lastprofile handelt.

Auch wenn die Verwendung elektrischer Speicher in Haushalten zweifelsfrei viele Vorteile für das gesamte intelligente Stromnetz mit sich bringt, ist der Preis für solche Systeme nach wie vor ein großer Nachteil. Dieses Problem gestehen auch Kalogridis *et al.* ein. Daher stellen sie in [KFB11] und [KD12] verbesserte Verfahren vor, welche mit kleineren, weniger leistungsfähigen und damit auch günstigeren Batterien vergleichbare Ergebnisse vorzeigen können. Darüber hinaus sind die verwendeten Systeme in der Lage, zur Stabilitätsverbesserung des Netzes beizutragen. Allen Vorteilen zum Trotz dürften die Kosten nach wie vor ein Hinderungsgrund für einen großflächigen Einsatz dieser Technologie sein. Davon abgesehen lässt sich mit einem solchen System aber eine deutliche Verbesserung der Privatsphäre erreichen.

3.2.2.6. Platform for Privacy Preferences (P3P)

Das *World Wide Web Consortium* (W3C) ist das Gremium für Standardisierung im World Wide Web. Unter anderem mit HTML, XHTML und XML hat das W3C viele allgegenwärtige Technologien der digitalen Welt standardisiert. Im Jahr 2002 veröffentlichte das Konsortium die *Platform for Privacy Preferences* (P3P) [Wor02], welche als ein Standard zum Austausch von Datenschutzinformationen etabliert werden sollte. Der Standard ist in [Wor02] detailliert beschrieben und wird im Folgenden kurz erläutert.

Hauptsächliches Ziel der Plattform ist es, Nutzern von Webseiten auf standardisierte Weise zu vermitteln, welche personenbezogenen Daten beim Besuch einer Webseite anfallen und wie diese verwendet werden. Diese Informationen sollen sowohl in maschinen- als auch in menschenlesbarer Form zur Verfügung stehen. Dementsprechend können Entscheidungsprozesse auf Seiten der Nutzer automatisiert umgesetzt werden, so dass das Studieren der Datenschutzbedingungen für die Besucher einer Webseite entfallen kann.

Der Standard sieht vor, dass serverseitig eine XML-Datei zur Verfügung gestellt wird, welche Informationen bezüglich der potenziell betroffenen personenbezogenen Daten enthält. Dies umfasst neben der Art der Daten, auch den Verwendungszweck, die Empfänger und Weiteres. Clientseitig entscheidet der Nutzer oder ein vom Nutzer verwendetes System, ob diese *P3P Policy* akzeptiert wird. Private Daten bleiben unter Kontrolle der Nutzer und wer-

den nur an zertifizierte Externe weitergegeben, falls ein tatsächlicher Bedarf nachgewiesen werden kann.

Obwohl der Standard vor über zehn Jahren veröffentlicht wurde, hat er bis heute keine große Verbreitung erfahren. Nur wenige Webseiten bieten die Nutzung von *P3P* an. Auch die meisten der weitverbreiteten Browser bieten keine Unterstützung für den Standard. Die Weiterentwicklung wurde seitens W3C ebenfalls eingestellt. Dementsprechend ist zu vermuten, dass sich *P3P* auch in Zukunft nicht durchsetzen kann.

Eine Anwendung für Energiedaten war durch das W3C nicht explizit vorgesehen. Das Prinzip des Standards ließe sich aber auch auf diese übertragen. Würde die Grundidee von *P3P* für Energiedaten verwendet, wäre zumindest eine Aufklärung über die Datenschutzfolgen, wenn auch kein Schutz der Privatsphäre möglich. Nutzer würden informiert über die Art der zu übertragenden Energiedaten, den geplanten Verwendungszweck und den Empfänger. Es wäre jedoch kein garantierter Schutz, da Vertrauen in die Ehrlichkeit des Dienstebetreibers und die Korrektheit der Angaben erforderlich sind. Bei absichtlich oder versehentlich falsch erstellten *P3P Policies* oder davon abweichendem Verhalten der Dienstleister könnten Energiedaten nicht geschützt werden. Es könnte technisch nicht verhindert werden, dass die Energiedaten für andere als die angegebenen Zwecke verwendet werden. Sehr wahrscheinlich würde ein abweichendes Verhalten gar nicht erkannt werden.

3.2.2.7. Usage Control

Usage Control ist ein Verfahren aus dem Bereich der digitalen Rechteverwaltung (engl. *Digital Rights Management* (DRM)). Es handelt sich um mehr als eine reine Zugriffskontrolle, da nicht nur festgelegt wird, wer auf die Daten zugreifen darf, sondern auch in welcher Weise sie genutzt werden dürfen. Pretschner *et al.* stellen in ihren Arbeiten [PHB06; Pre+08] ein Konzept für ein solches System vor.

Zu schützende Daten werden von einem *Data Consumer* bei einem *Data Provider* angefragt. Der Anfrage wird nur stattgegeben, wenn die zugehörige *Policy* es erlaubt und von allen beteiligten Parteien akzeptiert wird. In dieser Richtlinie sind Pflichten aufgeführt, welche der Anfragende beziehungsweise der Datenempfänger zu erfüllen hat. Beispiele für solche Pflichten sind die

Lösung nach Ablauf einer Zeitspanne oder eine Meldung an den Datenurheber, sobald die Daten verwendet werden. Eine Vielzahl weiterer Pflichten kann im Rahmen des Regelwerks festgelegt werden.

Das die Pflichten betreffende Verhalten kann zum Teil kontrolliert und zum Teil nur beobachtet werden. Im Unterschied zu nur beobachteten Pflichten kann der Datenurheber bei kontrollierbaren Pflichten sicher sein, dass diese durchgesetzt werden können. Beobachtbare Pflichten können hingegen nicht erzwungen werden. Es ist lediglich möglich, einen Verstoß zu erkennen und gegebenenfalls zu ahnden.

Dieses abstrakte Rahmenwerk muss zur Umsetzung in ein konkretes System überführt werden. Je nach Implementierung sind die potenziell zu formulierenden Pflichten nur beobachtbar oder auch kontrollierbar beziehungsweise durchsetzbar. Die Sicherheit und Nutzbarkeit eines solchen Systems ist damit nicht nur vom Konzept, sondern auch in hohem Maße von der konkreten Implementierung abhängig. Alle zu schützenden Daten müssen innerhalb des Systems gehandhabt werden, sodass dieses stets die Kontrolle behält. Sobald die Daten das System verlassen, entzieht sich der Nutzer der Daten der Kontrolle des Systems und untergräbt damit den Schutz der Daten. Im Umkehrschluss bedeutet dies, dass alle Operationen, die auf den Daten ausgeführt werden sollen, im System verfügbar sein müssen. Ansonsten wäre die Nutzbarkeit der Daten stark eingeschränkt.

Prinzipiell wäre *Usage Control* auch für Energiedaten einsetzbar. In einem mit entsprechenden Pflichtendefinitionen ausgestatteten System könnten Energiedaten vorgehalten und verarbeitet werden. Über die Regelwerke hinaus müssten aber auch alle potenziell gewünschten Datenoperationen unterstützt werden. Dies ist insbesondere im Bereich von Energiedaten eine große Herausforderung, da es zahlreiche Verfahren zur Analyse von Energiedaten gibt, welche zum Teil unterschiedlichste Ansätze verfolgen. Darauf hinaus müsste das System allen Beteiligten – Datenurhebern und Datenanfragenden gleichermaßen – zur Verfügung gestellt werden. Die Sicherheit der Daten kann schließlich nur dann gewährleistet werden, wenn die Daten das System nicht verlassen. Somit müssten alle Beteiligten das System nutzen oder auf einen Datenzugriff verzichten.

3.2.3. Zusammenfassung

Der Überblick zu den verwandten Arbeiten zeigt, dass sehr unterschiedliche Vorstellungen davon existieren, wie Energiedaten verwaltet und geschützt werden sollen. Insbesondere im wirtschaftlichen Bereich besteht oft ein Interessenkonflikt zwischen den Betroffenen der Energiedatenerfassung und der Industrie. Viele der Ansätze vernachlässigen die von der Datenerfassung betroffenen Personen beziehungsweise den Schutz deren Privatsphäre. Ansätze, die dem Schutz von Energiedaten dienen sollen, vernachlässigen im Gegenzug jedoch häufig die technischen Aspekte des Umgangs mit den Daten. Naturgemäß konzentrieren sich viele Arbeiten auf einzelne Teilespekte und blenden das restliche System aus.

Im folgenden Kapitel wird daher aus unterschiedlichen Blickwinkeln eine umfangreiche Anforderungsanalyse durchgeführt. Diese hat das Ziel, einen Anforderungskatalog zu formulieren, welchen ein Energiedatenmanagementsystem erfüllen sollte – aus technischer Sicht wie auch aus Nutzersicht.

Anforderungsanalyse

In diesem Kapitel wird eine umfangreiche Anforderungsanalyse durchgeführt. Ziel der Untersuchung ist die Ausarbeitung sowohl von technischen als auch von nutzergetriebenen Anforderungen.

Zunächst erfolgt eine Untersuchung des Energiedatenlebenszyklus. Dieser deckt von der Erfassung über die Übertragung, die Speicherung, die Analyse, die Verteilung bis hin zur Löschung alle Phasen des Lebenszyklus ab. Auf diese Weise können die technischen Anforderungen systematisch erarbeitet werden. Anschließend werden die Interessen der Nutzer beleuchtet. Charakteristisch für Energiedaten ist einerseits deren Personenbezug und die damit verbundene Gefahr für die Privatsphäre der betroffenen Personen. Andererseits erlauben Energiedaten auch im gewerblichen Umfeld Rückschlüsse auf den Betriebszustand oder die Leistungsfähigkeit von Maschinen oder Produktionsstraßen. Dementsprechend sind der Datenschutz und die Wahrung der Privatsphäre im Fokus.

Den Abschluss des Kapitels bildet eine Zusammenfassung der erarbeiteten Anforderungen, die als Ausgangsbasis für die Konzeption eines nutzerorientierten Energiedatenmanagementsystems dienen kann.

4.1. Technische Datenlebenszyklusanalyse

Energiedatenmanagementsysteme müssen eine Reihe von technischen Anforderungen erfüllen, um möglichst nutzbringend eingesetzt werden zu können. Insbesondere weisen Energiedaten einige Besonderheiten auf, welche sie von vielen anderen Arten von Daten unterscheiden. Dementsprechend ist eine systematische Untersuchung der spezifischen Anforderungen nötig. Die Da-

tenlebenszyklusanalyse stellt eine Möglichkeit dar, Energiedaten systematisch zu untersuchen und aus den Ergebnissen Anforderungen abzuleiten.

4.1.1. Der Datenlebenszyklus als Werkzeug

Das *Data Life Cycle Lab Energy* im Helmholtz-Projekt *Large Scale Data Management and Analysis* (LSDMA) [Wez+12; Jun+14; Jun+15; Jun+16] bildete den Rahmen, in welchem diese Arbeit zum größten Teil entstanden ist. In LSDMA existieren weitere *Data Life Cycle Labs*, welche zentrale Einheiten zur Untersuchung von Datenlebenszyklen in ihrem jeweiligen Forschungsbereich darstellen. Die dort durchgeföhrten Untersuchungen verfolgen zwei Hauptziele. Einerseits sollen domänenspezifische Lösungen und Methoden entwickelt werden, welche im Bereich des jeweiligen *Data Life Cycle Labs* Anwendung finden können. Andererseits sollen aus einem *Data Life Cycle Lab* entstammende Lösungen im Idealfall auch in anderen Bereichen Anwendung finden können. Im *Data Life Cycle Lab Energy* werden vor allem Energiedaten, deren Lebenszyklen, sowie Projekte mit Bezug zu Energiedaten untersucht. Die im Rahmen der Tätigkeit im *Data Life Cycle Lab Energy* gewonnenen Erkenntnisse fließen in besonderem Maße in die folgenden Abschnitte ein.

Eines der zentralen Werkzeuge in den *Data Life Cycle Labs* ist das Datenlebenszyklusmodell. Es ist als Rahmenwerk für einen prototypischen Datenlebenszyklus zu verstehen, welcher aus sechs Phasen besteht. Die Phasen sind zeitlich aufeinander folgend angeordnet, aber nicht zwangsläufig strikt getrennt. Vielmehr sind sie Anhaltspunkte, um den Datenlebenszyklus systematisch zerlegen und untersuchen zu können. Abbildung 4.1 stellt das Modell eines solchen Datenlebenszyklus dar. Im Folgenden werden die einzelnen Phasen kurz erläutert.

Erfassung Der Zyklus beginnt mit der Erfassung der Daten. Untersucht werden in diesem ersten Abschnitt unter anderem deren Ursprung, die Art der Erfassung, die Datenqualität, die Datenmenge und die verwendeten Datenformate.

Übertragung Nach der Erfassung der Daten folgt mit deren Übertragung die zweite Phase. Im Fokus der Untersuchung sind hier Übertragungsprotokolle und -datenformate, ebenso wie Frequenz und Qualität der Übertragung.

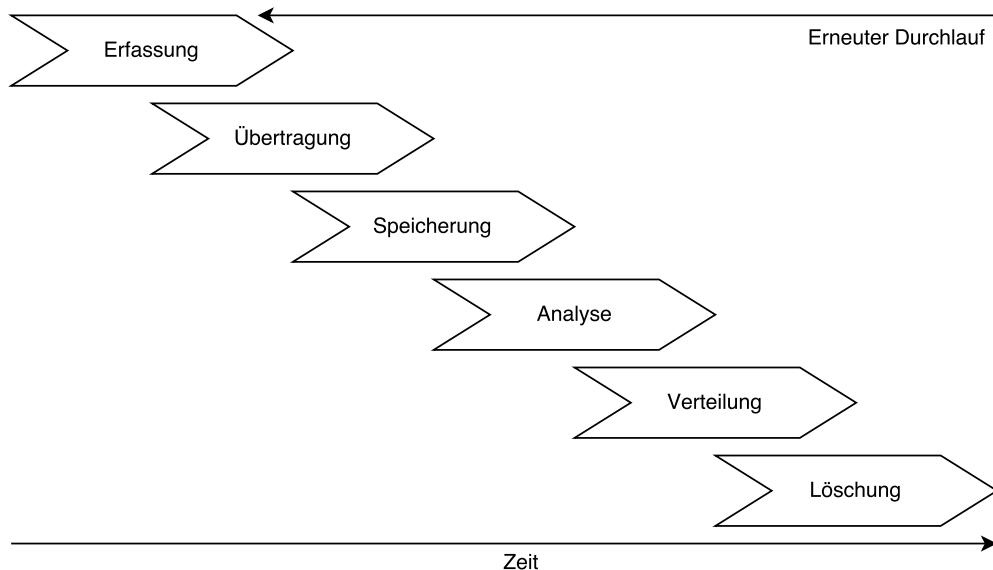


Abbildung 4.1.: Dieses Datenlebenszyklusmodell ist angelehnt an die im LSDMA Data Life Cycle Lab Energy [vgl. Wez+12; Jun+14; Jun+15; Jun+16] verwendeten Modelle. Der Datenlebenszyklus besteht aus sechs, zeitlich aufeinander folgenden Phasen, welche jedoch nicht strikt getrennt werden müssen. Im Rahmen einer Analyse können die einzelnen Phasen auch mehrfach durchlaufen werden.

Speicherung Die übertragenen Daten werden in der nächsten, der dritten, Phase gespeichert. In dieser Phase werden neben der Art der gespeicherten Daten auch die verwendeten Datenformate, Speichersysteme sowie Anforderungen an diese betrachtet.

Analyse In der vierten Phase, der Analyse, werden die Daten ihrem Hauptverwendungszweck, der Verarbeitung, zugeführt. Ziel der Untersuchung dieses Abschnittes ist es, den Zweck der Analysen, die geplante Vorgehensweise und die verwendeten Methoden zu bestimmen.

Verteilung Im vorletzten Abschnitt wird betrachtet, inwiefern die ursprünglich übertragenen, aber auch die bearbeiteten Daten, weiterverteilt werden. Diese Phase ist der zweiten Phase ähnlich, zieht aber insbesondere externe Parteien mit ein, die Daten aus dem Datenlebenszyklus erhalten. Daher müssen hier zusätzlich auch die Datenempfänger betrachtet werden.

Lösung Eine Iteration des Lebenszyklus endet mit der Untersuchung der Lösung der Daten. Im Fokus steht hier, ob und welche Daten tatsächlich gelöscht werden.

Zweifelsohne weisen die meisten realen Datenlebenszyklen eine komplexere Struktur auf als die in Abbildung 4.1 gezeigte. Die strikte Trennung wie im Modell wird in der Realität nur selten zu finden sein. Vielmehr ist es durchaus möglich, dass sich Phasen überlappen oder mehrfach wiederholen. So ist es beispielsweise durchaus üblich, dass im Rahmen der Analysephase Daten entstehen, welche wiederum im System erfasst werden, wodurch ein neuer Zyklus entsteht. Dennoch bietet das Modell einen wertvollen Ansatz, um die Untersuchung von Daten und deren Lebenszyklen systematisch und nach Abschnitten getrennt durchführen zu können und zu vereinheitlichen. Auch auf Energiedaten und deren Management lässt sich die Datenlebenszyklusanalyse anwenden. Die Definition der VDI-Richtlinie 4602 zu „Energiedatenmanagement“ (vgl. Abschnitt 2.3.1) enthält insbesondere auch die Begriffe Bereitstellung, Übermittlung, Verarbeitung und Speicherung von Daten. An dieser Stelle sind die Parallelen zu einigen der oben genannten Phasen offenkundig.

Im Folgenden werden die einzelnen Phasen des Datenlebenszyklus von Energiedaten beleuchtet. Aus den Ergebnissen der Datenlebenszyklusanalyse werden wiederum technische Anforderungen abgeleitet, welche ein Energiedatenmanagementsystem erfüllen muss. Die Anforderungen werden jeweils durchnummieriert und kurz am Ende einer jeden Phase zusammengefasst.

4.1.2. Phase I: Erfassung

Die Erfassung der Daten wird in der ersten Phase des Datenlebenszyklus untersucht. Typische Fragen wären unter anderem die folgenden:

- Welche Daten werden erfasst?
- Wie werden die Daten erfasst?
- Welche Qualität weisen die Daten auf?
- Was ist das bevorzugte Format?

4.1.2.1. Quellen für Energiedaten

Um die oben genannten Fragen zu beantworten, werden im Folgenden verschiedene Arten von Energiedatenquellen untersucht.

Intelligente Stromzähler Intelligente Stromzähler (engl. *smart meter*) erfassen in der Regel den Stromverbrauch eines Haushaltes oder eines Unternehmens am Netzanschlusspunkt. Auf dem Markt sind diverse Geräte von verschiedenen Anbietern in jeweils unterschiedlichen Ausstattungen verfügbar. Die intelligenten Zähler des Unternehmens *EasyMeter GmbH* dienen im Folgenden als Beispiel für solche intelligenten Stromzähler [Eas16]. Das Unternehmen bietet sie in mehreren, unterschiedlichen Varianten an.

In der kleinsten Ausstattungslinie besitzt der Zähler lediglich eine elektronische Anzeige und eine Impulsschnittstelle. Diese Schnittstelle gibt über eine Leuchtdiode 10 000 Impulse pro Kilowattstunde ab¹, so dass ein zusätzlich anzuschließendes Gerät die Impulse erfassen und auswerten muss. Der Zähler selbst zeigt zwar den aktuellen Zählerstand sowie die aktuelle Leistung auf einer Anzeige an, eine weitere Auswertung und gegebenenfalls Kommunikation müsste aber ein Zusatzgerät übernehmen. Darüber hinaus hat die Impulsschnittstelle weitere Nachteile. Die durchschnittliche Leistung lässt sich nur über den zeitlichen Abstand der Impulse interpolieren, da diese lediglich anzeigen, dass eine gewisse Energiemenge verbraucht wurde. Auch der Zählerstand wird nicht direkt übermittelt, sondern kann lediglich durch einen Startwert und fortlaufende Addition der verbrauchten Energie bestimmt werden. Dies bedeutet aber, dass bei Unterbrechungen der Aufzeichnung zwangsläufig Fehler bei der Bestimmung des Zählerstands entstehen, die nur durch manuelles Angleichen wieder behoben werden können.

Modelle mit höherwertiger Ausstattung verfügen dagegen nicht nur über eine Impulsschnittstelle, sondern auch über eine uni- oder bidirektionale Kommunikationsmöglichkeit mit externen Geräten über die sogenannte *Smart Message Language (SML)* [Bun13a]. Der Zähler sendet im Zweisekundentakt eine sogenannte *SML-Nachricht* mittels einer Leuchtdiode. Diese hexadezimal kodierten Nachrichten können von entsprechenden Geräten empfangen und

¹Die Anzahl der Impulse ist modellabhängig. Üblich sind unter anderem auch 800, 1000 oder 2000 Impulse pro Kilowattstunde.

ausgewertet werden. Im Gegensatz zur Impulsschnittstelle werden mit der *Smart Message Language* deutlich mehr Informationen zur Verfügung gestellt. Die Nachrichten enthalten neben der aktuellen Leistung unter anderem auch den Zählerstand sowie die eindeutige Kennung des Zählers. Auch wenn es technisch problemlos möglich wäre, ist es typischerweise nicht vorgesehen, dass Endkunden die Daten direkt vom intelligenten Stromzähler abgreifen. Stattdessen erhält für gewöhnlich ein Messstellenbetreiber die erfassten Verbrauchsdaten und kann diese wiederum dem Endkunden zur Verfügung stellen. Dies geschieht für gewöhnlich nicht über die *Smart Message Language*, sondern über interaktive Webseiten beziehungsweise kommaseparierte Dateien (vgl. 3.1.2.2).

Nicht nur intelligente Stromzähler, wie der zuvor beschriebene, können Energiedaten am Netzanschlusspunkt liefern. Auch traditionelle, elektromechanische Ferraris-Zähler können mit zusätzlichen Geräten ausgestattet werden, welche die Markierung auf der sich drehenden Zählerscheibe erfassen. Jede Umdrehung entspricht dabei einer festgelegten Energiemenge – beispielsweise 1/75 kWh. Das angeschlossene Zusatzmodul liefert dann, ganz ähnlich wie die Impulsschnittstelle eines intelligenten Zählers, Informationen über den Energieverbrauch. Auch diese Art der Erfassung weist die Nachteile der Impulszähler auf. Erschwerend kommt hinzu, dass die potenziell ungenaue Erfassung der Zählerscheibe eine weitere Fehlerquelle darstellt.

Die Produkte bereits eines Herstellers zeigen, dass zahlreiche Möglichkeiten existieren, wie die Daten intelligenter Stromzähler erfasst und verarbeitet werden können. Dementsprechend groß ist auch die potenzielle Zahl unterschiedlicher Darstellungen. Die Spanne reicht von einzelnen Ereignissen wie Energieverbrauchsimpulse über komplettete *SML-Nachrichten* bis hin zu ausgewerteten Daten des Messstellenbetreibers. Unabhängig vom konkreten Format ist aber allen Daten gemeinsam, dass sie in Form einer Zeitreihe interpretierbar sind: Jedem Messpunkt, sei es ein Impuls oder eine (aufbereitete) *SML-Nachricht*, kann ein eindeutiger Zeitstempel zugeordnet werden.

Auch wenn aus Sicht der Haushalte Metadaten keine große Rolle zu spielen scheinen, darf ihre Wichtigkeit nicht unterschätzt werden. Innerhalb eines Haushaltes gibt es in der Regel nur einen intelligenten Stromzähler, daher müssen erfasste Daten in dem Fall zwangsläufig von diesem kommen. Wird aber ein Mehrparteienhaus mit zentraler Energiedatenhaltung betrachtet,

gestaltet sich der Fall nicht mehr so einfach. Die einzelnen Datenströme müssen jeweils wieder einem Zähler zugeordnet werden können. Selbst wenn in den Daten eine entsprechende Identifikationsmöglichkeit gegeben ist (beispielsweise bei *SML-Nachrichten*), muss das Identifikationsmerkmal noch einer Wohnung zugeordnet werden können. Dementsprechend sollten die Daten mit aussagekräftigen Metadaten versehen werden.

Intelligente Zwischenstecker Während intelligente Stromzähler den aggregierten Stromverbrauch eines Haushalts oder Unternehmens messen, eignen sich intelligente Zwischenstecker (engl. *smart plugs*) dazu, den Stromverbrauch einzelner Geräte direkt zu erfassen. Allerdings ist zu beachten, über welche Funktionen solche Zwischenstecker tatsächlich verfügen. Am Markt werden zum Teil auch Zwischensteckdosen als „*smart plugs*“ beworben, obwohl sie lediglich ein- beziehungsweise ausgeschaltet werden können, den Stromverbrauch aber nicht erfassen. Die hier besprochenen intelligenten Zwischensteckdosen sind hingegen dadurch charakterisiert, dass sie auch den Stromverbrauch der angeschlossenen Geräte messen und übertragen können.

In Abschnitt 3.1.2.3 wurden bereits die vom niederländischen Unternehmen *Plugwise* vertriebenen Messgeräte vorgestellt. Sie stellen ein typisches Beispiel für intelligente Zwischensteckdosen dar. Bei Verwendung des herstellereigenen Programms *Source* kann ein Endkunde die Daten im Rahmen der gegebenen Möglichkeiten auswerten. Eine weitere Verwendung in externen Programmen ist jedoch nicht unmittelbar vorgesehen. Daher werden von unabhängigen Entwicklern einige quelloffene Bibliotheken angeboten, welche es erlauben, die Messgeräte ohne Verwendung von *Plugwise Source* direkt anzusprechen und auszulesen. Je nach Bibliothek unterscheiden sich die verwendeten Formate, nicht jedoch die erfassten Daten. Ausgegeben werden unter anderem ein Zeitstempel, eine eindeutige Identifikation der jeweiligen Zwischensteckdose sowie die aktuell gemessene Leistung. An dieser Stelle wird deutlich, dass nicht nur die erfassten Daten selbst, sondern auch die zugehörigen Metadaten von großer Bedeutung sind. Während die Identifikationskennung eine eindeutige Zuordnung der Daten zu einer Steckdose ermöglicht, geht aus ihr nicht direkt hervor, welches Gerät gemessen wurde. Diese Information muss somit zusätzlich zu den eigentlichen Messdaten

vorgehalten werden. Ohne Metadaten wäre eine Zuordnung möglicherweise im Nachhinein nicht mehr zweifelsfrei möglich, womit die Daten an Wert verlieren.

Bei den Daten der intelligenten Zwischensteckdosen handelt es sich um real gemessene Verbräuche der angeschlossenen Geräte. Prinzipiell vergleichbar sind allerdings auch Ergebnisse aus Verfahren zur Informationsextraktion aus Energiedaten (vgl. Abschnitt 4.2.2). Im Erfolgsfall liefern solche Verfahren zerlegte Lastgänge, welche zwar zentral und aggregiert erfasst, dann aber in einzelne Geräte aufgespalten wurden. Dementsprechend können diese algorithmisch ermittelten Daten analog zu real gemessenen Daten behandelt werden. Auch hier müssen Metadaten gepflegt werden.

Bei näherer Betrachtung sind Parallelen zu den Daten aus intelligenten Stromzählern erkennbar. Doch obwohl es sich bei allen Daten um Energiedaten handelt, unterscheiden sie sich im Detail. Gemeinsam ist ihnen aber, dass sie in Zeitreihenform darstellbar sind.

Weitere Quellen für Energiedaten In den vorherigen Abschnitten wurden mit intelligenten Stromzählern und intelligenten Zwischensteckern Messgeräte für elektrische Energie vorgestellt. Neben den angeführten Beispielen gibt es zahlreiche weitere Produkte, welche vergleichbare Daten zum Stromverbrauch liefern. Doch elektrische Energie ist nicht die einzige Energieform, für welche intelligente Zähler existieren. Auch für Gas und Fernwärme gibt es entsprechende (intelligente) Zähler. Ebenso wie ihre elektrischen Pendants verfügen auch diese über die Möglichkeit, den Gesamtverbrauch an Gas beziehungsweise Wärme zu erfassen und zu kommunizieren.

Bei dem von Cohn *et al.* [Coh+10] entwickelten Gerät namens *GasSense* handelt es sich um eine besondere Art eines intelligenten Gaszählers. Der Sensor ersetzt allerdings nicht den klassischen Gaszähler, sondern wird zusätzlich am bereits vorhandenen Druckminderer angeschlossen. An diesem erfasst er mit einem Mikrofon die beim Gasverbrauch entstehenden Geräusche, welche wiederum ausgewertet werden können. Anhand typischer Vibrationen wie sie beispielsweise beim Öffnen oder Schließen von Ventilen sowie bei Zündvorgängen entstehen, können unterschiedliche Verbraucher identifiziert werden. Dies erlaubt eine feingranulare Analyse des Gasverbrauchs in einem Haushalt.

Die Abschätzung des Verbrauchs ist konstruktionsbedingt nicht genau genug, um den klassischen Zähler zu ersetzen. Dennoch weist das Verfahren eine mit 95 % sehr hohe Erkennungsrate auf, weshalb es sich gut als Ergänzung zu einem (intelligenten) Gaszähler eignet. Auch diese Gasverbrauchs- beziehungsweise Nutzungsdaten sind Energiedaten. Obwohl sie anders geartet sind als die zuvor besprochenen elektrischen Energiedaten, sollten auch solche Datentypen in einem Energiedatenmanagementsystem verwaltet werden können.

Der *HydroSense*-Sensor von Froehlich *et al.* [Fro+09] arbeitet nach einem ähnlichen Prinzip, wird aber zur Erfassung des Wasserverbrauchs eingesetzt. Zu diesem Zweck verfügt *HydroSense* über einen Drucksensor, welcher an einen beliebigen, geöffneten Wasserhahn angeschlossen wird. Sobald im Haushalt Wasser verbraucht wird – sei es an einem Wasserhahn oder durch ein Gerät – erkennt der Sensor eine Druckschwankung im hausinternen Leitungsnetwork. Beim Schließen eines Ventils entsteht erneut eine Druckschwankung. Diese Ereignisse können genutzt werden, um Verbraucher und deren Nutzungsduer zu identifizieren. Aus diesen Erkenntnissen kann wiederum die entnommene Wassermenge geschätzt werden. Die von Froehlich *et al.* erzielten Erkennungsraten liegen bei über 97 %. Auch wenn Wasser ein Energieträger sein kann, handelt es sich bei Leitungswasser nicht um Energie im engeren Sinne. Dennoch sind Wasserverbrauchsdaten mit Energiedaten eng verwandt. Wasser wird in jedem Haushalt und den allermeisten Unternehmen täglich benötigt. Darüber hinaus lässt der Wasserverbrauch detaillierte Rückschlüsse auf die betroffenen Personen zu: Duschverhalten, Toilettennutzung sowie die Verwendung von Geschirrspülern oder Waschmaschinen sind einfach zu erkennen [Fro+09]. Dementsprechend sollten auch diese personenbezogenen Daten ähnlich wie Energiedaten behandelt werden und gegebenenfalls in einem geeigneten (Energie-)Datenmanagementsystem vorgehalten werden.

Intelligente Stromzähler und Zwischenstecker erfassen in der Regel elektrische Energie oder auch die Leistung. Insofern stellt der *Electrical Data Recorder* von Maass *et al.* [Maa+12] einen Spezialfall dar. Dieser Sensor misst nicht den Strom, sondern den Verlauf der Spannung (und damit auch die Frequenz) mit einer zeitlichen Auflösung von bis zu 25 kHz. Während die Frequenz Aussagen über die globale Netzsituation zulässt, erlaubt die Spannung Rückschlüsse auf den regionalen oder gar lokalen Zustand. Die hohe Frequenz

gestattet eine detaillierte Analyse des Spannungsverlaufs und die Erkennung von Oberwellen. Allerdings bringt sie den Nachteil großer Datenmengen mit sich. In Abhängigkeit der tatsächlich aufgezeichneten Auflösung fallen pro *Electrical Data Recorder* bis zu 17 GB Daten am Tag an. Solche Datenmengen lassen sich vermutlich auch in naher Zukunft nicht in einem generischen Energiedatenmanagementsystem verwalten. Wahrscheinlich werden auch weiterhin darauf spezialisierte Systeme nötig sein, welche für diese hohen Datenraten ausgelegt sind. Es darf jedoch davon ausgegangen werden, dass Geräte wie der *Electrical Data Recorder* keine flächendeckende Verbreitung erfahren werden. Einerseits sind sie teurer als normale intelligente Zähler und andererseits verfolgen sie mit der Erfassung der Spannungsqualität einen anderen Anwendungszweck.

Auch das *ElectriSense* genannte Messgerät von Gupta *et al.* [GRP10] stellt einen Sonderfall dar. Es ist spezialisiert auf die Erkennung elektromagnetischer Interferenz, welche unter anderem von Schaltnetzteilen ausgeht und sich über das Stromnetz im gesamten Haus verbreiten kann. Sein Einsatz wird in Abschnitt 4.2.2.3 genauer beleuchtet.

4.1.2.2. Physikalische Größen

Die verschiedenen Quellen von Energiedaten zeigen, wie heterogen die eigentlich verwandten Daten sein können. Dies führt oftmals zu unsauber verwendeten Begriffen und damit zu Verwirrung. So ist beispielsweise meist nicht die physikalische Größe *elektrische Stromstärke* gemeint, wenn von „Stromverbrauch“ die Rede ist. Vielmehr wird oft die Nutzung elektrischer *Energie* so bezeichnet. In Tabelle 4.1 ist eine Auswahl verschiedener physikalischer Größen und der zugehörigen Einheiten aufgeführt, wie sie typischerweise im Kontext von Energiedaten verwendet werden.

Ein klassischer Ferraris-Zähler misst elektrische Energie und zeigt diese als einzige Größe mithilfe eines Zählwerks an. Dementsprechend dürfte in Haushalten diese Größe auch die am meisten beachtete sein – schließlich bestimmt sie im klassischen Energiesystem direkt die Höhe der Stromrechnung am Ende einer Abrechnungsperiode. Der umgangssprachlich dafür verwendete Begriff „Stromverbrauch“ ist zwar nicht korrekt, hat sich aber eingebürgert.

4.1. Technische Datenlebenszyklusanalyse

Tabelle 4.1.: Energiedaten sind zum Teil sehr heterogen. Dementsprechend werden unterschiedliche physikalische Größen und zugehörige Einheiten verwendet.

Physikalische Größe	Einheit
Elektrische Energie	Kilowattstunden (kWh)
Scheinleistung	Voltampere (VA)
Wirkleistung	Watt (W)
Blindleistung	Var (var)
Elektrische Spannung	Volt (V)
Elektrische Stromstärke	Ampere (A)
Netzfrequenz	Hertz (Hz)
Volumen	Kubikmeter (m^3)
Volumenstrom	Kubikmeter pro Sekunde ($m^3 s^{-1}$)
Masse	Kilogramm (kg)
Massenstrom	Kilogramm pro Sekunde ($kg s^{-1}$)
Temperatur	Grad Celsius ($^{\circ}C$)

Energie und Leistung hängen direkt miteinander zusammen. Energie wird als Integral der Leistung über die Zeit erfasst. Während klassische Ferraris-Zähler lediglich indirekt die Leistung angeben können (über die Geschwindigkeit, mit der sich die Zählscheibe dreht), sind intelligente Stromzähler in der Lage, diese direkt auszugeben. In Haushalten wird in der Regel ausschließlich die sogenannte *Wirkleistung* gemessen. Dies ist der Anteil, welcher vollständig in andere Leistungsformen umgesetzt werden kann. Die in größeren Unternehmen oft ebenfalls erfasste *Blindleistung* kann hingegen nicht in andere Formen umgewandelt werden. Sie hat ihre Ursache in induktiven und kapazitiven Verbrauchern. Diese bauen magnetische beziehungsweise elektrische Felder auf, welche wiederum periodisch Energie aufnehmen und anschließend ans Netz abgeben. Auf diese Weise entsteht eine – Phasenverschiebung genannte – zeitliche Differenz des Verlaufs von Spannung und Strom. Die Blindleistung steht zwar nicht zur Umwandlung zur Verfügung, muss aber dennoch bereit-

gestellt werden. Dementsprechend wird bei großen Verbrauchern (vor allem in der Industrie) zusätzlich zur Wirkleistung auch die Blindleistung gemessen und abgerechnet. Diese beiden Größen zusammen ergeben wiederum die *Scheinleistung*. Außer der Art der Leistung spielt auch das Vorzeichen eine Rolle. Typischerweise werden aus Verbrauchersicht positive Leistungsangaben als Verbrauch gewertet. Die Produktion beziehungsweise Rückspeisung wird hingegen mit negativen Werten dargestellt. Die elektrische Leistung wird wiederum durch die Höhe der elektrischen Spannung und des fließenden Stroms bestimmt. Obwohl auch diese Größen von Interesse sein können, werden sie in Haushalten seltener erfasst als die elektrische Leistung oder Energie. Dies gilt auch für die Netzfrequenz der Wechselspannung, welche die Anzahl an Perioden pro Sekunde angibt.

Die bisher beschriebenen Größen sind vor allem elektrischer Natur. Es gibt jedoch weitere Größen, die zwar auf den ersten Blick nicht den Energiedaten zugehörig scheinen, aber dennoch zu diesen gehören. Dies ist beispielsweise bei Gas der Fall. Auch wenn dort typischerweise nach verbrauchten Kilowattstunden (und damit nach Energie) abgerechnet wird, erfolgt eigentlich eine Volumenmessung. Der Gesamtverbrauch wird als Volumen angegeben, der aktuelle Verbrauch – falls durch den Zähler gemessen – durch Volumen pro Zeit. Aus dem gemessenen Volumen wird für die Abrechnung dann die genutzte Energie berechnet. Die beiden Größen können nicht äquivalent verwendet werden, da sich über die Zeit die Zusammensetzung des Gases und damit dessen Brennwert ändern kann. Analog wäre die Verwendung von Masse beziehungsweise Massenstrom bei flüssigen Brennstoffen möglich. Während für elektrische Speicher durchaus auch die verfügbare elektrische Energie in Kilowattstunden angegeben werden kann, ist dies bei thermischen Speichern häufig umständlich. Dementsprechend wird oft die Temperatur im Speicher als Messgröße verwendet. Somit stellen zum Teil auch Temperaturmessreihen Energiedaten dar.

Neben den in Tabelle 4.1 aufgeführten Größen existieren einige weitere, welche als Energiedaten in Betracht kommen. Die Vielfalt zeigt deutlich, welche Bandbreite unterschiedlicher Informationen in einem Energiedatenmanagementsystem abgebildet werden muss. In [EEN15] werden weitere Größen beschrieben, welche als Energiedaten angesehen werden können. Insbesondere wird dort auch auf zugehörige Anwendungsfälle eingegangen.

4.1.2.3. Weitere technische Details

Wie zuvor gezeigt, gibt es eine Vielzahl verschiedener Datenquellen, welche Energiedaten mit unterschiedlicher Bedeutung liefern. Daraus resultiert wiederum eine ebenfalls große Zahl technischer Details, welchen Beachtung geschenkt werden muss. An dieser Stelle werden zwei typischerweise auftretende Besonderheiten erklärt.

Zeitstempel Die Werte eines Energiedatenstroms liegen zeitlich geordnet vor. Dies wird zumeist mithilfe von Zeitstempeln gelöst. Das verwendete Zeitstempelformat ist jedoch keineswegs eindeutig. Die folgenden Zeitstempel sind Beispiele für tatsächlich verwendete Formate:

- 2016-05-04T13:12:11.000Z
- 2016-05-04T13:12:11Z
- 2016-05-04T13:12:11+00:00
- 2016-05-04T15:12:11+02:00
- 04.05.2016 15:12:11
- 1462360331
- ...

Obwohl sich die ersten vier Zeitstempel unterscheiden, entsprechen sie der ISO-Norm 8601 und sind damit äquivalent. Die Angabe von Jahr, Monat, Tag, Stunde, Minute, Sekunde und gegebenenfalls Millisekunden erlaubt ein aufsteigendes Sortieren. Darüber hinaus ist die verwendete Zeitzone angefügt. Z und +00:00 sind gleichwertig und entsprechen der *Koordinierten Weltzeit* (engl. *Coordinated Universal Time (UTC)*). Der vierte Zeitstempel weist eine Zeitzone von +02:00 auf, was der *Mitteleuropäischen Sommerzeit (MESZ)* entspricht. Mit dieser Art von Zeitstempel ist der jeweilige Zeitpunkt eindeutig definiert, auch wenn die Zeitstempel auf den ersten Blick unterschiedlich aussehen mögen. Der fünfte Zeitstempel, der in der alltäglichen Verwendung durchaus üblich ist, ist hingegen bereits nicht mehr eindeutig. Handelt es sich möglicherweise

um eine amerikanische Schreibweise, bei welcher der Monat vor dem Tag kommt? In welcher Zeitzone wurde der Zeitstempel erfasst – Weltzeit oder Lokalzeit? Der letzte hier gezeigte Zeitstempel liegt in *Unixzeit* vor. Diese gibt die verstrichenen Sekunden seit dem 1. Januar 1970 um 00:00 Uhr UTC an. Somit ist auch diese Zeitform eindeutig, wenn auch nicht menschenlesbar.

Es scheint logisch, eine eindeutige, dazu noch menschenlesbare und ge normte Darstellung zu wählen – beispielsweise in Form der Darstellung gemäß ISO 8601. In der Praxis werden jedoch aus unterschiedlichen Gründen oft abweichende Formate genutzt. Daraus können Fehler entstehen, welche einfach hätten vermieden werden können. Solche typischen Fehler treten jeweils am letzten Sonntag im März und im Oktober zu Tage. An diesen Tagen wird die Zeit umgestellt, so dass bei Verwendung der Lokalzeit eine Stunde zu fehlen scheint beziehungsweise eine Stunde doppelt vorkommt. Bei einer Interpretation der Daten müssen solche potenziellen Fehlerquellen beachtet und beseitigt werden.

Dateiformate In der Praxis kommen darüber hinaus viele unterschiedliche Dateiformate vor. Üblich sind unter anderem proprietäre Binärformate, kommaseparierte Dateien sowie XML- und JSON-Dateien. Die verschiedenen Formate weisen jeweils Vor- und Nachteile auf, so dass sich vermutlich auch zukünftig keines dauerhaft durchsetzen kann. Selbst bei vermeintlich simplen Dateiformaten, wie kommaseparierten Dateien, gibt es zwischen unterschiedlichen Datenquellen oft Inkompatibilitäten. Welcher Zeichensatz wird verwendet? Welches Trennzeichen wird verwendet (Komma, Tabulator, Semikolon)? Was ist das verwendete Datumsformat? Was ist der Dezimaltrenner? Darüber hinaus weisen insbesondere Daten mit hoher Messfrequenz zum Teil Besonderheiten auf. Zur Reduktion des Speicherbedarfs beziehungsweise Overheads werden beispielsweise Zeitstempel ausgespart und nur in größeren Abständen aufgeführt. Dazwischenliegende Stempel müssen dementsprechend interpoliert werden.

Es ist offensichtlich, dass durch die Vielzahl an Quellen, die unterschiedlichen physikalischen Größen und die Eigenarten der jeweiligen Produkte eine Unmenge an Datenformaten existiert. Ein geeignetes Energiedatenma-

nagementsystem sollte in der Lage sein, die unterschiedlichsten Arten von Energiedaten zu unterstützen.

4.1.2.4. Resultierende Anforderungen

Die Analyse der ersten Phase des Datenlebenszyklus hat gezeigt, dass Energiedaten überaus heterogen sein können. Es gibt nicht nur unterschiedlichste Datenquellen, sondern auch unterschiedliche Arten von Daten und viele abweichende Datenformate. Bei der Erfassung der Daten müssen diese Gegebenheiten berücksichtigt werden. Ein Energiedatenmanagementsystem sollte nach Möglichkeit keine Einschränkungen hinsichtlich der Datenquellen auferlegen. Alle eingehenden Daten sollten in ein einheitliches Format überführt werden. Dabei muss darauf geachtet werden, dass die Daten mit entsprechenden Metadaten angereichert werden, so dass ihr Informationsgehalt erhalten bleibt und sie auch weiterhin sinnvoll genutzt werden können. Insbesondere bei unterschiedlichen physikalischen Größen in den erfassten Daten ist dieser Schritt unerlässlich. Auch kann es sinnvoll sein, den Daten einen entsprechenden Bezeichner zur eindeutigen Identifikation zuzuweisen, so dass ein eindeutiges Referenzieren der Daten möglich ist.

Je nach Art der Erfassung können die Daten Fehler enthalten und gegebenenfalls lückenhaft sein. Auch die Übertragungsart spielt eine große Rolle (vgl. Abschnitt 4.1.3). Ein Energiedatenmanagementsystem sollte diesbezüglich robust sein und auch fehlerhafte Daten verwalten können. Lücken in den Daten sollten aber falls möglich erkannt und markiert werden. Diese aus der ersten Phase resultierenden Anforderungen sind in den Anforderungsübersichten 1 und 2 zusammengefasst.

Anforderung 1: *Unterstützung unterschiedlicher Datenquellen*

- Keine Einschränkungen bezüglich Datenquellen
 - Überführung in einheitliches Datenformat
 - Anreicherung mit Metadaten
-

Anforderung 2: *Robustheit gegenüber fehlenden und fehlerhaften Daten*

- Kein Abbruch bei Erfassung fehlerhafter Daten
 - Erkennung und Markierung potenzieller Lücken in den Daten
-

4.1.3. Phase II: Übertragung

In der zweiten Phase der Datenlebenszyklusanalyse wird die Übertragung der Daten untersucht. Typische Fragen wären unter anderem die folgenden:

- Welche Übertragungsprotokolle werden verwendet?
- Welchen Weg nehmen die Daten?

Dementsprechend werden die Untersuchungen dieser Phase in zwei Bereiche unterteilt. Zunächst wird untersucht, auf welche Weise die Daten konkret übertragen werden und welche Auswirkungen dies hat. Danach wird analysiert, welche Wege die Daten bei ihrer Übertragung nehmen.

4.1.3.1. Übertragungstechnologien und -protokolle

Energiedaten werden je nach verwendeten Geräten auf unterschiedliche Art und Weise transportiert. Im Folgenden wird ein knapper Überblick zu verwendeten Übertragungstechnologien und -protokollen gegeben.

Da bei der Übertragung von Energiedaten in der Regel keine großen Ansprüche an die Datenrate gestellt werden, steht eine breite Palette unterschiedlicher Technologien zur Verfügung. Sowohl drahtgebundene als auch drahtlose Kommunikationsmöglichkeiten finden Anwendung. Manche Messgeräte verfügen standardmäßig nur über rudimentäre Schnittstellen, können über entsprechende Zusatzmodule jedoch erweitert werden [vgl. u.a. Eas16]. So ist es möglich, je nach Anwendungsfall die geeignete Variante zu wählen. Übliche drahtgebundene Technologien sind unter anderem die folgenden [vgl. u.a. BP11; Fan+10; Fan+12; Fan+13; Gun+11]:

- Ethernet (meist in Verbindung mit dem Protokoll TCP/IP)

- Digital Subscriber Line (DSL)
- Powerline Communication (auch Powerline oder PowerLAN genannt)
- Modbus

Bei drahtlosen Übertragungen findet typischerweise eine der folgenden Technologien Anwendung [vgl. u.a. BP11; Fan+10; Fan+12; Fan+13; Gun+11]:

- Wireless LAN
- ZigBee
- Bluetooth
- Mobilfunk (GSM, GPRS, 3G, LTE)
- Wireless M-Bus

Neben den unterschiedlichen Übertragungstechnologien existieren noch zahlreiche Protokolle, welche auf diesen aufbauen. Aus diesem Grund ist die Heterogenität eine große Herausforderung bei der Bewältigung der eingehenden Daten [Ami12]. Die Situation ist vergleichbar mit der Anzahl und Heterogenität der Energiedatenquellen (vgl. Abschnitt 4.1.2). Es gibt zwar durchaus Bestrebungen, die Kommunikation zu vereinheitlichen [vgl. bspw. Daw+10; VPS12], erfolgreich durchsetzen konnte sich bisher jedoch noch kein Ansatz.

Ein Empfänger von Energiedaten muss dementsprechend mehrere Technologien und Protokolle unterstützen, um eine Vielzahl verschiedener Daten tatsächlich nutzen zu können. Dieser Herausforderung begegnen unterschiedliche Hersteller auf unterschiedliche Art und Weise. Der intelligente Stromzähler des Unternehmens *Discovery* [Dis16] verfügt beispielsweise über ein Modul, welches die Daten per Infrarotschnittstelle von der Ableseeinheit empfängt, dann aber über Ethernet und einen im Haus vorhandenen Router an die Server des Unternehmens überträgt. Alternativ kann auch ein Powerline-Communication-Modul verwendet werden. Bei den drahtlos kommunizierenden Sensoren von *Plugwise* [Plu16] wird hingegen mittels ZigBee ein Netzwerk in Baumstruktur aufgebaut. Ein dedizierter Hauptknoten kommuniziert dann

mit dem ZigBee-USB-Stick und überträgt die Daten aller anderen Messgeräte. Dies sind nur zwei von zahllosen weiteren Beispielen. Sie untermauern jedoch die Anforderung 1.

Insbesondere bei den drahtlosen Kommunikationsmöglichkeiten muss noch ein weiterer Aspekt beachtet werden. Diese sind je nach verwendetem Protokoll nicht oder nicht immer in der Lage, alle Daten fehlerfrei und ohne Verluste zu übertragen. ZigBee, Wireless LAN und Bluetooth sind zwar in der Lage, unter optimalen Bedingungen übliche Wohn- und zum Teil auch Geschäftsgebäude abzudecken. Abschirmende Baumaterialien und andere Störeinflüsse können die Kommunikation jedoch (zeitweise) einschränken und unterbrechen. Falls die Messgeräte über keine Möglichkeit zur Zwischen-speicherung verfügen, kann dies in fehler- und lückenhaften Daten resultieren. Diese Problematik bestärkt Anforderung 2.

An dieser Stelle wurde nur ein kurzer Überblick über Übertragungstechnologien und -protokolle gegeben. Weitere Informationen diesbezüglich sind unter anderem in [Fan+13] und in [Gun+11] zu finden.

4.1.3.2. Der Weg der übertragenen Daten

Zuvor wurde die Art der Datenübertragung betrachtet. Nun folgt die Analyse der möglichen Wege, die die Daten während ihrer Übertragung nehmen können. Diese Betrachtung erfolgt vor allem im Hinblick auf die Kontrollmöglichkeiten des Nutzers: Verlassen die Daten den Haushalt und entziehen sich damit der Kontrolle des Nutzers oder nicht?

Generell können drei Arten von Übertragungswegen unterschieden werden (vgl. Abbildung 4.2). Im ersten Fall werden die Daten im Haushalt erfasst und nur lokal übertragen und verarbeitet. Dem gegenüber steht der zweite Fall, in welchem Daten im Haushalt erfasst, dann aber zu einem Dienstleister außerhalb des Haushalts übertragen werden. Eine Kombination beider Extreme stellt den dritten Fall dar: Die Daten werden sowohl lokal übertragen und verarbeitet als auch an Dritte weitergegeben.

Fall 1: Lokale Übertragung Die Energiedaten werden ausschließlich lokal von einem oder mehreren Messgeräten zu einer Erfassungseinheit übertragen. Obwohl Energiedaten personenbezogene Daten sind und eigentlich beson-

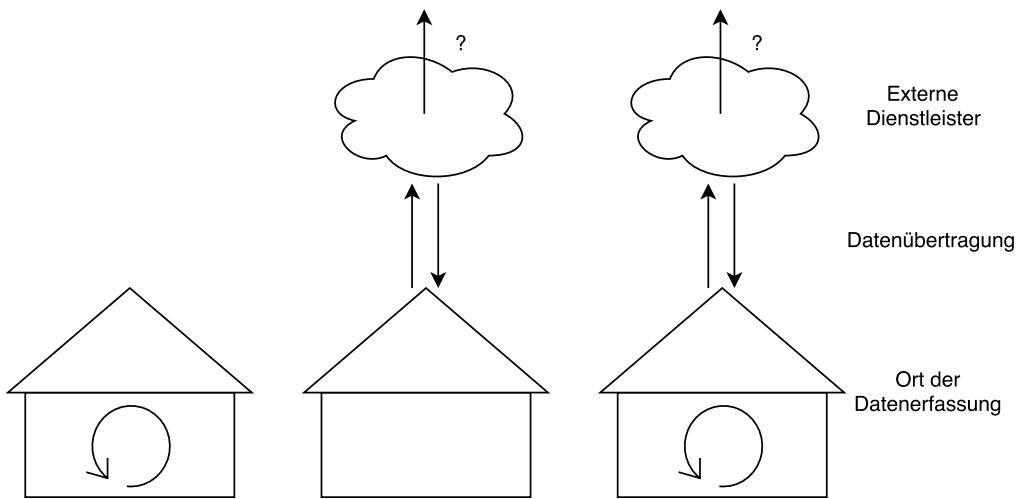


Abbildung 4.2.: Die Übertragungswege von Energiedaten können in drei Kategorien eingeteilt werden. Unterschieden wird zwischen einer rein lokalen Übertragung, einer Übertragung zu externen Parteien ohne lokale Übertragung, sowie einer Mischform der beiden vorherigen.

ders geschützt werden sollten, ist dieser Fall nur selten anzutreffen. Dies kann beispielsweise bei bestimmten Geräten der Fall sein, welche lediglich einen klassischen Ferraris-Zähler ablesen oder einen Zähler mit Impulsschnittstelle ohne weitere Kommunikationsmöglichkeiten auswerten. Es existiert eine Reihe von Open-Source-Projekten, welche sich der Entwicklung solcher Systeme widmet, die eine rein lokale Auswertung ermöglichen.

Darüber hinaus besteht die Möglichkeit, manche kommerziellen Produkte rein lokal zu nutzen. Am Beispiel der Sensoren von *Plugwise* (vgl. Abschnitt 3.1.2.3) wird dies deutlich. Es ist durchaus möglich, der Einheit, welche die Messwerte von den einzelnen Zwischensteckdosen erhält, den Netzzugang zu verwehren beziehungsweise eine alternative Anwendung zur Datenerfassung zu verwenden. Somit können die Daten das Haus nicht verlassen. Dies ist allerdings gemäß den Nutzungsbedingungen nur gegen Bezahlung einer jährlichen Gebühr ohne weitere Einschränkungen möglich.

Fall 2: Übertragung an Dritte Die Erkenntnisse, die sich aus Energiedaten gewinnen lassen, können für Unternehmen von großem Wert sein. Vermutlich ist das der Grund, warum bei zahlreichen kommerziellen Produkten die

Energiedaten zwar lokal erfasst, dann aber zu einem externen Dienstleister übertragen werden. Als Beispiel soll an dieser Stelle erneut das Unternehmen *Discovergy* (vgl. Abschnitt 3.1.2.2) dienen. Deren Stromzähler überträgt Daten über eine Infrarotschnittstelle an ein Zusatzmodul, welches als Schnittstelle zu einem im Haushalt befindlichen Router dient. Die Energiedaten werden auf die Server des Unternehmens übertragen und dort weiter ausgewertet. Der Nutzer greift dann über eine Internetseite auf die verarbeiteten Daten zu. Es wäre zwar technisch möglich, dass ein direkter Zugriff auf die Daten des Stromzählers erfolgt. Dies ist aber nicht vorgesehen.

Diese Art der Datenübertragung ist die am häufigsten anzutreffende. Die externe Datenübertragung bietet aus Sicht der Unternehmen den Vorteil, dass die Daten der Kunden ausgewertet werden können, was nicht der Fall wäre, wenn die Daten im Haushalt verblieben. Den Kunden werden zum Teil Mehrwertdienste angeboten, welche die Dienstleistung des Unternehmens attraktiver machen sollen. Sie zielen oft darauf ab, die Komplexität vor den Kunden zu verbergen und eine „einfache“ Auswertung zu ermöglichen. Beispielsweise wird damit geworben, den Energieverbrauch „im Haus“ auch von unterwegs über ein Smartphone auswerten zu können. Insbesondere technisch unbedarften Personen dürfte dabei aber häufig nicht klar sein, dass die Daten den Haushalt dazu verlassen müssen und auf einer Internetseite eines Dienstleisters dargestellt werden.

Die Übertragung der Daten an einen externen Dienstleister bringt einen weiteren potenziellen Nachteil für die betroffenen Personen mit sich. Die Nutzer haben keine realistische Möglichkeit zu prüfen, ob die Daten an weitere Parteien weitergeleitet werden (dargestellt durch „?“ in Abbildung 4.2). Sie sind darauf angewiesen, dem Unternehmen, welches ihre Daten erhält, zu vertrauen. Dieses Vertrauen muss sich einerseits auf die Einhaltung der gegebenen Datenschutzerklärungen erstrecken. Andererseits muss aber auch der Sicherheit der verwendeten Systeme vertraut werden.

Fall 3: Mischform Beim dritten Fall handelt es sich um eine Mischform der beiden vorherigen Fälle. Die Daten werden einerseits lokal übertragen und ausgewertet, andererseits aber auch an Externe weitergeleitet. Erneut kann hier *Plugwise* (vgl. Abschnitt 3.1.2.3) als Beispiel für einen kommerzi-

ellen Anbieter dienen. Einerseits werden die Daten auf einen lokalen, dem Kunden gehörenden, Rechner übertragen und können dort ausgewertet werden. Andererseits erfolgt eine Übertragung an den Hersteller der Messgeräte. Da die Daten bereits lokal ausgewertet werden können, ist nicht klar, ob die Übertragung der Daten an den Hersteller einen Mehrwert für den Kunden darstellt.

Dieser Fall zeigt, dass die Übertragungswege für Nutzer oft intransparent sein können. Auch wenn augenscheinlich die Daten lokal übertragen werden, werden sie jedoch gleichzeitig auch an externe Server übertragen. Die im zweiten Fall geäußerten Bedenken gelten natürlich auch hier analog.

4.1.3.3. Resultierende Anforderungen

Die Untersuchungen der zweiten Phase des Datenlebenszyklus unterstreichen die Anforderungen 1 und 2 (vgl. Abschnitt 4.1.2.4). Darüber hinaus ergeben sich weitere Anforderungen, die ein Energiedatenmanagementsystem erfüllen sollte.

Zunächst sollte das Energiedatenmanagementsystem eine lokale Sammelstelle für alle Daten eines Haushalts sein. Unabhängig davon, was mit den Daten geschehen soll, müssen sie zuerst zu einer Stelle unter der Kontrolle der Nutzer übertragen werden. Von dieser aus kann dann systematisch entschieden werden, ob die Daten an Externe weitergegeben werden sollen oder nicht. Wird die Übertragung der Daten nicht von Anfang an kontrolliert, bedeutet dies unweigerlich einen Kontrollverlust für den Nutzer.

Darüber hinaus muss ein Energiedatenmanagementsystem für Transparenz und Sicherheit bei der Datenübertragung sorgen. Standardmäßig dürfen keine Daten die Kontrolle des Nutzers verlassen, ohne dass dieser davon erfährt. Werden die Daten vom zentralen Sammelpunkt an weitere Stellen übertragen, muss dies transparent und für den Nutzer nachvollziehbar geschehen. Darüber hinaus darf die Datenübertragung ausschließlich auf gesicherten Kanälen erfolgen, so dass ein unbefugtes Abgreifen der Daten unterbunden oder zumindest signifikant erschwert wird (vgl. Anforderung 9).

Die Anforderungen sind in den Anforderungsübersichten 3 und 4 zusammengefasst.

Anforderung 3: Lokale Sammelstelle unter Kontrolle des Nutzers

- Datenerfassung innerhalb der Kontrolle der Nutzer an einer lokalen Stelle
 - Daten dürfen nicht parallel durch Dritte beziehungsweise extern erfasst werden
-

Anforderung 4: Transparenz der Datenübertragungswege

- Keinerlei externe Übertragung ohne explizites Einverständnis der Nutzer
 - Nachvollziehbarkeit aller Übertragungen
-

4.1.4. Phase III: Speicherung

Die Untersuchung der dritten Phase des Datenlebenszyklus betrifft die Speicherung. Typische Fragen wären unter anderem die folgenden:

- Welche Datenmengen müssen gespeichert werden?
- Wo werden die Daten gespeichert?
- Welche Speichertechnologien bieten sich für Energiedaten an?

In den weiteren Unterabschnitten werden diese Fragestellungen systematisch untersucht. Zunächst erfolgt eine Abschätzung der anfallenden Datenmenge in Haushalten.

4.1.4.1. Abschätzung der Datenmengen

In der Literatur [vgl. u.a. Fan+13; Gun+11] finden sich unterschiedliche Angaben zu üblichen Datenraten in intelligenten Stromnetzen. Die Abschätzungen beziehen sich in der Regel auf die maximal erreichbaren Übertragungsraten

der unterschiedlichen Technologien, welche in Verwendung sind. Die Angaben reichen von wenigen Kilobit pro Sekunde (beispielsweise GSM) bis zu mehreren Megabyte pro Sekunde (beispielsweise Powerline Communication).

An dieser Stelle soll hingegen eine Abschätzung des Datenaufkommens anhand der zu erwartenden Daten erfolgen. Betrachtet wird ein Haushalt oder ein kleineres Unternehmen mit jeweils mehreren Quellen von Energie-daten. Es wird angenommen, dass keine der Quellen zeitlich hochauflöste Energiedaten liefert. Die dazu nötigen Messgeräte sind teuer und erfordern leistungsfähige Datenverarbeitungsmöglichkeiten. Dementsprechend ist davon auszugehen, dass diese Geräte keine weite Verbreitung erfahren werden. Da es sich um eine Abschätzung der Obergrenze handelt, wird weiter angenommen, dass in jeder Sekunde ein Datenpaket mit einem oder mehreren Messwerten erfasst werden muss. Da es sich bei den Messwerten in der Regel um Gleitkommazahlen handelt, ist eine angenommene Größe von einem Kilobyte mehr als angemessen. Somit ergeben sich folgende Annahmen:

- Ein Haushalt beziehungsweise ein kleineres Unternehmen verfügt über maximal zwanzig Datenquellen.
- Jede Datenquelle liefert maximal sekündlich neue Datenpakete.
- Ein Datenpaket hat eine maximale Größe von einem Kilobyte.

Diese Annahmen ergeben eine Datenrate von 20 KB/s. Pro Tag beläuft sich die Datenmenge demnach auf rund 1,7 GB und jährlich addiert sich das auf bis zu 630 GB. Auch wenn diese Datenmengen auf den ersten Blick groß erscheinen mögen, sind sie dennoch problemlos mit heute üblicher Technologie zu bewältigen. Digitale Aufnahmegeräte für Überwachungskameras beispielsweise müssen bereits bei einzelnen Kameras deutlich größere Datenraten bewältigen. Darüber hinaus handelt es sich um eine vorsichtige Abschätzung des Datenvolumens – in der Praxis dürften die einzelnen Pakete deutlich kleiner als ein Kilobyte sein. Darüber hinaus können die Daten im Zuge der Speicherung gegebenenfalls komprimiert werden. Auch die Einschränkung auf niederfrequente Messungen ist realitätsnah. Sollten dennoch Messungen mit hoher Messfrequenz vorgenommen werden, ist ein entsprechend leistungsfähiges System aber unumgänglich [vgl. u.a. Maa+12]. Dahingegen sind die

in Haushalten oder kleineren Unternehmen typischerweise zu erwartenden Datenmengen auch mit Standardhardware problemlos zu bewältigen.

4.1.4.2. Speicherort

Der Speicherort hängt eng mit den unterschiedlichen Übertragungswegen der Energiedaten zusammen (vgl. Abbildung 4.2). Die vergleichsweise geringen Datenraten erlauben bei Verwendung eines Breitbandinternetanschlusses auch problemlos das Übertragen der Daten zu externen Dienstleistern. Dementsprechend kann die Fragestellung nach dem Speicherort analog zu den Übertragungswegen beantwortet werden.

Aus Nutzersicht kann es durchaus von Vorteil sein, die Daten bei einem externen Dienstleister zu speichern. Für die Nutzer entfallen dann Anschaffung und Unterhalt der Hardware sowie das Anfertigen von Backups. Allerdings bringt diese Art der Speicherung den Nachteil mit sich, dass der Haushalt die Kontrolle über die Daten verliert. Sollte ein Energiedatenmanagementsystem externe Speichermöglichkeiten nutzen, müssten die Daten verschlüsselt sein, damit die Privatsphäre der Nutzer gewahrt bleibt. Andernfalls wäre der Schutz lediglich auf das Vertrauen gegenüber dem Dienstleister beschränkt.

4.1.4.3. Formen der Datenspeicherung

Energiedaten liegen typischerweise in Form einer Zeitreihe vor oder können in eine solche umgewandelt werden. Zu den Energiedaten gehören jedoch weitere Informationen, welche die eigentlichen Daten genauer beschreiben – die Metadaten. Es wäre theoretisch möglich, jeden einzelnen Datenpunkt um die zugehörigen Metadaten (bspw. Datenquelle, verwendete Einheiten, ...) zu erweitern und diese gemeinsam abzuspeichern. Dieses Vorgehen würde aber zum wiederholten Abspeichern derselben Daten führen und wäre äußerst ineffizient. Dementsprechend ist es üblich, die Metadaten getrennt vorzuhalten und die eigentlichen Zeitreihendaten zu referenzieren. Daher bietet es sich an, die Speicherung der Metadaten sowie der Zeitreihendaten getrennt zu betrachten.

Speicherung der Metadaten Die Metadaten enthalten Informationen über die jeweiligen Zeitreihen. Dabei kann es sich beispielsweise um die Datenquelle, die physikalische Größe, welche gemessen wurde, oder auch den Besitzer der Daten handeln. Aufgrund der geringen Größe der Metadaten ist der Speicherbedarf nicht kritisch. Wichtiger ist, dass die Metadaten effizient durchsuchbar sind und die gewünschten Zeitreihendaten einfach gefunden werden können. Daher bieten sich unter anderem zwei Arten von Datenbanken für die Metadatenhaltung an. Zum einen relationale Datenbanken, zum anderen dokumentenorientierte Datenbanken.

Relationale Datenbanken sind etablierte Werkzeuge und seit Jahrzehnten im Einsatz. Die Daten unterliegen einem festen Datenbankschema und werden in tabellenähnlicher Form gespeichert. Bekannte Vertreter sind neben vielen weiteren *SQLite*² und *MySQL*³. Dokumentenorientierte Datenbanken weisen kein festes Datenbankschema auf und sind dementsprechend flexibler einzusetzen. Dort werden die Daten nicht in tabellarischer Form gespeichert, was nachträgliche Änderungen an den Daten vereinfachen kann. Im Gegensatz zu relationalen Datenbanken handelt es sich aber um eine noch recht junge Technologie, deren populärster Vertreter *MongoDB*⁴ sein dürfte. Unabhängig von der konkreten Form der Datenspeicherung sollte ein Energiedatenmanagementsystem ein effizientes Durchsuchen der Metadaten ermöglichen.

Speicherung der Zeitreihen Zeitreihen könnten auch in relationalen oder dokumentenorientierten Datenbanken gespeichert werden. Ein Messpunkt entspricht damit einer Zeile in einer relationalen Datenbank beziehungsweise einem Dokument in einer dokumentenorientierten Datenbank. Diese Art der Speicherung kann bei einer großen Zahl von Messpunkten – wie sie auch in Haushalten durchaus vorkommen kann – jedoch ineffizient und langsam sein [Jia+09]. Daher bietet es sich an, explizit auf Zeitreihen zugeschnittene Lösungen zu verwenden.

Wie die Bezeichnung bereits verrät, stehen mit Zeitreihendatenbanken spezialisierte Produkte für Zeitreihen zur Verfügung. Die Datenbanken sind nicht nur auf eine effiziente Speicherung getrimmt, sondern ermöglichen

²<https://www.sqlite.org/>

³<https://www.mysql.de/>

⁴<https://www.mongodb.com/de>

auch Operationen auf den Zeitreihen, welche bei deren Analyse hilfreich sein können. Wie die dokumentenorientierten Datenbanken ist auch diese Art von Datenbanken recht jung. Beispielhaft zu nennen sind *OpenTSDB*⁵ und *InfluxDB*⁶.

Alternativ kann eine dateibasierte Lösung für die Speicherung von Zeitreihen verwendet werden. Neben den klassischen kommaseparierten Dateien und weiteren Standardformaten stehen auch auf große Datenmengen spezialisierte Lösungen zur Verfügung. *HDF5*⁷ ist ein Beispiel für ein solches Dateiformat. Als Dateiformat bietet es im Gegensatz zu den Zeitreihendatenbanken zwar keine direkten Operationen auf den Zeitreihen an, ermöglicht dafür aber eine äußerst effiziente Ein- und Ausgabe großer Datenmengen.

Wahl einer geeigneten Speicherform Die Auswahl einer Speicherform ist stark abhängig von der konkreten Ausgestaltung eines Energiedatenmanagementsystems. Erschwerend kommt hinzu, dass insbesondere die vergleichsweise jungen Speicherlösungen wie spezialisierte Zeitreihendatenbanken einem starken Wandel unterliegen. Zusätzlich entstehen im Kontext der *NoSQL*-Datenbanken neue Produkte, welche sich möglicherweise ebenfalls als geeignet erweisen. Aus diesem Grund beschränkt sich dieser Abschnitt auf eine kurze Übersicht potenzieller Kandidaten für die Speicherung von Energiedaten. Eine mögliche Lösung wird im Rahmen des Demonstrators in Abschnitt 5.3.3 aufgezeigt.

4.1.4.4. Resultierende Anforderungen

Die in Haushalten und kleinen Unternehmen anfallenden Energiedaten erreichen üblicherweise kaum eine Menge, die die Handhabung der Daten signifikant erschwert. Daher resultieren aus der Datenrate und der Datenmenge zunächst keine besonderen Anforderungen.

Die Analyse der dritten Phase des Datenlebenszyklus hat gezeigt, dass bei der Speicherung von Energiedaten zwischen den eigentlichen Zeitreihendaten und den Metadaten unterschieden werden sollte. Während bei den

⁵<http://opentsdb.net/>

⁶<https://influxdata.com/>

⁷<https://www.hdfgroup.org/HDF5/>

4.1. Technische Datenlebenszyklusanalyse

Metadaten eine effiziente Suche im Vordergrund steht, sollten Zeitreihendaten möglichst effizient gespeichert und auch wieder ausgelesen werden können. Dementsprechend sollte ein Energiedatenmanagementsystem für die jeweilige Art von Daten eine geeignete Speicherform ermöglichen.

Da es in Haushalten aber auch in Unternehmen das Bestreben geben kann, Energiedaten wegen ihrer Größe nicht vor Ort, sondern bei einem Dienstleister zu speichern, sollten die Daten bei einer externen Speicherung besonders geschützt werden. Dies ist beispielsweise durch die Verwendung geeigneter Verschlüsselungsverfahren möglich. Daher sollte ein Energiedatenmanagement bei der Verwendung externer Speicher eine Verschlüsselung ermöglichen.

Die Anforderungen sind in den Anforderungsübersichten 5 und 6 zusammengefasst.

Anforderung 5: Unterstützung einer geeigneten Speicherform in Abhängigkeit der Datenart

- Energiedaten bestehen aus Zeitreihen und Metadaten
 - Zeitreihen erfordern eine effiziente Speicherung
 - Metadaten erfordern eine gute Durchsuchbarkeit
 - Gegebenenfalls Nutzung externer Speicher
-

Anforderung 6: Verschlüsselung bei externer Speicherung

- Optionale externe Speicherung kann von Nutzern erwünscht sein (bspw. für vereinfachtes Backup), auch wenn lokale Speicherung präferiert werden sollte
 - Bei externer Speicherung Schutz vor unberechtigtem Zugriff durch Verschlüsselung
-

4.1.5. Phase IV: Analyse

Bei der vierten Phase der Untersuchung des Datenlebenszyklus steht die Verarbeitung der Daten – deren Analyse – im Vordergrund. Typische Fragen zu diesem Abschnitt des Datenlebenszyklus wären unter anderem die folgenden:

- Welchen Zweck verfolgen die Analysen der Energiedaten?
- Welche Operationen sind dazu vorrangig nötig?

4.1.5.1. Zweck der Analysen

Es gibt zahlreiche Anwendungsfälle für die Auswertung und Verarbeitung von Energiedaten. Nachfolgend wird eine Auswahl möglicher Ziele einer Energiedatenanalyse angelehnt an [JKD12] aufgelistet.

Beobachten des Lastverlaufs und Erstellung von Prognosen Energiedaten ermöglichen einen detaillierten Einblick in das Verhalten der beobachteten Haushalte beziehungsweise Bewohner. Eine Analyse der Daten ermöglicht es, Verhaltensmuster besser zu verstehen und dadurch genauere Prognosen zu erstellen [vgl. u.a. JKD12].

Visualisierung des Energieverbrauchs Geschickte Aufbereitung und Visualisierung können dazu führen, dass Bewohner besser verstehen, wie sie ihren Energieverbrauch beeinflussen können. Eine zuvor abstrakte Größe wird so auch für Laien besser verständlich, was sich wiederum in positiven Verhaltensänderungen niederschlagen kann [vgl. u.a. CTG08; Wei+12b; WN07].

Effizienzanalyse und -beratung Die Auswertung der Energiedaten ermöglicht es, sowohl die Effizienz der Geräte als auch deren Nutzung zu bewerten. Dem Nutzer können anschließend Hinweise bezüglich seines Verhaltens oder seiner Geräte gegeben werden, um die Energieeffizienz zu steigern [vgl. u.a. Wei+12a; Wei+12b]. Auch die vorausschauende Instandhaltung (engl. *predictive maintenance*) auf Basis von Energiedaten fällt in diese Kategorie.

Lastmanagement Eine Auswertung der Energiedaten kann Potentiale für Lastmanagement aufzeigen, indem nach einer möglichen Flexibilität des Lastgangs gesucht wird. Die Nutzung solcher Potentiale ermöglicht den Übergang von einer verbrauchsorientierten Produktion zu einer produktionsorientierten Nachfrage [vgl. u.a. Hir15].

Abrechnung des Energieverbrauchs Energiedaten werden auch zu Abrechnungszwecken ausgewertet. Insbesondere können auch zeitvariable Tarife oder Tarife mit Leistungsgrenzen zum Einsatz kommen.

Entdeckung von Straftaten In Energiedaten enthaltene Informationen geben zum Teil Anhaltspunkte für die Entdeckung von Straftaten. Dies kann beispielsweise Stromdiebstahl in Form einer Umgehung des Stromzählers sein. Plötzlich signifikant gesunkener Stromverbrauch gäbe in Folge Anlass zur Vermutung, dass der Zähler umgangen wird. Darüber hinaus können bestimmte Verbrauchsmuster beispielsweise auf den Betrieb von Drogenplantagen hinweisen [vgl. u.a. LMW10].

Erstellung von Nutzerprofilen Eine Auswertung der Energiedaten erlaubt die Ableitung von Nutzer- und Verhaltensprofilen. Diese können für personalisierte Werbung, Marktforschung oder auch fragwürdigere Zwecke verwendet werden [vgl. u.a. LMW10].

Diese Auflistung erhebt nicht den Anspruch, vollständig zu sein, stellt aber in Kürze die wichtigsten Arten der Energiedatennutzung dar. Einzelne Aspekte werden im Laufe der Arbeit näher beleuchtet (u.a. in den Abschnitten 2.2.3 und 4.2.3). Auch wenn die Analysen auf sehr unterschiedliche Art und Weise erfolgen können, weisen sie dennoch gemeinsame Operationen auf, welche im nächsten Abschnitt untersucht werden.

4.1.5.2. Operationen auf Energiedaten

Bei der Verarbeitung von Energiedaten funktionieren die einzelnen Verfahren im Detail sehr unterschiedlich. Insbesondere bei der Extraktion von Informationen aus Energiedaten (vgl. Abschnitt 4.2.2) werden zum Teil aufwendige

Berechnungen durchgeführt. Je nach Ansatz können diese aber vollkommen unterschiedlicher Natur sein. Aus diesem Grund erscheint es wenig sinnvoll, in einem Energiedatenmanagementsystem alle denkbaren Operationen zu ermöglichen. Allerdings sollte ein Grundstock zur Verfügung gestellt werden, welcher die einzelnen Verfahren unterstützen kann. Dies betrifft insbesondere die vorbereitenden Schritte, welche die meisten Verfahren gemeinsam haben. Folgende Operationen sollten zumindest in grundlegender Funktion auf den Zeitreihen der Energiedaten möglich sein.

Selektion (nach Datenquelle, nach Zeitraum) Häufig sind aus einer großen Menge von Energiedaten nur gewisse Bereiche von Interesse. Dementsprechend müssen die gewünschten Bereiche selektiert und extrahiert werden können. Die Auswahl nach Datenquellen beziehungsweise Geräten kann anhand der Metadaten erfolgen. Je nach Speicherform wird somit eine Datei, eine Tabellenspalte oder ähnliches ausgewählt. Aus dieser räumlichen Auswahl muss anschließend der gewünschte Zeitraum extrahiert werden können. Diese zeitliche Selektion sollte sowohl möglichst feingranular als auch effizient durchführbar sein.

Umrechnung in andere Abtastrate (Resampling) Nicht immer werden die Energiedaten mit einer festen Frequenz erfasst. Darüber hinaus sind je nach Anwendungsfall unterschiedliche Auflösungen nötig. Aus diesem Grund ist es erforderlich, die Energiedaten mit einer anderen Abtastrate darzustellen (engl. *resampling*). Wenn die zeitliche Auflösung erhöht wird, müssen fehlende Werte geeignet interpoliert werden. Wenn die zeitliche Auflösung hingegen verringert wird, müssen aus den bestehenden Werten geeignete neue Werte errechnet werden.

Addition, Subtraktion Oft ist nicht eine einzelne Zeitreihe von Interesse, sondern die Summe oder Differenz mehrerer. Dies kann beispielsweise der Fall sein, wenn der Gesamtverbrauch aller Haushaltsgeräte bestimmt werden soll, aber ohne die Beleuchtung. Eine sequentielle beziehungsweise zeilenweise Berechnung aller Werte ist in diesem Fall weniger effizient als eine vektorbasierte Lösung.

Multiplikation, Division Bei der Umrechnung der Zeitreihen in verschiedene Einheiten kann es nötig sein, die Messwerte zu multiplizieren. Ebenso kann beispielsweise zur Bildung von Durchschnitten eine Division nötig sein. Dementsprechend sollten auch diese Operationen effizient durchführbar sein. Mit diesen Operationen lässt sich bereits ein breites Spektrum von Anwendungsfällen abdecken. Insbesondere wird durch sie aber eine geeignete Vorverarbeitung der Daten ermöglicht, so dass diese in der gewünschten Qualität weitergegeben werden können. Somit ist der Grundstein gelegt, um durch geschickte Aufbereitung nur so viele Daten wie notwendig weiterzugeben.

4.1.5.3. Resultierende Anforderung

Energiedaten werden zu unterschiedlichen Zwecken analysiert und ausgewertet. Die große Zahl möglicher Methoden und Verfahren macht es unmöglich, alle Analysen durch das Energiedatenmanagementsystem ausführen zu lassen. Viele Verfahren haben jedoch Gemeinsamkeiten bei der Vorverarbeitung von Energiedaten, welche durchaus lokal durchgeführt werden kann. Ein Energiedatenmanagementsystem sollte demnach über die entsprechenden Operationen verfügen, um diese Schritte lokal durchführen zu können. Die Anforderung ist in der Anforderungsübersicht 7 zusammengefasst.

Anforderung 7: Lokale Vorverarbeitung der Daten

- Anforderungsgerechte Aufbereitung der Rohdaten (bspw. Reduktion der Aussagekraft und Qualitätsreduktion)
 - Bereitstellung grundlegender Operationen zur Vorbereitung der Analysen
-

4.1.6. Phase V: Verteilung

In der fünften Phase der Datenlebenszyklusanalyse wird die Verteilung der Daten betrachtet. Im Gegensatz zur zweiten Phase werden nicht die einge-

henden Daten untersucht, sondern die vom System ausgehenden. Typische Fragestellungen sind unter anderem die folgenden:

- An wen werden die Daten weitergegeben?
- Welche Daten werden weitergegeben?
- Welche Datenformate und Protokolle werden verwendet?

4.1.6.1. Potenzielle Datenempfänger und Daten von Interesse

Das im Juni 2016 vom Bundestag verabschiedete Messstellenbetriebsgesetz (vgl. Abschnitt 3.2.1.3) sieht eine Reihe von Akteuren vor, welche Zugriff auf die Daten eines intelligenten Stromzählers erhalten müssen, um klar definierte Aufgaben erfüllen zu können. Zu dieser Gruppe gehören beispielsweise Messstellenbetreiber, Netzbetreiber und Bilanzkreisverantwortliche. Darüber hinaus kommt als Datenempfänger potenziell jede Stelle in Betracht, welche über eine schriftliche Einwilligung des Nutzers verfügt [Deu16a]. Denkbar wären an dieser Stelle Forscher und Gerätehersteller, aber auch Werbetreibende und viele weitere Dienstleister. Während die im Gesetz explizit aufgeführten Parteien die Daten für wohldefinierte Zwecke erhalten müssen, ist die Weitergabe an andere Parteien nicht verpflichtend für den Nutzer. Dem Nutzer steht es – nicht zuletzt durch das Recht auf informationelle Selbstbestimmung – frei, zu wählen, wer die Daten erhalten darf.

Neben intelligenten Zählern gibt es jedoch weitere Quellen für Energiedaten, welche nicht zwangsläufig vom Messstellenbetriebsgesetz abgedeckt sind. Dementsprechend genießen sie auch nicht unbedingt denselben Schutz wie Daten von intelligenten Stromzählern. Umso wichtiger ist es für den Nutzer, bewusst entscheiden zu können, welche Daten weitergegeben werden dürfen und welche nicht. Dieser Aspekt wird ausführlich in Abschnitt 4.2.4 behandelt.

Aus technischer Sicht ergibt sich das Problem, dass der Kreis potenzieller Datenempfänger kaum eingeschränkt werden kann. Verschiedene Interessenten mit unterschiedlichen Ansprüchen an die Daten kommen als Empfänger in Frage. Die steigende Verbreitung von Energiemessgeräten wird dazu führen, dass neue Geschäftsfelder erschlossen und neue Dienstleistungen angeboten werden. Im Umkehrschluss wird damit auch die Zahl an Dateninteressenten

steigen. Ein Energiedatenmanagementsystem sollte die vorverarbeiteten Daten – sofern der Nutzer es gestattet – möglichst flexibel weitergeben können. Je nach Empfänger sind unterschiedliche Qualitäten und Formate nötig, welche wiederum bereitgestellt werden sollten.

4.1.6.2. Dateiformate und Protokolle zur Weitergabe

Ebenso wie der Kreis der Dateninteressenten lässt sich auch die Wahl der möglichen Dateiformate und Protokolle kaum einschränken. Die konkrete Ausgestaltung ist abhängig von der anfragenden Partei. Diese Situation ist vergleichbar mit den ersten beiden Phasen des Datenlebenszyklus – der Erfassung und Übertragung (vgl. Abschnitt 4.1.2 und 4.1.3). Da dort bereits eine detaillierte Analyse durchgeführt wird, ist der folgende Abschnitt kurzgefasst.

Analog zur Übertragung können auch bei der Verteilung unterschiedliche Datenformate nötig sein. Ein Energiedatenmanagementsystem sollte geeignete Lösungen in Abhängigkeit von der anfragenden Partei anbieten. Das System nimmt damit gleichsam die Funktion einer Schnittstelle ein, welche die Energiedatenquellen mit den Dateninteressenten verbindet. Dementsprechend sollte ein Energiedatenmanagementsystem bezüglich der Datenausgabe ebenso wenig Einschränkungen haben wie bei der Dateneingabe. Dies impliziert jedoch keineswegs, dass die erfassten und gespeicherten Daten oder gar die Rohdaten unverarbeitet weitergegeben werden. Die Daten müssen zuvor eingehend analysiert werden, um ihren Informationsgehalt abzuschätzen. Gemäß Nutzerwunsch muss nach der Vorverarbeitung eine Qualitätsreduktion vorgenommen werden, um den Informationsgehalt geeignet zu reduzieren (vgl. Abschnitt 4.2.4). Dies erlaubt eine effektive Kontrolle der Informationen, die den Einflussbereich des Nutzers verlassen.

Sobald die Daten übertragen werden, besteht die Gefahr, dass sie von unbefugten Dritten abgefangen werden. Aus diesem Grund müssen sie auf dem Transportweg angemessen geschützt werden. Die Risiken einer unverschlüsselten Übertragung dürften hinreichend bekannt sein (vgl. Abschnitt 3.1.2.2), weshalb eine Transportverschlüsselung vorausgesetzt wird. Im Zuge der Verschlüsselung sollten nicht nur die zu übertragenden Energiedaten verschlüsselt werden, sondern die gesamte Kommunikation inklusive eingehender Anfragen.

4.1.6.3. Resultierende Anforderungen

An dieser Stelle wird vor allem der technische Aspekt der Datenweitergabe betrachtet. Aus dieser Sicht ergibt sich in der Phase der Verteilung hauptsächlich ein Anspruch an die Flexibilität. Ein Energiedatenmanagementsystem muss die Energiedaten – falls vom Nutzer genehmigt – möglichst in der vom Anfragenden gewünschten Datenqualität und im richtigen Format zur Verfügung stellen. Die Übertragungswege der Daten müssen gesichert werden, um ein Abhören der Daten zu verhindern oder zumindest zu erschweren. Diese Anforderungen sind in den Anforderungsübersichten 8 und 9 zusammengefasst.

Da eine Weitergabe personenbezogener Daten auch die Privatsphäre der Nutzer betrifft, werden deren Interessen gesondert in Abschnitt 4.2 untersucht.

Anforderung 8: Unterstützung unterschiedlicher Datenausgabeformate

- Vergleichbar mit Erfassung und Übertragung
 - Flexible Unterstützung unterschiedlicher Formate
 - Keine Weitergabe unverarbeiteter (Roh-)Daten
-

Anforderung 9: Sicherheit der Datenübertragungswege

- Daten sind bei Übertragung gefährdet
 - Jegliche Datenübertragung (inkl. Datenanfragen) muss geschützt werden
-

4.1.7. Phase VI: Löschung

In der letzten Phase der Datenlebenszyklusanalyse wird die Löschung der Daten untersucht. Gleichzeitig endet mit ihr eine Iteration des Datenlebenszyklus. Typische Fragestellungen sind unter anderem die folgenden:

- Erfolgt überhaupt eine Löschung?
- Werden Daten gelöscht?
- Wie und wann werden die Daten gelöscht?

4.1.7.1. Löschung bereits übertragener Daten

Das Bundesdatenschutzgesetz sieht vor, dass personenbezogene Daten – meist auf Antrag der Betroffenen oder nach Ablauf von Fristen – gelöscht werden müssen (vgl. Abschnitt 3.2.1.1). Auch das neue Messstellenbetriebsgesetz (vgl. Abschnitt 3.2.1.3) enthält mehrere die Löschung von Daten betreffende Paragrafen. In Abhängigkeit der jeweiligen (Markt-)Rolle werden dort unterschiedliche Regeln für die Verwendung und Löschung der Daten angegeben. Damit ist für betroffene Personen zumindest der regulatorische Rahmen festgelegt, in welchem eine Löschung der Daten verlangt werden kann. Es handelt sich allerdings nicht um eine technische Vorkehrung, die sicherstellt, dass die Daten auch tatsächlich gelöscht werden. Dementsprechend müssen die Nutzer den entsprechenden Stellen vertrauen, dass eine fällige Löschung auch tatsächlich umgesetzt wird. Darüber hinaus gelten die Vorgaben des Messstellenbetriebsgesetzes zunächst nur für intelligente Energiezähler und nicht zwangsläufig auch für andere Quellen von Energiedaten wie beispielsweise intelligente Zwischensteckdosen. Offensichtlich gilt hier das Prinzip, dass Daten die Kontrolle der Nutzer nach Möglichkeit gar nicht erst verlassen dürfen. Sobald dies geschieht, kann eine Weiterverteilung ebenso wenig unterbunden werden, wie eine Löschung durchgesetzt werden kann. Werden keine Technologien wie *Usage Control* (vgl. Abschnitt 3.2.2.7) eingesetzt, muss sich ein Nutzer ausschließlich auf regulatorische Möglichkeiten verlassen. Dieser Aspekt und die daraus resultierenden Anforderungen werden im kommenden Abschnitt 4.2 ausführlich erläutert.

4.1.7.2. Löschung lokaler Daten

Wie in der Untersuchung der dritten Phase gezeigt, fallen bei der Speicherung von Energiedaten keine extrem großen Datenmengen an (vgl. Abschnitt 4.1.4.1). Der verfügbare Speicherplatz dürfte somit keinen zwingenden Anlass zur

Lösung von Daten liefern. Es ist jedoch möglich, dass die gespeicherten Daten für den Nutzer nicht mehr von Relevanz sind, weshalb eine Lösung in regelmäßigen Abständen in Erwägung gezogen werden sollte. Sobald die Daten für den Nutzer nicht mehr von Wert sind, sollten sie gelöscht werden. Daten, welche gelöscht wurden, können auch nicht mehr von unbefugten Dritten missbraucht werden, so dass eine regelmäßige Lösung den Datenschutz und die Privatsphäre verbessern kann. Dementsprechend sollte ein Energiedatenmanagementsystem Regeln zur Lösung der lokal gespeicherten Daten vorhalten. Eine optionale Archivierung stellt eine Möglichkeit dar, die Daten außerhalb des Energiedatenmanagementsystems vorzuhalten, sollte der Nutzer in Zukunft doch noch auf die Daten zugreifen wollen.

4.1.7.3. Resultierende Anforderungen

Da ein Nutzer für bereits weitergebene Daten eine Lösung kaum⁸ durchsetzen kann, sollte die Menge der verteilten Daten so klein wie möglich gehalten werden. Diese Anforderung ist vor allem nutzergetrieben und wird deshalb im folgenden Abschnitt 4.2 ausführlich ausgearbeitet. Für lokale Daten ist eine regelmäßige Prüfung und gegebenenfalls Lösung hingegen sinnvoll. Ein Energiedatenmanagementsystem sollte dem Nutzer daher in regelmäßigen Abständen eine Lösung der Daten vorschlagen und – insofern der Nutzer einwilligt – diese Daten auch tatsächlich löschen. Die Anforderung ist in Anforderungsübersicht 10 zusammengefasst.

4.1.8. Fazit zur technischen Datenlebenszyklusanalyse

Mit der Datenlebenszyklusanalyse werden vor allem die technischen Aspekte des Umgangs mit Energiedaten untersucht. Aus der Analyse der sechs Phasen ergeben sich Besonderheiten von Energiedaten und technische Anforderungen, die ein Energiedatenmanagementsystem erfüllen muss. Energiedaten zeichnen sich oft in besonderem Maße durch ihre Heterogenität aus. Sie stammen aus unterschiedlichsten Quellen, liegen in verschiedenen Formaten vor und werden mit diversen Protokollen übertragen. Dies gilt sowohl für die in ein

⁸Die Verwendung von *Usage Control* (vgl. Abschnitt 3.2.2.7) böte gegebenenfalls diese Möglichkeit, ist jedoch nicht immer praktikabel.

Anforderung 10: *Lösung nicht mehr benötigter Daten*

- Vorgabe sinnvoller Löschregeln (bspw. in Abhängigkeit vom Alter der Daten)
 - Regelmäßige Prüfung, ob Nutzer Daten löschen möchte
 - Lösung nur auf ausdrücklichen Wunsch
 - Optionale Archivierung in ausgelagertem System
-

System eingehenden Daten als auch für die ausgehenden. Dementsprechend sollte ein Energiedatenmanagementsystem flexibel auf die verschiedenen Arten von Daten reagieren können. Diese und weitere technische Anforderungen sind in den Anforderungsübersichten 1 bis 14 zusammengefasst.

Es gilt jedoch, nicht nur die technischen Herausforderungen zu beachten. Energiedaten weisen in der Regel einen starken Personenbezug auf und können die Privatsphäre der betroffenen Personen gefährden. Dementsprechend sollten auch die Interessen der Betroffenen bei der Ausarbeitung von Anforderungen in Betracht gezogen werden. Im folgenden Abschnitt werden die Auswirkungen von Energiedaten auf die Privatsphäre detailliert untersucht und die sich ergebenden Anforderungen ausgearbeitet.

4.2. Privatsphäre und Energiedaten

Energiedaten sind als personenbezogene Daten ein schützenswertes Gut. Eine Analyse der Energiedaten mit entsprechenden Methoden ermöglicht umfangreiche Rückschlüsse auf die betroffenen Personen, wodurch eine Gefahr für die Privatsphäre entsteht. Im Folgenden werden solche Methoden untersucht, die es ermöglichen, Informationen aus Energiedaten zu extrahieren. Die Untersuchungen haben das Ziel, zu erkennen, welche Informationen auf welche Art extrahiert werden können und wie die Nutzer davor geschützt werden können. Aus diesen Erkenntnissen werden nutzerorientierte Anforderungen an ein Energiedatenmanagementsystem abgeleitet.

4.2.1. Begriffsabgrenzung

Verfahren zur Extraktion von Informationen aus Energiedaten werden unter verschiedenen, verwandten Begriffen zusammengefasst. In der Literatur findet vor allem der Begriff *nonintrusive load monitoring* Anwendung, wobei *nonintrusive appliance load monitoring* ebenfalls verwendet wird. Die entsprechenden Abkürzungen *NILM*, *NALM* beziehungsweise *NIALM* können synonym verwendet werden [Ber+11; Har92; NL96; Par+12; ZR11]. Der *Monitoring*-Aspekt bezieht sich in diesem Kontext ausdrücklich nicht nur auf eine reine Beobachtung der Daten, sondern bedeutet auch, dass die Daten weitergehend analysiert und verarbeitet werden. Im Folgenden wird daher anstelle von *Energiemonitoring* der sehr allgemein gehaltene Ausdruck der Extraktion von Informationen genutzt, um damit die diversen Methoden abzudecken – unabhängig davon, in welcher Form konkret Informationen extrahiert werden. Gemeint ist damit folglich die Gesamtheit aller Verfahren zur Gewinnung von Informationen aus erfassten Energiedaten.

4.2.2. Verfahren zur Extraktion von Informationen aus Energiedaten

Bereits in den 1980er-Jahren hat Hart als Pionier auf seinem Gebiet an Verfahren zur Erfassung und Zerlegung von Lastgängen gearbeitet, mit deren Hilfe er weitere Informationen über die in Haushalten verwendeten Geräte und das Verhalten der Bewohner erlangen wollte [Har89]. Die zugrundeliegende Idee ist es, den Stromverbrauch nicht nur als Energietransport, sondern auch als eine Art von Informationsübertragung anzusehen. Die übertragenen Informationen müssen demzufolge nur auf die richtige Art und Weise interpretiert werden, um Rückschlüsse über Verhalten und verwendete Geräte ziehen zu können [Har92]. Dieser Ansatz ist der Grundstein für eine Vielzahl verschiedener Abwandlungen und Spezialisierungen, welche in ihrer Summe mittlerweile ein gesamtes Forschungsfeld bilden. Die Menge an Veröffentlichungen ist zu groß, als dass an dieser Stelle eine erschöpfende Analyse möglich wäre. Daher wird im Folgenden auf Basis des Übersichtsartikels von Zeifmann und Roth [ZR11] aufgezeigt, wie sich die Verfahren unterscheiden

und welches Potential sie aufweisen. Zuvor wird jedoch erläutert, welche Rolle die Messfrequenz bei den vorgestellten Verfahren spielt.

4.2.2.1. Die Rolle der Messfrequenz

Verfahren zur Extraktion von Informationen aus Energiedaten können im Hinblick auf die Messfrequenz der erfassten Daten in zwei Lager eingeteilt werden. Auf der einen Seite die Methoden, welche mit einer relativ niedrigen zeitlichen Auflösung im Bereich von Sekunden bis Minuten arbeiten. Auf der anderen Seite finden Verfahren mit hoher Messfrequenz in der Größenordnung bis zu einigen Kilohertz Anwendung [EE15; ZR11].

Verfahren, welche mit geringeren Frequenzen um 1 Hz arbeiten, können gemäß Nyquist-Shannon-Abtasttheorem bei weitem nicht alle Informationen erfassen, die in einem sinusförmigen Spannungsverlauf mit einer Frequenz von 50 Hz enthalten sind. Noch weniger ist bei einer so geringen Abtastrate erkennbar, ob und falls ja, welche Oberwellen vorhanden sind [vgl. u.a. Maa+12]. Dementsprechend können bei dieser Auflösung nur gröbere Merkmale der Stromnutzung betrachtet werden. Dabei handelt es sich beispielsweise um den Verlauf von Wirk- und Blindleistung mit einer Abtastrate im Sekundenbereich. Für eine solche zeitlich geringer aufgelöste Erfassung der Daten sprechen einige Vorteile. So sind nicht nur die Datenraten kleiner und damit die Anforderungen an Übertragung und Speicherung geringer. Darüber hinaus sind insbesondere auch die Geräte zur Erfassung der Daten günstiger [VKD13].

Werden die Daten hingegen mit einer hohen zeitlichen Auflösung im Bereich von einigen Kilohertz erfasst, sind weitaus mehr Informationen zugänglich. Es ist nicht mehr nur der grobgranulare Verlauf von Wirk- und Blindleistung erkennbar, sondern beispielsweise auch Abweichungen vom idealen sinusförmigen Verlauf der Wechselspannung in Form von Oberwellen. Werden diese zusätzlichen Informationen in die Analyse mit einbezogen, können weitere Charakteristika des Stromverbrauchs identifiziert werden, die bei der Geräteerkennung hilfreich sein können. Die Erhöhung der Messfrequenz liefert somit nicht einfach nur feinere Daten, sondern macht damit auch Effekte sicht- und damit nutzbar, die andernfalls verborgen blieben. Diese Vorteile bringen jedoch den Nachteil mit sich, dass die Datenraten im Vergleich um Größenordnungen höher sind als bei Verfahren mit niedriger Messfrequenz.

Die Anforderungen an die Analysesysteme steigen im gleichen Maße. Des Weiteren sind Messgeräte, welche in der Lage sind, hochauflöste Daten zu erfassen, tendenziell deutlich teurer als solche, welche nur mit niedrigen Frequenzen arbeiten [Kim+11].

Es ist offensichtlich, dass die Messfrequenz eine signifikante Auswirkung auf die Erkennbarkeit von Geräten haben kann. Anhand unterschiedlicher Gerätetypen kann dieser Umstand dargelegt werden. Wie von Hart *et al.* ursprünglich vorgeschlagen und von Zeifman und Roth aufgegriffen, können elektrische Haushaltsgeräte hinsichtlich ihrer Verbrauchsmuster in vier Kategorien eingeteilt werden [ZR11].

Dauerverbraucher (i) Die verbrauchte Wirk- und Blindleistung dieser Geräte ist dauerhaft annähernd konstant. Darüber hinaus ist diese Art von Geräten durchgehend in Betrieb und wird nicht ausgeschaltet. Permanente Beleuchtung, mit Netzspannung versorgte Rauchmelder, dauerhaft laufende Heizungspumpen und ähnliche Geräte gehören zu dieser Kategorie.

An/aus-Geräte (ii) Diese Geräte haben lediglich zwei Betriebszustände. Entweder sind sie ein- oder ausgeschaltet. Typische Vertreter sind (nicht dimmbare) Leuchten, aber auch einstufige (Heiz-)Lüfter sowie Wasserkocher und Toaster.

Geräte mit mehreren Zuständen (iii) Geräte, welche dieser Gruppe angehören, durchlaufen in ihrem Betrieb eine Reihe von Zuständen in einer definierten Reihenfolge. Angehörige dieser Kategorie sind unter anderem Waschmaschinen, Spülmaschinen und Trockner. Auch Kühlgeräte gehören zu dieser Kategorie.

Kontinuierlich variable Verbraucher (iv) Diese Geräte weisen keinen konstanten Verbrauch auf und befinden sich nicht in definierten Zuständen. Vielmehr kann ihr Verbrauch kontinuierlich variieren. Beispiele aus dieser Gruppe sind dimmbare Leuchten, aber auch diverse elektrische Werkzeuge wie Bohrmaschinen oder stufenlos einstellbare Lüfter und Pumpen.

Verfahren, welche lediglich gering aufgelöste Daten zu Wirk- und Blindleistung verwenden, beruhen meist auf irgendeiner Form der Kantenerkennung [EE14]. Daher sind sie für gewöhnlich nicht in der Lage, Geräte der

Kategorie i zu erkennen, da deren Stromverbrauch offensichtlich keine Kanten aufweist. Auch Geräte der Kategorie iv sind so für gewöhnlich nicht identifizierbar, da sie ebenfalls keine oder keine stark ausgeprägten Kanten im Stromverbrauch haben. Die Gerätetypen ii beziehungsweise iii hingegen weisen im Betrieb in der Regel deutliche Sprünge und dadurch auch Kanten auf, die eine Identifikation selbst bei geringeren zeitlichen Auflösungen ermöglichen. Werden hochauflöste Daten hinzugezogen, die weitere Eigenschaften wie Oberwellen oder Einschaltverhalten sichtbar machen können, ist auch eine korrekte Identifikation konstanter (i) beziehungsweise beliebig variabler Geräte (iv) möglich [ZR11].

Dies zeigt, dass bereits die Messgeräte und ihre Fähigkeit, hochauflöste Daten zu erfassen, eine entscheidende Rolle bei der Informationsextraktion spielen. Wenn diese Daten gar nicht erst erfasst werden, können nach heutigem Stand der Technik gewisse Geräte auch nicht erkannt werden. Im Umkehrschluss bedeutet dies aber auch, dass – falls solche Daten erfasst werden – detaillierte Profile erzeugt werden können. Aus Nutzersicht sollten solche Daten nicht oder nur aus guten Gründen weitergegeben werden, da sie sehr sensible Informationen enthalten können. Ein Energiedatenmanagementsystem sollte folglich Daten nicht oder nur entsprechend der Präferenzen der Nutzer weitergeben.

4.2.2.2. Überblick über verschiedene Verfahren zu niedrig aufgelösten Daten

Im Folgenden werden verschiedene Verfahren vorgestellt, mit deren Hilfe Informationen aus Energiedaten mit niedriger Auflösung – beispielsweise im Bereich von einigen Minuten – extrahiert werden können.

Pionierarbeiten Die klassischen Verfahren im Sinne von Hart *et al.* basieren auf einer einfachen Kantenerkennung und einem anschließenden Clustering. Zunächst werden dazu im Verlauf des Stromverbrauchs mehr oder weniger abrupte Sprünge gesucht. Diese Sprünge werden als Einschalten eines Gerätes interpretiert, woraufhin dann dessen – nun neu hinzugekommener – Stromverbrauch erfasst wird. Dazu werden beim Verfahren von Hart *et al.* sowohl Wirk- als auch Blindleistung gemessen. Die erhaltenen Werte werden

in einem zweidimensionalen Raum einer Clusteranalyse unterzogen, um sie Geräten zuordnen zu können [Har89; Har92]. Diese ersten Verfahren waren bereits in der Lage, einige ein- und ausschaltbare Geräte (Kategorie ii) zuverlässig zu erkennen. Werden allerdings gleichzeitig viele Geräte genutzt oder unterschiedliche Geräte mit ähnlichem Stromverbrauch verwendet, wie es beispielsweise bei Schaltnetzteilen der Fall ist, versagen sie. Darüber hinaus sind die Verfahren empfindlich gegenüber plötzlich auftretenden, kurzen Spitzen. Nachteilig ist weiterhin, dass in einem neuen Umfeld ein erneutes Training notwendig ist, um Geräte zuverlässig erkennen zu können [ZR11].

Verbesserung durch Training Weiss *et al.* [Wei+12a] wenden ein ähnliches Verfahren an, beziehen aber die Bewohner eines Hauses in das Training des Systems mit ein. Die Nutzer werden per Smartphone dazu aufgefordert, Geräte zu verwenden beziehungsweise an- und auszuschalten. Das Messsystem kann die dadurch ausgelösten Flanken und Profile erkennen und in einer Datenbank ablegen. Diese auf die jeweiligen Haushalte zugeschnittenen Profile erlauben Erkennungsraten um 90 %. Während zum gegenwärtigen Zeitpunkt in neuen Haushalten erneut eine Trainingsphase nötig ist, ist es durchaus denkbar, auf diese zu verzichten, wenn die Datenbank einen ausreichenden Umfang an Profilen hat. Auch Ruzzelli *et al.* [Ruz+10] beziehen den Nutzer in das initiale Training des Systems mit ein. Die gelernten Profile werden an ein externes System übertragen, so dass mit wachsender Datenbankgröße der nötige Trainingsumfang reduziert werden kann.

Markov-Modelle Diverse Arbeiten [vgl. u.a. Par+12; KJ12; Kim+11] haben das Ziel, den Trainingsaufwand am Einsatzort zu vermeiden und trotzdem ausreichend genau zu arbeiten. Die Ansätze unterscheiden sich signifikant von den klassischen Verfahren. Im Vorfeld werden dazu Verbrauchsmodelle auf Basis sogenannter *Hidden Markov Models* entwickelt. Ein *Hidden Markov Model* ist – vereinfacht dargestellt – ein stochastisches Modell eines Systems, in welchem die eigentlichen Zustände nach außen verborgen sind. Erkannt werden können lediglich die Auswirkungen der Zustände, welche dann über Wahrscheinlichkeitsverteilungen den möglichen Zuständen zugeordnet werden müssen. Diese Art von Modellen wird bereits seit längerem erfolgreich in

der Spracherkennung eingesetzt. In den letzten Jahren erfolgte die Übertragung des Prinzips auf Energiedaten. Bereits jetzt handelt es sich bei dieser Art von Verfahren um sehr vielversprechende Kandidaten für eine effiziente und erfolgreiche Extraktion von Informationen aus Energiedaten [KJ12]. Die Verfahren erreichen mit *Hidden Markov Models* meist bessere Erkennungsraten als andere klassische Verfahren. Diese hohen Erkennungsraten sind auch möglich, wenn lediglich niedrig aufgelöste Energiedaten verwendet werden [Kim+11]. Auch der Einsatz unter realistischen Bedingungen wurde bereits erprobt. Parson *et al.* [Par+12] haben ihr Verfahren nicht nur an einem Standarddatensatz (vgl. Abschnitt 3.1.1.1 bzw. [KJ11]) getestet, sondern auch seine Einsetzbarkeit in einem Realweltzenario demonstriert. Das verwendete Modell wurde lediglich mit aggregierten Daten über den Gesamtverbrauch trainiert. Trotzdem konnten gute Erkennungsraten erreicht werden, welche vergleichbar sind mit der Qualität von Modellen, die mit detaillierten Einzellastgangsdaten trainiert wurden. Trotz der spärlichen Datenbasis konnte eine Vielzahl von Vorgängen im Realhaushalt korrekt erkannt werden.

Künstliche neuronale Netze Prudenzi [Pru02] wählt mit künstlichen neuronalen Netzen einen weiteren Ansatz. Bei künstlichen neuronalen Netzen handelt es sich um Modelle, welche die Funktionsweise von Nervenzellverbindungen biologischer Gehirne nachahmen. Verwendet werden hier Daten mit einer geringen zeitlichen Auflösung von 15 Minuten. Die Daten müssen aufgrund des verwendeten Algorithmus bereits blockweise vorliegen, für einen Echtzeiteinsatz mit kontinuierlich erfassten Werten ist dieses Verfahren somit nicht geeignet. Die gespeicherten Daten werden in [Pru02] in Blöcke zu je acht Stunden aufgeteilt und jeweils in ein künstliches neuronales Netz gespeist, welches daraufhin die antrainierten Muster erkennen soll. Im Rahmen der Veröffentlichung wurde die Untersuchung auf Waschmaschinen, Geschirrspülmaschinen und die elektrische Warmwasserbereitung beschränkt. Mit den angegebenen Einschränkungen werden etwa 90 % der Geräteprofile korrekt erkannt. Offensichtlich gibt es noch Potential für Verbesserungen, da nur eine Auswahl typischer Haushaltgeräte erkannt werden kann. Dennoch zeigt die Arbeit eindrücklich, dass selbst mit vergleichsweise wenigen Messpunkten – 32 je Block von acht Stunden beziehungsweise alle 15 Minuten – Aussagen

über die Gerätenutzung in einem Haushalt möglich sind. Insbesondere die aufgrund ihrer charakteristischen Profile gut zu erkennenden Haushaltsgeräte wie Waschmaschinen und Geschirrspülmaschinen erlauben ihrerseits wiederum Rückschlüsse auf das Verhalten von Hausbewohnern. Somit können – je nach angestrebtem Zweck der Analysen – bereits zeitlich gering aufgelöste Daten umfangreiche Informationen liefern.

Genetische Algorithmen Baranski *et al.* [BV04] nutzen mit genetischen Algorithmen wiederum ein anderes Verfahren. Die Daten entstammen dabei einem einfachen Messgerät, welches mit einer niedrigen Frequenz den Stromverbrauch am Haushaltsanschlusspunkt erfasst. Dem Verfahren liegt die Idee zugrunde, dass sich der Stromverbrauch durch einen endlichen Automaten darstellen lässt. Mit einer Kombination diverser Clusteringverfahren und genetischer Algorithmen versucht das Verfahren, einen endlichen Automaten zu finden, welcher das gemessene Lastprofil erzeugt. Die Zustände des endlichen Automaten geben wiederum die vermuteten Zustände der modellierten Haushaltsgeräte wieder. Von Nachteil sind einerseits der hohe Rechenaufwand und andererseits die vergleichsweise schlechten erzielten Ergebnisse.

4.2.2.3. Überblick über verschiedene Verfahren zu höher aufgelösten Daten

Die zuvor beschriebenen Verfahren nutzen Energiedaten mit niedriger zeitlicher Auflösung. Darüber hinaus gibt es Daten mit einer hohen zeitlichen Auflösung im Kilonahertzbereich (vgl. auch Abschnitt 4.2.2.1). Inspiriert von den Ideen von Hart *et al.* [Har92] wurden innerhalb kurzer Zeit auch Daten mit einer hohen Messfrequenz untersucht, wobei zum Teil enorme Fortschritte erzielt werden konnten [NL96].

Erste Ansätze für Energiedaten mit hoher Messfrequenz Exemplarisch für Analysen von Daten mit hoher Messfrequenz sei die Veröffentlichung von Laughman *et al.* [Lau+03] genannt. Sie beschreibt ein Messsystem, welches die elektrische Leistung mit einer Frequenz von rund 8 kHz erfasst. Mit dieser Auflösung versuchte man, folgendes Problem zu umgehen. Werden lediglich die durchschnittliche Wirk- und Blindleistung verschiedener Geräte

betrachtet, so unterscheiden sich diese zwischen verwandten Gerätetypen häufig kaum. Eine Clusteranalyse wie bei den klassischen Verfahren ist somit nicht möglich. Allerdings kann mit der hohen nutzbaren zeitlichen Auflösung auch der Verlauf der Leistungsaufnahme detailliert betrachtet werden. Dies erlaubt die Betrachtung des nicht perfekt sinusförmigen Verlaufs des Stromflusses bei Wechselstrom und damit eine Charakterisierung der verschiedenen Geräte. Insbesondere die Analysen von Oberwellen und der Einschaltvorgänge zeigen Unterschiede zwischen den Geräten auf. Somit lassen sich selbst bei einem scheinbar konstanten durchschnittlichen Stromverbrauch Geräte identifizieren.

Künstliche neuronale Netze In jüngerer Vergangenheit wurde die Idee der Verwendung neuronaler Netze erneut aufgenommen und weiter verbessert. So nutzen Srinivasan *et al.* [SNL06] künstliche neuronale Netze, um hochaufgelöste Energiedaten zu analysieren. Sie konzentrieren sich dabei auf Geräte, welche mit niedrig aufgelösten Daten kaum identifizierbar wären: Nicht zuletzt aufgrund ihres geringen Stromverbrauchs und einem annähernd flachen Profil ohne ausgeprägte Kanten sind Verbraucher wie Computer, Monitore, Lampen und ähnliche nur schwer mit klassischen Methoden zu identifizieren. Die charakteristischen Oberwellen ermöglichen laut den Arbeiten von Srinivasan *et al.* jedoch sehr gute Erkennungsraten von zum Teil annähernd 100 %. Selbst wenn die Messwerte von Rauschen überlagert sind, wird noch von durchschnittlichen Erkennungsraten von 70 % berichtet.

Hochaufgelöste Daten zu elektromagnetischer Interferenz Gupta *et al.* [GRP10] erreichen bei einigen Gerätetypen ebenfalls hervorragende Erkennungsraten von über 90 %. Zum Einsatz kommt ein *ElectriSense* genanntes Messgerät, welches die von vielen Geräten ausgehende elektromagnetische Interferenz messen kann. Die Analog-Digital-Wandlung erfolgt mit einer Abtastrate von 1 MHz. Durch eine softwaregestützte Spektralanalyse wird anschließend die Interferenz bestimmt. Diese geht in der Regel von Schaltnetzteilen aus, welche den Stromkreis, an dem sie angeschlossen sind, mit elektromagnetischer Strahlung „verschmutzen“. Im Unterschied zu vielen anderen Verfahren muss das Messgerät nicht am Hausanschlusspunkt zum

Einsatz kommen, sondern kann an jeder Steckdose im Haus verwendet werden. Es wird schließlich nicht der Stromverbrauch, sondern lediglich die Interferenz gemessen. Dadurch, dass das Verfahren nicht auf das Erfassen von Einschaltvorgängen angewiesen ist, sind auch dauerhaft eingeschaltete Geräte gut zu erkennen. Da der Profilverlauf ebenfalls keine Rolle spielt, können auch Geräte erkannt werden, die mit anderen Verfahren nicht identifizierbar sind, da sie beispielsweise keine Kanten im Verbrauchsprofil verursachen. Im Umkehrschluss sind aber Geräte, welche keine elektromagnetischen Interferenzen verursachen, nicht erkennbar. Dadurch sehen die Autoren dieses Verfahren auch nicht zwangsläufig als Konkurrenz zu existierenden Methoden, sondern vielmehr als Erweiterung. Da aber ein steigender Anteil aktueller elektrischer Verbraucher über Schaltnetzteile verfügt, ist zu vermuten, dass zukünftig mehr Geräte über ihre elektromagnetische Strahlung identifizierbar sein werden.

4.2.2.4. Weitere Verfahren

Die für eine hohe zeitliche Auflösung nötigen Messgeräte sind in der Regel teuer. Auch sind in vielen Haushalten bereits intelligente Stromzähler vorhanden. Dementsprechend ist ein breiter Einsatz zusätzlicher, teurer Messgeräte eher unwahrscheinlich. Aus diesem Grund setzen Liao *et al.* [LSS14] auf niedrig aufgelöste Daten, welche mit weiteren Informationen von sehr günstigen Umgebungssensoren angereichert werden. Die erfasste Luftfeuchtigkeit und Raumtemperatur in Kombination mit den gemessenen Energiedaten erlauben umfangreiche Rückschlüsse auf das Verhalten der beobachteten Hausbewohner. In über 80 % der Fälle wird korrekt erkannt, ob die betroffenen Personen Essen bereiten, sich ausruhen oder das Bad benutzen. Denkbar ist der Einsatz dieser Technologie beispielsweise bei der Unterstützung pflegebedürftiger Personen. Diese Arbeit zeigt eindrucksvoll, dass zusätzliche, einfache und günstige Sensoren eine deutliche Verbesserung in der Erkennung von Nutzerverhalten ermöglichen.

Marchiori *et al.* [Mar+11] erkennen ebenfalls, dass Messungen ausschließlich am Hausanschlusspunkt nur eine begrenzte Genauigkeit liefern können. Da die Kosten hochauflösender Messgeräte einen breiten Einsatz verhindern und die flächendeckende Ausstattung mit intelligenten Zwischensteckdosen

ebenfalls zu teuer ist, setzen Marchiori *et al.* auf einen Kompromiss. An jeder Sicherung und somit in jedem Stromkreis eines Haushalts wird jeweils ein Messgerät installiert, welches mit 1 Hz Blind- und Wirkleistung erfasst. Diese Installation ist zwar teurer als eine einzelne Messstelle am Hausanschlusspunkt, jedoch günstiger als das Ausstatten aller Geräte mit Messeinheiten oder die Verwendung hochauflösender Geräte. Durch die feinere räumliche Aufteilung ist ein besseres Erkennen von Geräten möglich, da beispielsweise Großverbraucher wie Waschmaschinen oft einen eigenen Stromkreis besitzen. Darüber hinaus werden die Profile kleiner Verbraucher tendenziell weniger von anderen Geräten überlagert und sind damit besser zu identifizieren.

4.2.2.5. Zwischenfazit

Dieser nicht erschöpfende Überblick zu den unterschiedlichen Arten von Verfahren zeigt, dass in großem Umfang Informationen aus den gemessenen Energiedaten extrahiert werden können. Für verschiedene denkbare Konstellationen existieren jeweils mehrere Ansätze, mithilfe derer Informationen zu Geräten oder Nutzern extrahiert werden können. Nicht nur hochauflöste Daten können zu guten Erkennungsraten führen, sondern auch aggregierte, zeitlich niedrig aufgelöste Daten. Des Weiteren erlaubt insbesondere die Kombination mit weiteren Informationsquellen umfangreiche Rückschlüsse [vgl. u.a. TIL04].

Darüber hinaus ist bemerkenswert, dass ein großer Teil der Veröffentlichungen zum Thema *nonintrusive (appliance) load monitoring* die Vorzüge der Verfahren betont, dabei aber das Missbrauchspotential und die Gefahren für die betroffenen Personen verschweigt [EE15]. Dieser Konflikt kann jedoch in Zukunft an Brisanz gewinnen, wenn zwischen Energiesystemstabilität und Privatsphäre abgewogen werden muss [San+11].

4.2.3. Konkrete Auswirkungen auf die Privatsphäre

Viele Arbeiten zur Extraktion von Informationen aus Energiedaten schweigen sich über die damit verbundenen Gefahren aus. Es gibt jedoch auch zahlreiche Veröffentlichungen zu Energiedaten, die sich vor allem mit den negativen Aspekten der Datenerfassung beschäftigen [vgl. u.a. CPW10; GJL12; MM09;

Khu+10; Mül10]. Die möglichen Auswirkungen auf die Privatsphäre der Betroffenen sind vielfältig und werden zum Teil nur in Ansätzen beschrieben. Es ist aber unstrittig, dass die Erfassung von Energiedaten unter anderem folgende Gefahren für betroffene Personen birgt [The14; Sid+12]:

- Unerlaubte Weitergabe von Energiedaten oder damit verbundenen personenbezogenen Daten an Dritte
- Profilerstellung bezüglich Nutzerverhalten und anschließende nicht erwünschte Nutzung der abgeleiteten Informationen (Marktforschung, Werbung, Versicherungseinstufung, ...)
- Echtzeitüberwachung von Haushalten bei Verfügbarkeit von Echtzeitdaten

Auch wenn diese *abstrakten* Gefahren in der Literatur häufig aufgeführt werden, dürften die konkreten Konsequenzen für die Privatsphäre den meisten Personen nicht bewusst sein [Sid+12]. Im Folgenden werden daher einige *konkrete* Gefahren aufgezeigt, um zu unterstreichen, welche Ausmaße die Analysen annehmen können und dass es sich nicht um eine rein theoretische und abstrakte Bedrohung handelt.

Auch zeitlich niedrig aufgelöste Daten lassen umfangreiche Rückschlüsse über die Personen eines Haushalts und ihr Verhalten zu. Molina-Markham *et al.* [Mol+10] zeigen, dass bereits Daten mit einer Auflösung im Minutenbereich ausreichend sind, um die Privatsphäre signifikant zu verletzen. War eine eigentlich krankgemeldete Person tatsächlich zu Hause? Wie ist die Schlafqualität der Bewohner? Haben die Bewohner rechtzeitig ihren Weg zur Arbeit angetreten? Welche Ernährungsgewohnheiten haben die Bewohner? Diese und weitere Fragen können vergleichsweise einfach beantwortet werden, wenn Daten zum Stromverbrauch am Hausanschlusspunkt vorliegen. Detaillierte Informationen von einzelnen Geräten oder gar hochauflöste Daten sind dafür nicht nötig.

Die Arbeit von Greveler *et al.* [GJL12] zeigt, dass es möglich ist, das Fernsehprogramm zu identifizieren, welches die Bewohner konsumieren. Dazu ist es bei einigen TV-Geräten ausreichend, den Stromverbrauch des Fernsehers

etwa alle zwei Sekunden zu erfassen und zu analysieren. Je nach Bildhelligkeit schwankt der Stromverbrauch und das entstandene Muster kann entsprechend zugeordnet werden – diese Eigenschaft ist allerdings bauart- und modellabhängig. Die in der Arbeit verwendeten Daten stammen aus einem marktüblichen intelligenten Stromzähler, wie er in einem normalen Haushalt installiert ist.

Quinn nennt in [Qui09] eine Reihe von konkreten Auswirkungen, die ein Missbrauch von Energiedaten auf die Privatsphäre haben kann. Darunter sind Aspekte, die zwar sehr persönlich, aber für Außenstehende vermutlich weniger relevant sind, wie beispielsweise das Duschverhalten. Darüber hinaus gibt es aber einige Aspekte, die von großem Interesse für Externe, wie Versicherungen, sein können. Wie steht es um den Schlaf und damit die Fahrtüchtigkeit der Bewohner? Brechen sie rechtzeitig zu ihren Zielen auf oder lässt der Zeitpunkt der Abfahrt vermuten, dass die Bewohner zu schnell fahren müssen, um ihr Ziel noch rechtzeitig zu erreichen? Werden Haushaltsgeräte vergessen oder genutzt, obwohl niemand im Haus ist? Zusätzlich ist eine Art soziales Profil aus den Energiedaten ablesbar. Wird aufwendig gekocht oder nur die Mikrowelle genutzt? Wie viele Stunden am Tag läuft der Fernseher? Wie oft wird Wäsche gewaschen, wie oft gesaugt?

Auch eine mögliche Religionsangehörigkeit der Bewohner ist anhand der Energiedaten erkennbar [LW08; JKD12]. Werden religiöse Fastenzeiten eingehalten? Wird zu Zeiten des Gottesdienstes zusätzliche Energie verbraucht oder ist nur die Grundlast erkennbar? Wird an bestimmten Religiösen Feiertagen viel oder überhaupt nicht gekocht? Energiedaten können Hinweise darauf geben, ob und wenn ja welcher Religion ein Haushalt angehört. Auch die Einhaltung gewisser religiöser Regeln könnte über Energiedaten geprüft werden.

Seit Jahren werden bereits Stromverbrauchsdaten bei der Verbrechensbekämpfung verwendet. Der dauerhaft hohe Stromverbrauch von Wachstumslampen und der Bewässerung haben bereits mehrmals dazu geführt, dass illegale Drogenplantagen entdeckt wurden [LMW10].

Auch wenn es einige Anwendungsfälle gibt, in denen Nutzer von der Auswertung der Daten profitieren, zeigen die genannten Beispiele, dass Energiedaten ein enormes Missbrauchspotential bergen. Insbesondere große Datenmengen, die massenhafte Auswertungen möglich machen, stellen eine enorme

Verlockung für Dritte dar, die die Daten einfach und schnell zu Geld machen möchten [LMW10]. Selbst wenn die Daten nur für einen bestimmten Grund erhoben werden, besteht stets die Gefahr, dass sie für andere – nicht dem Kundenwunsch entsprechende – Zwecke missbraucht werden [CPW10]. Sind die Daten aber einmal weitergegeben, entziehen sie sich der Kontrolle der betroffenen Personen. Das Prinzip des „Nichts-zu-verbergen-Habens“ greift nicht, da vollkommen unklar ist, ob die Daten nicht doch zu irgendeinem Zeitpunkt böswillig oder mit heute noch unbekannten Verfahren ausgewertet werden [Raj+11; Sta10].

Es ist jedoch nicht nur im Interesse von Privatpersonen, dass ihre Energiedaten geschützt werden. Auch die Akteure der Energiewirtschaft haben ein Interesse daran, dass Energiedaten nicht manipuliert werden können. Dieser Aspekt wird zum Teil auch als Datensicherheit bezeichnet. Falls es potenziellen Angreifern gelänge, nicht nur lesenden, sondern auch schreibenden Zugriff auf die Daten zu erhalten, wären gravierende Angriffe möglich. Manipulierte Energieverbrauchsdaten schlagen sich beispielsweise direkt in der Abrechnung nieder und können damit immense Kosten verursachen. Durch ein massenhaftes Fälschen der Daten könnte auch das gesamte Energienetz instabil werden, da die wahrgenommene Netzsituation von der tatsächlichen stark abweicht. Dementsprechend können intelligente Stromzähler und ihre Daten ein hochattraktives Angriffsziel sein, welches auch im Interesse der Energiewirtschaft geschützt werden muss [vgl. u.a. MM09]. Dieser Aspekt gilt umso mehr, wenn große Datenmengen zentral erfasst und gespeichert werden [SK12].

4.2.4. Resultierende Anforderungen

Aus den durchgeführten Untersuchungen ergeben sich weitere Anforderungen an ein Energiedatenmanagementsystem, welche aus Nutzersicht erfüllt werden müssen. Die Anforderungen werden analog zu den Anforderungen aus der Datenlebenszyklusanalyse in Anforderungsübersichten zusammengefasst.

Energiedaten können mit einer Vielzahl von Methoden ausgewertet werden. Je nach zeitlicher Auflösung und Art von Datenquellen bieten sich unterschiedlichste Verfahren an. Generell gilt jedoch: Je mehr Daten weitergegeben

werden und je höher deren Qualität ist, desto mehr Informationen können potenziell über die betroffenen Personen und ihr Verhalten extrahiert werden. Dementsprechend muss aus Sicht der von der Datenerfassung betroffenen Personen ein Energiedatenmanagementsystem die Weitergabe der Daten massiv einschränken beziehungsweise an die Präferenzen der Betroffenen anpassen. Ein rein rollenbasiertes Zugriffsmanagement wie in [Sah+14] ist nicht ausreichend, da bestimmten Anfragergruppen pauschal Zugriff gewährt würde. Vielmehr muss jede Datenanfrage durch die Nutzer – die Betroffenen – geprüft werden und eine Weitergabe darf ausschließlich nach expliziter Zustimmung erfolgen [vgl. u.a. AA13, S. 37]. Diese Anforderung ist in Anforderungsübersicht 11 zusammengefasst.

Anforderung 11: Einschränkung der Datenweitergabe

- Vielzahl unterschiedlicher Methoden zur Auswertung von Energiedaten erschwert Schutz der Privatsphäre
 - Weitergabe von Energiedaten erhöht die potenzielle Missbrauchsgefahr
 - Keinerlei Datenweitergabe ohne explizite Zustimmung durch die von der Datenerfassung betroffenen Personen (Nutzer)
-

Einerseits erlauben Energiedaten, welche einen größeren Zeitraum überspannen, meist Aussagen bezüglich Verhalten und Gewohnheiten der betroffenen Personen – selbst dann, wenn die zeitliche Auflösung vergleichsweise gering ist. Andererseits erlauben Daten mit hoher zeitlicher Auflösung oft eine Erkennung der verwendeten Haushaltsgeräte, auch wenn der überspannte Zeitraum relativ klein ist. Dementsprechend spielt nicht nur der abgedeckte Zeitraum, sondern auch die zeitliche Auflösung eine große Rolle. Falls ein Nutzer der Weitergabe von Energiedaten zustimmt, sollte diese nur in der geringstmöglichen zeitlichen Auflösung erfolgen. Sie muss darüber hinaus gegebenenfalls mit der anfragenden Partei ausgehandelt werden. Außerdem sollten weitere sinnvolle Möglichkeiten zur Qualitätsreduktion von Energiedaten durch das Energiedatenmanagementsystem angeboten werden [vgl. u.a.

REC13]. Diese Anforderung ist in Anforderungsübersicht 12 zusammengefasst.

Anforderung 12: *Reduktion der Aussagekraft auf ein akzeptables Mindestmaß vor einer Weitergabe*

- Potenzieller Informationsgehalt der Energiedaten steigt mit zeitlicher sowie räumlicher Auflösung
 - Manche Eigenschaften sind nur bei hohen Auflösungen sichtbar
 - Daten nur in möglichst geringer Aussagekraft weitergeben (Reduktion der Auflösung, künstliches Rauschen, ...)
-

Falls Nutzer einer Datenfreigabe zustimmen, sollten sie dies gut informiert und nach einer bewussten Entscheidung tun. Ein Energiedatenmanagementsystem sollte sie dabei unterstützen, zu erkennen, welche Konsequenzen eine Datenfreigabe für die eigene Privatsphäre haben kann. Sobald die Nutzer dazu in der Lage sind, können sie den potenziellen Mehrwert gegen die Risiken einer Datenweitergabe abwägen. Auf diese Weise sind sie in der Lage, eine wohl informierte Entscheidung zu treffen. Diese Anforderung ist in Anforderungsübersicht 13 zusammengefasst.

Anforderung 13: *Aufklärung über mögliche Auswirkungen einer Datenfreigabe auf die Privatsphäre*

- Wert bzw. Informationsgehalt der Energiedaten ist Nutzern meist unklar
 - Auswirkungen auf die Privatsphäre können von Nutzern in der Regel nicht abgeschätzt werden
 - Durch Aufklärung über mögliche Konsequenzen eine bewusste Entscheidung ermöglichen
-

Gegebenenfalls ist es sinnvoll, auch bereits in ihrer Aussagekraft reduzierte Energiedaten zusätzlich zu schützen, bevor eine Weitergabe erfolgt (vgl.

Abschnitt 3.2.2). Beispielsweise könnte ein in der zeitlichen Auflösung reduzierter Datensatz zusätzlich durch die Verwendung von *Usage Control* (vgl. Abschnitt 3.2.2.7 sowie [PHB06; Pre+08]) gesichert werden. Aus Nutzersicht sollte ein Energiedatenmanagementsystem auch solche zusätzlichen Schutzmaßnahmen unterstützen. Diese Anforderung wird in Anforderungsübersicht 14 zusammengefasst.

Anforderung 14: *Unterstützung zusätzlicher, externer Schutzmechanismen*

- Falls eine Datenweitergabe erfolgen soll, kann neben einer Qualitätsreduktion zusätzlicher Schutz sinnvoll sein
 - Flexible Einsatz weiterer Technologien beziehungsweise externer Systeme zum Schutz der Daten ermöglichen (bspw. *Usage Control*)
-

4.3. Weitere Nutzerinteressen

Falls ein Energiedatenmanagementsystem auch langfristig von seinen Nutzern eingesetzt werden soll, muss es entsprechend attraktiv gestaltet werden. Systeme zur Rückmeldung des Energieverbrauchs und insbesondere das Nutzerverhalten diesbezüglich wurden bereits umfangreich untersucht [NR09]. Beispielsweise analysieren Hargreaves *et al.* [HNB10; HNB13] in ihren aktuel-leren Arbeiten das Verhalten von Bewohnern, welche über einen Zeitraum von zwölf Monaten Rückmeldungen von einem dedizierten Gerät zu ihrem Energieverbrauch bekommen. Die Autoren stellen fest, dass nach einer anfänglich intensiven Nutzung die Energierückmeldungen weniger Beachtung erfahren und schließlich im Alltag untergehen. Zwar lässt sich nach Hargreaves *et al.* durch die Rückmeldungen ein verbessertes Verständnis für die Nutzung von Energie erreichen, dies führt jedoch nicht zwangsläufig auch zu einer Verbrauchsreduktion.

Es ist denkbar, dass ein ähnlicher Effekt bezüglich Energiedaten und ihren Auswirkungen auf die Privatsphäre eintreten kann. Für ein Energiedatenmanagementsystem würde dies bedeuten, dass die Nutzer nach einem anfängli-

chen Interesse die Verwendung eines solchen Systems reduzieren oder gar einstellen und damit keine oder – abhängig von der jeweiligen Ausgangssituation – alle Datenanfragen genehmigt werden. Unter welchen Umständen eine solche Situation tatsächlich eintritt und wie sie gegebenenfalls verhindert werden könnte, müsste detailliert untersucht werden. Ohne einen Feldtest, in welchem eine dahingehende Analyse erfolgen kann, sind keine verlässlichen Aussagen möglich. Trotzdem sollte im Allgemeinen versucht werden, die Verwendung eines Energiedatenmanagementsystems so weit wie möglich zu vereinfachen und so komfortabel wie möglich zu gestalten.

Es bietet sich an, bereits in Haushalten beziehungsweise in Unternehmen vorhandene Geräte zu nutzen, um die Benutzeroberfläche des Energiedatenmanagementsystems darzustellen. Dies könnten beispielsweise Notebooks oder Smartphones sein, welche heutzutage vermutlich in den meisten Haushalten zu finden sind. Mobile Endgeräte werden darüber hinaus bereits erfolgreich zur Rückmeldung des Energieverbrauchs eingesetzt [vgl. u.a. Wei+12b].

Sollen möglichst viele Arten von Endgeräten erreicht werden, sind auf die jeweilige Plattform zugeschnittene Anwendungen aufgrund des hohen Entwicklungsaufwands nicht mehr geeignet. Sogenannte Webanwendungen stellen dagegen eine Möglichkeit dar, nicht für jeden potenziellen Endgerätetyp beziehungsweise jedes Betriebssystem eine Applikation pflegen zu müssen [CL11]. Die eigentliche Anwendung wird nicht auf dem anzeigenden Gerät ausgeführt, sondern auf einem den Dienst bereitstellenden Server. Lediglich Teilaufgaben wie beispielsweise die Visualisierung werden durch das jeweilige Endgerät übernommen.

Für Energiedatenmanagementsysteme ergibt sich damit die Anforderung, dass die Nutzung einfach und komfortabel sein sollte. Insbesondere sollte der Kreis potenzieller Endgeräte zur Darstellung der Benutzeroberfläche nicht stark eingeschränkt werden. Diese Anforderung wird in Anforderungsübersicht 15 zusammengefasst.

Vermutlich existieren weitere Anforderungen, die unter Umständen auch stark von der jeweiligen Nutzergruppe abhängen können. Im Rahmen eines Feldtests oder mittels Umfragen könnten diese zukünftig gegebenenfalls erarbeitet werden. An dieser Stelle scheint eine zuverlässige Aussage diesbezüglich jedoch nicht möglich.

4.4. Zusammenfassung der Anforderungsanalyse

Anforderung 15: *Unterstützung verschiedener Endgeräte zur Darstellung der Benutzeroberfläche*

- Nutzung eines Energiedatenmanagementsystems muss einfach und komfortabel sein
 - Zugriff auf die Benutzeroberfläche sollte mit bereits vorhandenen Geräten (Notebook, Smartphone, ...) möglich sein
-

4.4. Zusammenfassung der Anforderungsanalyse

In diesem Kapitel wird eine umfangreiche Analyse der Anforderungen an ein nutzerorientiertes Energiedatenmanagementsystem vorgestellt. Zunächst wird dazu der Datenlebenszyklus von Energiedaten eingehend analysiert und technische Anforderungen erarbeitet. Anschließend werden die Extraktion von Informationen aus Energiedaten und die daraus resultierenden Auswirkungen auf die Privatsphäre betrachtet. Auch hieraus ergeben sich Anforderungen, die ein Energiedatenmanagementsystem erfüllen sollte. Zuletzt fließen noch weitere mögliche Interessen der potenziellen Nutzer in die Untersuchung ein.

Die Gesamtheit der Ergebnisse ergibt einen Anforderungskatalog, der bei der Ausarbeitung eines geeigneten Energiedatenmanagementsystems als Grundlage dient. Die in diesem Kapitel erarbeiteten Anforderungen werden in Tabelle 4.2 zusammengefasst. Der Anforderungskatalog erfasst damit nicht nur die technischen Anforderungen, sondern berücksichtigt insbesondere auch nutzergetriebene Ansprüche an ein Energiedatenmanagementsystem.

Tabelle 4.2.: Aus der Anforderungsanalyse ergibt sich ein Anforderungskatalog, welcher als Grundlage für die Ausarbeitung eines geeigneten Energiedatenmanagementsystems dienen kann. Diese Tabelle gibt eine Übersicht und verweist auf die Seiten, auf welchen die jeweiligen Anforderungen besprochen werden.

Nr.	Anforderungsbezeichnung	Seite
1	Unterstützung unterschiedlicher Datenquellen	79
2	Robustheit gegenüber fehlenden und fehlerhaften Daten	80
3	Lokale Sammelstelle unter Kontrolle des Nutzers	86
4	Transparenz der Datenübertragungswege	86
5	Unterstützung einer geeigneten Speicherform in Abhängigkeit der Datenart	91
6	Verschlüsselung bei externer Speicherung	91
7	Lokale Vorverarbeitung der Daten	95
8	Unterstützung unterschiedlicher Datenausgabeformate	98
9	Sicherheit der Datenübertragungswege	98
10	Lösung nicht mehr benötigter Daten	101
11	Einschränkung der Datenweitergabe	115
12	Reduktion der Aussagekraft auf akzeptables Mindestmaß vor Weitergabe	116
13	Aufklärung über mögliche Auswirkungen einer Datenfreigabe auf die Privatsphäre	116
14	Unterstützung zusätzlicher, externer Schutzmechanismen	117
15	Unterstützung verschiedener Endgeräte zur Darstellung der Benutzeroberfläche	119

Konzept und Umsetzung

Dieses Kapitel widmet sich der Konzeption und Umsetzung eines nutzerorientierten Energiedatenmanagementsystems. Zunächst wird das *Data Custodian Service* genannte Konzept eines solchen Systems vorgestellt. Es sieht eine Reihe von Komponenten vor, welche jeweils spezifische Aufgaben im Hinblick auf die Verwaltung und Handhabung von Energiedaten erfüllen. Anschließend werden die Beziehungen der einzelnen relevanten Entitäten wie beispielsweise Nutzern, Dateninteressenten und Datenanfragen im Rahmen eines Datenmodells beschrieben. Eine erste Version des Konzepts wurde in den Vorarbeiten [RS14; Rig+14] veröffentlicht. Dem konzeptionellen Teil des Kapitels schließt sich die Beschreibung einer möglichen Umsetzung an. In diesem Kontext werden unterschiedliche Werkzeuge und Bibliotheken vorgestellt, welche in dem, in dieser Arbeit entwickelten, Demonstrator verwendet werden, um die vorgesehenen Funktionalitäten zu realisieren. Des Weiteren werden Möglichkeiten zur Bewertung von Datenanfragen sowie zur Reduktion der Datenqualität beziehungsweise zur Verminderung der Aussagekraft der Daten präsentiert. Das Kapitel schließt mit einer Darstellung typischer Arbeitsabläufe und wie diese im Demonstrator umgesetzt werden.

5.1. Architektur des Data Custodian Service

Dieser Abschnitt befasst sich mit dem *Data Custodian Service* genannten Konzept eines Energiedatenmanagementsystems. Im Fokus stehen an dieser Stelle die Struktur des Systems sowie die Aufgaben der jeweiligen Komponenten. Eine mögliche Umsetzung des vorgestellten Ansatzes wird im weiteren Verlauf des Kapitels vorgestellt.

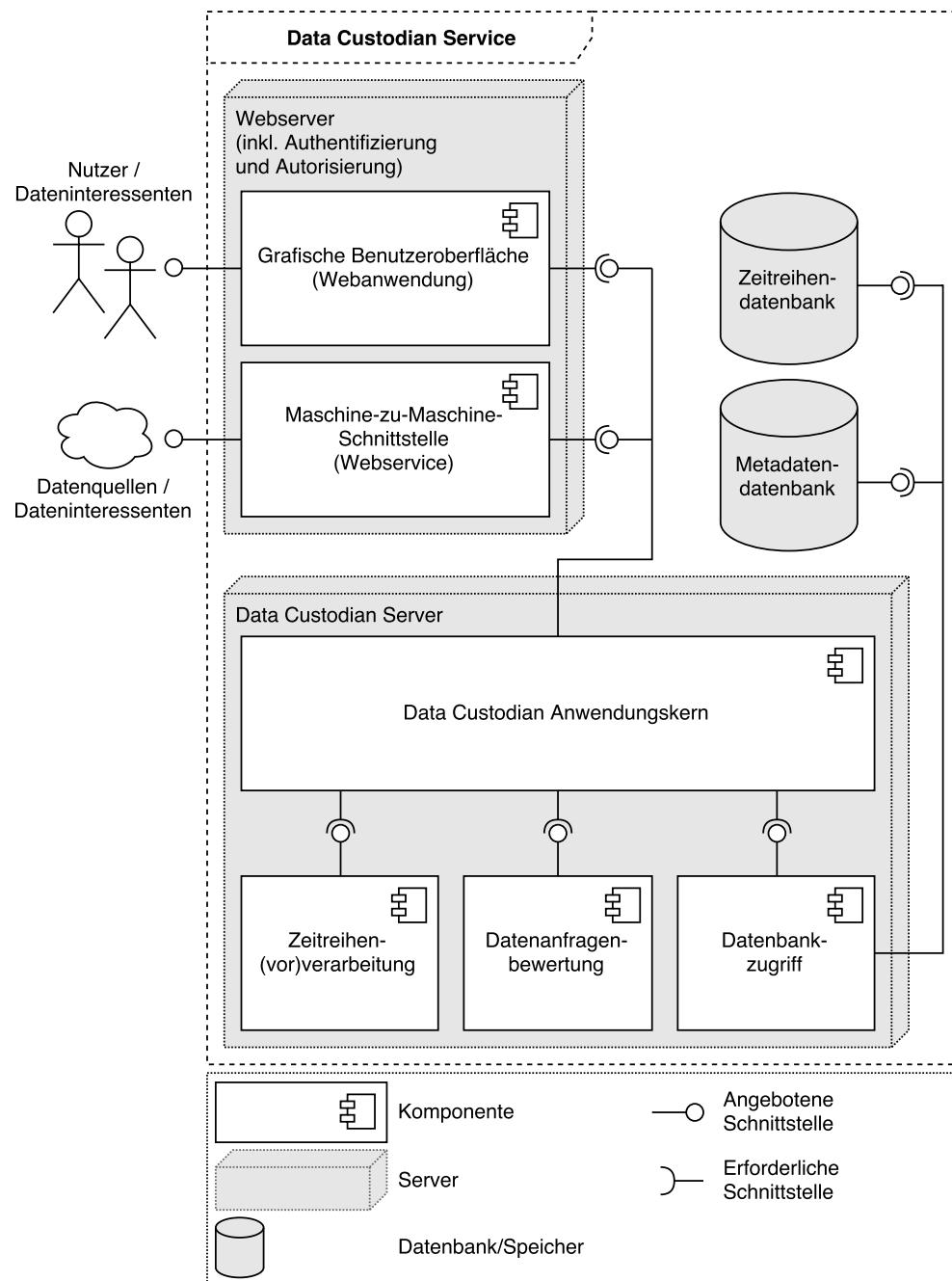


Abbildung 5.1.: Der Data Custodian Service besteht aus mehreren Komponenten, welche jeweils einen bestimmten Aufgabenbereich übernehmen. Diese Abbildung zeigt die Architektur des Energiedatenmanagementsystems in Form der einzelnen Komponenten und deren Zusammenspiel. Die Darstellung ist angelehnt an ein Komponentendiagramm mit sogenannter Lollipop-Notation.

Das *Data Custodian Service* genannte Energiedatenmanagementsystem besteht aus einer Reihe verschiedener Komponenten, welche jeweils eigene Aufgabenbereiche übernehmen. Der modulare Aufbau ermöglicht den einfachen und schnellen Austausch einzelner Teile und damit eine zügige Anpassung an neue oder geänderte Anforderungen. Die Architektur des *Data Custodian Service* ist in Abbildung 5.1 dargestellt. Die Darstellung ist ein UML-Komponentendiagramm, welches die einzelnen Komponenten eines Systems und ihre Zusammenhänge in *Lollipop*-Notation zeigt¹. Von Komponenten angebotene Schnittstellen werden durch Kreise symbolisiert, benötigte Schnittstellen durch Klammern. Die Komponenten selbst sind entsprechend der UML-Spezifikation abgebildet.

Data Custodian Server

Der *Data Custodian Server* stellt das Back-End und somit den Unterbau des *Data Custodian Service* dar. Neben dem *Data Custodian Server* gibt es noch einen Webserver sowie zwei Datenbanken beziehungsweise Datenspeicher, welche jeweils in eigenen Abschnitten erklärt werden.

Der eigentliche *Data Custodian Server* enthält vier Module, welche hauptsächlich für den Umgang mit eingehenden Energiedaten und den dazugehörigen Anfragen verantwortlich sind. Die Koordination aller Vorgänge obliegt dabei dem *Data Custodian Anwendungskern*.

Data Custodian Anwendungskern

Der *Data Custodian Anwendungskern* ist das zentrale Modul des *Data Custodian Service*. Er ist verantwortlich für den Ablauf aller Prozesse im Energiedatenmanagementsystem und steuert das Zusammenspiel der anderen Komponenten. Alle eingehenden Energiedaten gelangen zum *Data Custodian Anwendungskern* und werden von ihm ihrer weiteren Verarbeitung zugeführt. Ebenso wird jede Anfrage von diesem Kernmodul ausgewertet und bearbeitet. Als Koordinator des Systems übt der *Anwendungskern* größtenteils eine steuernde Aufgabe aus.

¹Diese Art der Notation wird zum Teil auch als *Ball-and-Socket* bezeichnet. Weitere Informationen sind beispielsweise unter <http://www.ibm.com/developerworks/rational/library/dec04/bell/> zu finden.

Dementsprechend benötigt er weitere Komponenten, welche jeweils spezifische Aufgaben übernehmen: *Zeitreihenverarbeitung*, *Datenanfragenbewertung* und *Datenbankzugriff*. Außerdem bietet er Schnittstellen für den Anschluss von Front-Ends an.

Zeitreihen(vor)verarbeitung

Energiedaten fallen zwar in Form von Zeitreihen an, jedoch weisen sie zum Teil große Unterschiede auf. Dies betrifft vor allem die verwendeten Dateiformate und die unterschiedlichen physikalischen Größen, welche je nach Energiedatenquelle verwendet werden. Dementsprechend ist beim Einlesen eine Vorverarbeitung der Energiedaten und ein Überführen in ein einheitliches Datenformat nötig. Die Komponente *Zeitreihen(vor)verarbeitung* übernimmt diese Aufgaben.

Für die gezielte Reduktion der Aussagekraft von Energiedaten ist eine Veränderung der Zeitreihen beziehungsweise der darin gespeicherten Daten nötig. Hierbei kann es sich beispielsweise um die Reduktion der zeitlichen Auflösung oder das Hinzufügen künstlichen Rauschens handeln. Auch diese Aufgabe wird durch diese Komponente unterstützt.

Schließlich müssen die verarbeiteten Energiedaten vor einer Weitergabe an Externe noch in das gewünschte AusgabefORMAT übertragen werden. Auch diese Tätigkeit sollte von dem Modul *Zeitreihen(vor)verarbeitung* unterstützt werden.

In Summe sollte innerhalb des *Data Custodian Service* lediglich mit *einem* einheitlichen Format für die Zeitreihendaten gearbeitet werden müssen. Jegliche Manipulation dieser Daten sollte ausschließlich durch die hier beschriebene Komponente geschehen.

Datenanfragenbewertung

Von Dateninteressenten abgegebene Datenanfragen müssen im Hinblick auf ihre Auswirkungen auf die Privatsphäre der Nutzer bewertet werden. Die Komponente *Datenanfragenbewertung* übernimmt diese Aufgabe. Sie erhält vom *Data Custodian Anwendungskern* die zu bewertende Anfrage und beurteilt diese. Bewertungskriterien können unter anderem die gewünschte zeitliche

Auflösung, die betroffenen Geräte und der abgedeckte Zeitraum sein. Das Ergebnis der Auswertung kann von dem *Data Custodian Anwendungskern* verwendet werden, um den betroffenen Nutzern die Auswirkungen auf ihre Privatsphäre zu verdeutlichen.

Datenbankzugriff

Die eigentlichen Zeitreihendaten und die dazugehörigen Metadaten unterscheiden sich grundlegend. Während es sich bei den Zeitreihen um große Datenmengen handeln kann, ist der Speicherbedarf der Metadaten vergleichsweise klein. Die Zeitreihendaten müssen effizient abgespeichert und gelesen werden können, wohingegen bei den Metadaten eine gute Durchsuchbarkeit wichtig ist. Dementsprechend werden die Zeitreihen und die Metadaten getrennt voneinander in spezialisierten Speichern beziehungsweise Datenbanken vorgehalten. Der *Data Custodian Anwendungskern* nutzt für diesen Zweck das *Datenbankzugriff*-Modul. Dieses dient als Schnittstelle und Übersetzer zwischen der Anwendung und der *Zeitreihendatenbank* beziehungsweise *Metadatenbank*. Aus Sicht des Anwendungskerns soll der Datenzugriff möglichst transparent erfolgen.

Zeitreihendatenbank

Die *Zeitreihendatenbank* ist auf die Speicherung großer Mengen von Zeitreihen oder zeitreihenähnlichen Daten spezialisiert. Dabei muss es sich nicht zwangsläufig um eine Datenbank im klassischen Sinne handeln, entsprechend geeignete Binärformate können beispielsweise ebenfalls verwendet werden.

Metadatendatenbank

Die *Metadatendatenbank* dient der Speicherung der zu den Energiedaten gehörenden Metadaten. Dabei handelt es sich im Vergleich zu den Zeitreihendaten nicht um große Datenmengen, wohl aber um zusammenhängende Datensätze. Dementsprechend bietet sich eine relationale oder dokumentenorientierte Datenbank an. Eine effiziente Durchsuchbarkeit der Metadaten muss in jedem Fall gewährleistet sein.

Webserver

Der *Webserver* bietet die Umgebung für die Bereitstellung der Front-Ends des *Data Custodian Service*: die *Grafische Benutzeroberfläche* und die *Maschine-zu-Maschine-Schnittstelle*. Er ist gleichzeitig die Schnittstelle für Interaktionen mit Personen oder Maschinen über die Systemgrenzen hinweg. In diesem Kontext ist der Webserver auch der Kommunikationspartner für die Authentifizierung und Autorisierung der Interaktionspartner.

Grafische Benutzeroberfläche

Sowohl Nutzer als auch Dateninteressenten können über eine Webanwendung auf die *Grafische Benutzeroberfläche* zugreifen. Diese Komponente des *Webservers* erlaubt es Personen, mit dem *Data Custodian Service* über eine Webseite zu interagieren. Die eigentliche Anwendung wird in der Komponente auf dem *Webserver* ausgeführt, eine Visualisierung erfolgt jedoch im Browser der Nutzer beziehungsweise Dateninteressenten. Durch die Verwendung entsprechender Technologien ergibt sich der Vorteil, dass eine große Zahl unterschiedlicher Endgeräte erreicht werden kann, wobei diese lediglich über einen modernen Browser verfügen müssen.

Maschine-zu-Maschine-Schnittstelle

Im Gegensatz zur *Grafischen Benutzeroberfläche* ist die *Maschine-zu-Maschine-Schnittstelle* nicht für die Verwendung durch reale Personen vorgesehen, sondern für die Kommunikation mit anderen Rechnern. Dementsprechend werden definierte Datenstrukturen genutzt, um eine Automatisierung der Vorgänge zu vereinfachen und eine möglichst einfache Weiterverarbeitung der übertragenen Daten zu gewährleisten.

Interaktionspartner

Die zuvor beschriebenen Aspekte sind alle Bestandteile des *Data Custodian Service*. Eine Kommunikation und ein Datenaustausch über die Systemgrenzen hinweg ist lediglich über die Module *Grafische Benutzeroberfläche* und *Maschine-zu-Maschine-Schnittstelle* vorgesehen. Diese bieten – aus Systemsicht – externen

Interaktionspartnern die Möglichkeit, den *Data Custodian Service* zu verwenden. Bei den Interaktionspartnern sind wiederum zwei Gruppen zu unterscheiden: reale Personen und Geräte. Bei realen Personen kann es sich um *Nutzer* oder *Dateninteressenten* handeln. *Nutzer* verwenden den *Data Custodian Service*, um ihre Energiedaten zu verwalten und gegebenenfalls kontrolliert eine Weitergabe mancher Daten zu gestatten. *Dateninteressenten* können über den *Data Custodian Service* Energiedaten der Nutzer anfragen und bei entsprechender Genehmigung auch erhalten. Neben diesen realen Personen können auch andere Rechner mit dem *Data Custodian Service* kommunizieren. *Datenquellen* können nach einer entsprechenden Veranlassung durch den *Nutzer* ihre Daten direkt an den *Data Custodian Service* liefern. Analog muss es sich bei *Dateninteressenten* nicht um reale Personen handeln, da auch Rechner diese Anfragen automatisiert durchführen können.

Zusammenfassung zur Architektur

Das in Abbildung 5.1 dargestellte Konzept für den *Data Custodian Service* besteht aus mehreren Komponenten, welche jeweils einen bestimmten Aufgabenbereich übernehmen. Der *Data Custodian Anwendungskern* nimmt dabei eine koordinierende Rolle ein und steuert das Zusammenspiel der anderen Komponenten. Der modulare Aufbau ermöglicht es, einzelne Komponenten nach Bedarf auszutauschen und durch neue Module zu ersetzen, sodass bei geänderten oder neuen Anforderung eine zügige Anpassung gewährleistet werden kann. Im folgenden Abschnitt wird ein Datenmodell vorgestellt, welches in Kombination mit dem zuvor beschriebenen Architekturkonzept als Grundlage für eine Umsetzung dienen kann.

5.2. Datenmodell des Data Custodian Service

Das dem *Data Custodian Service* zugrundeliegende Datenmodell ist in Abbildung 5.2 dargestellt. Die Grafik zeigt die einzelnen Objekte und deren Beziehung zueinander in Form eines Entity-Relationship-Diagramms. Die Beziehungen sind in der Krähenfußnotation (auch *Martin-Notation* genannt) abgebildet. Bei dieser Notation werden die Objekte (dargestellt als Rechtecke) durch Kanten verbunden und die Kanten an jedem Ende mit zwei der folgen-

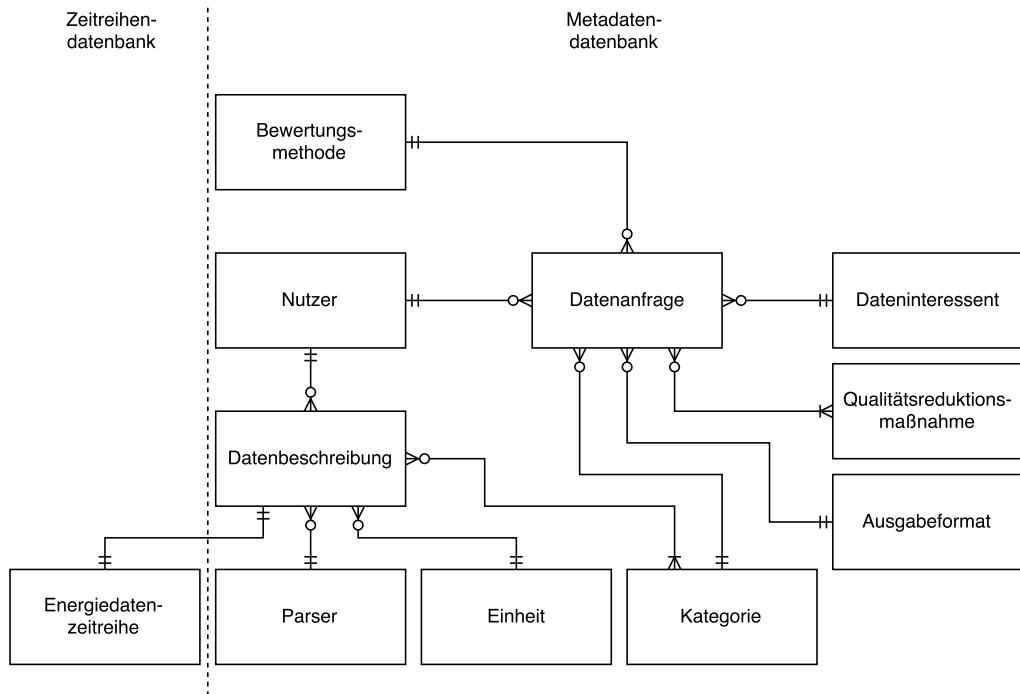


Abbildung 5.2.: Dieses Entity-Relationship-Modell zeigt das dem Data Custodian Service zugrundeliegende Datenmodell. Die Objekte und deren Zusammenhänge sind in der Krähenfußnotation (Martin-Notation) dargestellt. Das Symbol O entspricht null, I entspricht eins und < entspricht beliebig vielen. Das innere Symbol gibt das Minimum an, während das objektnähere Symbol das Maximum anzeigt. Abgelesen werden die Kardinalitäten jeweils am Objekt, zu welchem die Verbindung besteht.

den Symbole versehen: O, I, <. Das innere Symbol gibt jeweils das Minimum, das äußere – dem Objekt nähere – Symbol das Maximum der Kardinalitäten an. O steht für null, I für eins, < für beliebig viele. Abgelesen werden die Kardinalitäten jeweils am Objekt, zu welchem die Verbindung besteht.²

5.2.1. Beschreibung des Datenmodells

Wie in Abbildung 5.2 gezeigt, kann ein *Dateninteressent* im Minimum keine (O) und im Maximum beliebig viele (<) *Datenanfragen* an das System übermittelt

²Zur Verdeutlichung werden bei der ersten im Folgenden beschriebenen Beziehung in Klammern auch die entsprechenden Symbole genannt, bei den weiteren entfallen diese zugunsten einer besseren Lesbarkeit jedoch.

haben. Umgekehrt kann jeder *Datenanfrage* im Minimum ein (I) und im Maximum ebenfalls ein (I) – und damit exakt ein – *Dateninteressent* zugeordnet werden. Dies bedeutet weiterhin, dass eine *Datenanfrage* ohne *Dateninteressent* ebenso wenig existieren kann wie eine *Datenanfrage* mit mehreren *Dateninteressenten*. In der Grafik 5.2 ist dies ersichtlich durch die Verbindung der beiden Rechtecke, welche die Entitäten *Dateninteressent* und *Datenanfrage* darstellen, mit den entsprechenden Symbolen (O und < beziehungsweise I und I), welche wiederum die minimalen und maximalen Kardinalitäten darstellen.

Den *Dateninteressenten* gegenüber stehen die *Nutzer*. Sobald ein Nutzer dem *Data Custodian Service* Energiedaten zur Verfügung stellt, wird eine Instanz zur Beschreibung dieser Daten angelegt: die *Datenbeschreibung*. Da ein neuer *Nutzer* nicht zwangsläufig sofort Daten bereitstellen muss, können zu einem *Nutzer* auch keine *Datenbeschreibungen* existieren – oder (später) beliebig viele. Einer *Datenbeschreibung* kann jedoch immer exakt ein *Nutzer* zugeordnet werden.

Die eigentlichen Zeitreihendaten der Energiedaten werden als einziges Objekt in einer Zeitreihendatenbank gespeichert, alle anderen Objekte werden in der Metadatendatenbank vorgehalten. Damit jede *Energiedatenzeitreihe* eindeutig zugeordnet werden kann, wird ihre Referenz in der *Datenbeschreibung* vorgehalten. Eine *Energiedatenzeitreihe* ohne *Datenbeschreibung* kann ebenso wenig existieren wie eine *Datenbeschreibung* ohne *Energiedatenzeitreihe*.

Beim Einlesen der *Energiedatenzeitreihe* müssen in der Regel Daten aus unterschiedlichen Datenformaten eingelesen – gelesen – werden. Dementsprechend muss beim Bereitstellen der Daten ein *Parser* festgelegt werden, welcher in der Lage ist, die Rohdaten zu interpretieren und aufzubereiten. Zu jeder *Datenbeschreibung* gehört somit auch genau ein *Parser*. Umgekehrt muss aber nicht jedem *Parser* eine *Datenbeschreibung* zugeordnet werden können, da *Parser* existieren können, für deren Quellentyp noch keine Daten erfasst wurden.

Energiedaten können eine Vielzahl verschiedener Einheiten aufweisen. Da in der *Energiedatenzeitreihe* lediglich Zeitstempel und Gleitkommazahlen gespeichert werden, muss die *Einheit* zur Beschreibung der physikalischen Größe getrennt vorgehalten werden. Aus diesem Grund verfügt eine *Datenbeschreibung* stets über eine *Einheit*, welche die physikalische Größe beschreibt. Eine *Einheit* kann beliebig vielen *Datenbeschreibungen* zugeordnet werden – auch

keiner, da es *Einheiten* geben kann, die bisher in keiner Datenbeschreibung verwendet werden.

Zuletzt wird einer *Datenbeschreibung* noch mindestens eine *Kategorie* zugeordnet. Dabei handelt es sich beispielsweise um die Kategorien „Waschmaschine“, „Haushaltsgerät“, „Lastflexibles Gerät“ oder weitere. Insbesondere kann eine *Datenbeschreibung* auch mehrere Kategorien enthalten. Analog zum *Parser* und zur *Einheit* kann eine *Kategorie* zu beliebig vielen *Datenbeschreibungen* gehören – auch zu keiner.

Eine *Datenanfrage* kann stets genau einem *Dateninteressenten* zugeordnet werden, welcher diese dem *Data Custodian Service* übermittelt hat. Die *Datenanfrage* betrifft wiederum immer genau einen *Nutzer*. Der *Dateninteressent* muss in seiner *Datenanfrage* genau eine *Kategorie* benennen, aus welcher die Daten stammen sollen. Die Daten können vom *Dateninteressenten* in einem bestimmten *AusgabefORMAT* angefordert werden. Darüber hinaus muss der *Dateninteressent* eine oder mehrere *Qualitätsreduktionsmaßnahmen* wählen, welche verwendet werden, um die Datenqualität zu mindern. Der betroffene *Nutzer* wiederum muss für eine *Datenanfrage* wählen, mit welcher *Bewertungsmethode* die Auswirkungen auf die Privatsphäre bewerten werden sollen.

Eine *Qualitätsreduktionsmaßnahme*, ein *AusgabefORMAT*, eine *Kategorie* oder eine *Bewertungsmethode* muss nicht zwangsläufig eine korrespondierende *Datenanfrage* haben. Es können jeweils auch Objektinstanzen existieren, die (bisher) in keiner *Datenanfrage* verwendet werden.

Das Datenmodell stellt bewusst ein abstraktes Abbild der Realität dar. Beispielsweise wird an dieser Stelle nicht vorgeschrieben, wie eine *Qualitätsreduktionsmaßnahme* tatsächlich implementiert werden kann oder wie ein *AusgabefORMAT* umgesetzt werden muss. Vielmehr sind die Beziehungen der einzelnen Entitäten im Vordergrund. Konkrete Umsetzungsmöglichkeiten und Ausprägungen der einzelnen Entitäten werden detailliert in Abschnitt 5.3 beschrieben. Darüber hinaus wird in Anhang A gezeigt, wie das entworfene Datenmodell mithilfe der verwendeten Software-Komponenten in der Implementierung des Demonstrators umgesetzt wird.

5.2.2. Anmerkungen zum Datenmodell

Aus dem Datenmodell in Abbildung 5.2 ergibt sich eine Reihe von Besonderheiten im Hinblick auf die Verwaltung der Energiedaten und insbesondere den Umgang mit ihnen im Rahmen von Datenanfragen. Diese Aspekte werden im Folgenden erläutert.

Daten können nicht direkt referenziert werden

Aus Abbildung 5.2 wird ersichtlich, dass die *Datenanfrage* das verbindende Element im Datenmodell ist. Die *Datenanfragen* verbinden die beiden sich gegenüberstehenden Parteien: die *Nutzer* und die *Dateninteressenten*. Hervorzuheben ist dabei insbesondere, dass es *Dateninteressenten* nicht möglich ist, Energiedatensätze direkt zu referenzieren beziehungsweise anzufordern. Vielmehr muss ein *Dateninteressent* eine *Kategorie* und einen *Nutzer* angeben, wodurch potenziell interessante Datensätze identifiziert werden können. Dadurch ist es den *Dateninteressenten* nicht möglich, zu erkennen, welche Daten tatsächlich verfügbar wären, so dass auf diesem Umweg nicht unabsichtlich Informationen preisgegeben werden können. Ein *Dateninteressent* kann dadurch beispielsweise nicht erkennen, ob ein *Nutzer* ein bestimmtes Haushaltsgerät besitzt oder nicht: Beim Erstellen einer *Datenanfrage* können unabhängig von der tatsächlichen Verfügbarkeit der Daten beliebige *Kategorien* ausgewählt werden. Wenn der *Nutzer* die *Datenanfrage* ablehnt, muss er nicht angeben, dass die Daten nicht verfügbar sind. Dem *Dateninteressenten* wird damit nicht offenbart, ob der *Nutzer* solche Daten überhaupt nicht hat oder nur nicht preisgeben möchte.

Datenanfragen betreffen nur genau eine Kategorie

Im Rahmen einer *Datenanfrage* muss ein *Dateninteressent* genau eine *Kategorie* angeben, für welche passende Daten ausgegeben werden sollen. Dies hat hauptsächlich zwei Gründe. Einerseits existieren Kategorien, die andere einschließen, wie beispielsweise die Kategorie *Haushaltsgerät*, welche Trockner und Waschmaschinen gleichermaßen enthält. Passende Datenströme können dann aggregiert und als neue – einzelne – Zeitreihe ausgegeben werden: Die Zeitreihe würde dann die Summe aller betreffenden Haushaltsgeräte um-

fassen, jedoch keinen expliziten Hinweis auf die ursprünglichen Zeitreihen mehr enthalten. Andererseits ist ein *Dateninteressent* damit gezwungen, mehrere *Datenanfragen* abzusetzen, falls detailliertere Daten zu mehreren Geräten gewünscht sind. Diese Anfragen werden dann für den Nutzer individuell bewertet, wobei auch die Anfragehistorie in die Bewertung mit einfließen kann.

Auswahl der Bewertungsmethode

Die für eine *Datenanfrage* einzusetzende *Bewertungsmethode* ist ebenfalls als Entität im Datenmodell vorgesehen. Dem *Nutzer* wird so die Möglichkeit gegeben, aus verschiedenen Verfahren zur Bewertung der Auswirkungen auf die Privatsphäre zu wählen. Es ist denkbar, unterschiedliche Methoden anzubieten, welche sich jeweils beispielsweise an unterschiedlich risikofreudige *Nutzer* wenden. Abhängig von den individuellen Präferenzen können die *Nutzer* so die jeweils passende Methode anwenden lassen. Eine mögliche Umsetzung eines Bewertungsverfahrens wird in Abschnitt 5.3.4 detailliert beschrieben: Eigenschaften einer Datenanfrage wie die gewünschte zeitliche Auflösung, der angefragte Zeitraum, das Alter der Daten sowie die Kategorie, aus welcher die Daten stammen, werden zur Berechnung von Kennzahlen herangezogen, welche wiederum der Abschätzung der potenziellen Auswirkungen auf die Privatsphäre dienen.

Datenanfragen haben mindestens eine Qualitätsreduktionsmaßnahme

Eine *Datenanfrage* muss stets mindestens eine *Qualitätsreduktionsmaßnahme* aufweisen, wobei auch mehrere gleichzeitig ausgewählt werden können. Denkbare Maßnahmen sind hier beispielsweise das Hinzufügen von Rauschen oder die Reduktion der Messwertauflösung. Solche Maßnahmen haben einen Einfluss auf den Informationsgehalt der Energiedaten und damit auch auf die potenzielle Beeinträchtigung der Privatsphäre. Das Datenmodell erlaubt es auch, eine Kombination verschiedener Verfahren zum Einsatz zu bringen, wodurch gegebenenfalls eine signifikante Reduktion der Aussagekraft der Daten möglich ist. Konkrete mögliche Umsetzungen von Maßnahmen zur Reduktion der Datenqualität beziehungsweise der Aussagekraft der Daten werden in

Abschnitt 5.3.5 detailliert vorgestellt. Sollte ein *Dateninteressent* keine Qualitätsreduktion der Daten wünschen, wird im Rahmen der Modellierung die Identitätsfunktion als Maßnahme verwendet. Somit erfolgt zwar einerseits keine Veränderung der Datenqualität, andererseits ist dies aber explizit im Datenmodell ausgedrückt.

Aufteilung in zwei Arten von Datenbanken

In Abbildung 5.2 hebt die unterbrochene Linie die Trennung der Modelle nach Datenbanken hervor. Bis auf die *Energiedatenzeitreihe* werden alle Datenobjekte in einer Metadatendatenbank abgespeichert. Aufgrund der Datenstruktur bietet sich hierfür die Verwendung einer relationalen Datenbank an, in welcher die Verbindungen zwischen den Daten entsprechend abgebildet werden können. Zu beachten ist, dass zwei *many-to-many*-Beziehungen bestehen: *Datenanfrage* und *Qualitätsreduktionsmaßnahme* sowie *Datenbeschreibung* und *Kategorie*. Diese Beziehungen können in relationalen Datenbanken üblicherweise nicht direkt abgebildet werden, sondern benötigen Hilfstabellen, über welche die Zuordnung geschieht. Die beiden Hilfstabellen sind in Abbildung 5.2 aus Gründen der Einfachheit nicht dargestellt, sie sind aber für eine korrekte Zuordnung nötig. Sie befinden sich stets zwischen den Entitäten, welche durch die *many-to-many*-Beziehungen verbunden sind und somit zwischen *Datenanfrage* und *Qualitätsreduktionsmaßnahme* beziehungsweise *Datenbeschreibung* und *Kategorie*.

Die *Energiedatenzeitreihe* könnte ebenfalls in einer relationalen Datenbank gespeichert werden, allerdings ist diese Art der Speicherung mit wachsender Größe der Zeitreihen nicht effizient. Dementsprechend sollte eine auf Zeitreihen zugeschnittene Lösung verwendet werden, wobei es sich nicht um Datenbank im engeren Sinne handeln muss, sondern auch geeignete Binärformate verwendet werden können. Welche konkreten Produkte jeweils zum Einsatz kommen können wird in den Abschnitten 5.3.1 und 5.3.3 erläutert.

Erweiterbarkeit und Austauschbarkeit einzelner Funktionalitäten

Die Entitäten *Parser*, *Einheit*, *Ausgabeformat*, *Qualitätsreduktionsmaßnahme* und *Bewertungsmethode* hätten nicht unbedingt explizit als eigenständige Objekte

modelliert werden müssen. Alternativ wäre es möglich gewesen, diese als einzelnes Attribut oder gegebenenfalls als mehrere Attribute in den korrespondierenden Entitäten vorzusehen. Die explizite Modellierung als eigenständige Objekte hat jedoch den Vorteil, dass eine einfache Erweiterbarkeit und auch Austauschbarkeit gegeben ist. Am Beispiel des *Parser* wird dies deutlich. Soll eine neue Art von Datenquelle hinzugefügt werden, für die noch kein *Parser* existiert, muss nicht der existierende Programmcode erweitert werden. Stattdessen kann ein neues *Parser*-Objekt erstellt werden, welches – wie alle anderen *Parser*-Objekte auch – den entsprechenden, neu erstellen Quellcode beziehungsweise die Klasse referenziert. Somit ist eine zügige Anpassung an neue Anforderungen und veränderte Datenquellen oder auch *Ausgabeformate* möglich. Analog können effizient neue *Qualitätsreduktionsmaßnahmen* und *Bewertungsmethoden* bereitgestellt werden. Diese Aufgaben sollten dem Betreiber des *Data Custodian Service* (beispielsweise einem *Nutzer*) beziehungsweise einem Administrator vorbehalten sein. Da die Rolle des Administrators außerhalb des Systems steht, ist sie in Abbildung 5.1 auch nicht dargestellt. Dennoch müssen der Betrieb und die Wartung des Systems unbedingt von vertrauenswürdigen Parteien oder den *Nutzern* selbst übernommen werden, da über diese Zugriffsmöglichkeit ein Abgreifen oder Manipulieren der Daten möglich wäre.

5.3. Umsetzung des Konzepts in einem Demonstrator

Das in Abschnitt 5.1 beschriebene Konzept und das in Abschnitt 5.2 vorgestellte Datenmodell sind die Grundlage für die im Rahmen dieser Arbeit implementierte *Proof-of-Concept*-Anwendung. Die Umsetzung dient der Demonstration möglicher Funktionsweisen, ist jedoch in diesem Zustand ausdrücklich nicht für den Produktiveinsatz unter Realweltbedingungen vorgesehen.

5.3.1. Zugrundeliegendes Softwaregerüst

Der auf dem Konzept des *Data Custodian Service* basierende Demonstrator wurde in der Programmiersprache *Python* [Pyt16] implementiert. *Python* ist ei-

5.3. Umsetzung des Konzepts in einem Demonstrator

ne höhere, interpretierte Programmiersprache, welche im wissenschaftlichen Bereich weite Verbreitung erfährt. Darüber hinaus existieren zahlreiche Pakete und Module, welche die Funktionalität von *Python* erweitern und verschiedene Anwendungsbereiche abdecken. *Python* ist eine quelloffene Skriptsprache und die meisten Pakete unterliegen einer Lizenz für *freie* Software. Die damit einhergehende Möglichkeit der Überprüfung des Quellcodes ist einer der Wege, das Vertrauen der Nutzer in die verwendete Software zu verstärken [Sta10].

Das von der *Django Software Foundation* veröffentlichte und *Django* genannte *Web Framework*³ [Dja16a] stellt das Grundgerüst für den in dieser Arbeit entwickelten Demonstrator. Es handelt sich um ein Modul für *Python*, welches bei der Erstellung von Webanwendungen verwendet werden kann. Das Framework übernimmt zu diesem Zweck eine Reihe von typischen Aufgaben, die Webanwendungen üblicherweise mit sich bringen. Dies sind unter anderem die Folgenden:

- Eine objektrelationale Abbildung der Datenmodelle inklusive Anbindung an ein Datenbanksystem
- Eine von *Django* bereitgestellte Administrationsschnittstelle für die Datenmodelle
- Ein Vorlagensystem zur Aufbereitung und Darstellung von Webseiten
- Eine Zuordnung von URLs zu Sichten beziehungsweise Funktionen
- Ein Authentifizierungs- und Autorisierungssystem inklusive Nutzerverwaltung
- Diverse Verfahren zum Schutz vor Angriffen wie beispielsweise *Cross-Site-Scripting*

Die folgenden Ausführungen stützen sich auf die ausführliche Dokumentation zu *Django* [Dja16b]. Neben den im Folgenden beschriebenen Aspekten werden dort auch zahlreiche weitere Details zu *Django* und seiner Verwendung erläutert.

³<https://www.djangoproject.com/>

Objektrelationale Abbildung

Django verwendet zur Speicherung der Objekte eine relationale Datenbank. Die einzelnen Objekte werden dazu in Form von *Python-Klassen – Modellen* – definiert: Jede der Klassen weist einzelne Attribute – *Felder* genannt – auf, welche die zu dem jeweiligen Objekt gehörenden Datentypen festlegen. Bei den Datentypen kann es sich beispielsweise um Textfelder festgelegter Länge, Felder zur Speicherung von Zeitstempeln oder auch Fremdschlüsselbeziehungen handeln. Auf diese Weise lässt sich das gesamte Datenmodell aus Abbildung 5.2 inklusive der gezeigten Beziehungen erzeugen. In Anhang A finden sich weitere Details zur konkreten Umsetzung.

Das Datenmodell – aus Sicht von *Django* eine Sammlung von die Objekte beschreibenden Klassen – wird genutzt, um entsprechende Datenbankschemata zu erzeugen und die zugehörigen Tabellen in einer relationalen Datenbank anzulegen. Fortan stehen diverse *Python*-Funktionen für die Nutzung (Erzeugung, Änderung, Speicherung, Löschung, Suche, Filterung, ...) der Daten zur Verfügung. Das Ausführen von *SQL*-Befehlen ist zwar möglich, aber nicht nötig, da annähernd alle denkbaren Anfragen auch in nativem *Python*-Code formuliert werden können. Diese Vorgehensweise hat den Vorteil, dass die Nutzung der Datenbank aus Anwendungssicht vollkommen transparent geschieht und alle Objekte aus der Datenbank als native Objekte zur Verfügung stehen.

Darüber hinaus kann die zugrundeliegende Datenbank ausgetauscht werden, wobei beispielsweise die Produkte *SQLite*⁴, *MySQL*⁵ oder *PostgreSQL*⁶ zum Einsatz kommen können. Im Demonstrator wird *SQLite* als Datenbank verwendet. Diese Variante ist zwar weniger performant als viele andere relationale Datenbanken, dafür aber dateibasiert und serverlos. Bis auf die Zeitreihendaten weisen alle Objekte weder große Datenmengen auf, noch sind sie in extrem großer Zahl zu erwarten. Dementsprechend ist – zumindest für den Demonstrator – eine *SQLite*-Lösung ausreichend.

Für einen Produktiveinsatz mit einer großen Menge von Energiedaten ist *SQLite* nicht geeignet. Stattdessen sollte eine der oben genannten Lösungen

⁴<https://sqlite.org/>

⁵<https://www.mysql.de/>

⁶<https://www.postgresql.org/>

(MySQL oder PostgreSQL) zum Einsatz kommen, da diese für deutlich größere Datenmengen ausgelegt sind. Kritischer als die in der relationalen Datenbank gespeicherten Metadaten dürften jedoch die eigentlichen Zeitreihendaten sein. Deshalb werden die Zeitreihen – wie in Abbildung 5.2 ersichtlich – nicht in der relationalen Datenbank, sondern in einer dedizierten Zeitreihendatenbank gespeichert (vgl. Abschnitt 5.3.3).

Es existieren Erweiterungen für *Django*, welche es erlauben, die Daten symmetrisch zu verschlüsseln, bevor sie in einer relationalen Datenbank abgelegt werden⁷. Eine solche Vorkehrung würde es theoretisch ermöglichen, eine relationale Datenbank zu verwenden, welche außerhalb der Kontrolle des Nutzers ist, auch wenn dies nicht zu empfehlen ist. Es kann jedoch eine Maßnahme sein, um den Schutz vor Angriffen auf sensible Nutzerdaten weiter zu erhöhen.

Administrationsschnittstelle für Datenmodelle

Django nutzt die in *Python* erstellten Datenmodelle, um eine Administrationschnittstelle aufzubauen, welche entsprechend autorisierten Personen zur Verfügung steht. Welcher Person oder welchen Personen Zugriff auf die Administrationschnittstelle gewährt wird, muss im jeweiligen Anwendungsfall entschieden werden. Dies kann einerseits ein *Nutzer* des Systems oder ein (externer) Administrator sein, welchem die *Nutzer* vertrauen. Es sollte jedoch in keinem Fall ein *Dateninteressent* sein, da diesem sonst ein Einblick in alle vorhandenen Energiedaten möglich wäre. Hierbei handelt es sich offensichtlich um ein Problem, welches die meisten Computersysteme betrifft.⁸

Die Schnittstelle liegt für den Administrator in Form einer Webanwendung vor, welche es unter anderem erlaubt, Objekte anzulegen, zu ändern oder zu löschen. Dies betrifft vor allem folgende Objekte, die auch in Abbildung 5.2 dargestellt sind: *Nutzer*, *Dateninteressenten*, *Parser*, *Einheiten*, *Kategorien*, *Qualitätsreduktionsmaßnahmen* und *Ausgabeformate*.

⁷vgl. bspw. <https://django-fernet-fields.readthedocs.io/en/latest/>

⁸Dieser Aspekt wird als „Law #6: A computer is only as secure as the administrator is trustworthy.“ in Microsofts *Ten Immutable Laws Of Security (Version 2.0)* zusammengefasst (vgl. <https://technet.microsoft.com/en-us/library/hh278941.aspx>).

Nicht autorisierte Personen erhalten keinen Zugriff auf den Demonstrator, dementsprechend ist ein Benutzerkonto für die Verwendung nötig. Ein Administrator kann zwei verschiedene Arten von Konten erzeugen, indem er ein neues Objekt – entweder vom Typ *Nutzer* oder vom Typ *Dateninteressent* – anlegt. So gesehen sind alle Benutzerkonten auch nur ein Objekt wie alle anderen und werden entsprechend in der relationalen Datenbank vorgehalten.

Die weiteren Objekte *Parser*, *Einheit*, *Ausgabeformat*, *Qualitätsreduktionsmaßnahme* und *Bewertungsmethode* hätten – wie in Abschnitt 5.2.2 erklärt – auch als Attribute anderer Objekte modelliert werden können. An dieser Stelle zeigt sich jedoch der Vorteil der Modellierung als eigenständige Objekte. Dies erlaubt es dem Administrator, im laufenden Betrieb Erweiterungen am System vorzunehmen. Sollen dem Demonstrator beispielsweise Energiedaten von einem neuen Gerätetyp zur Verfügung gestellt werden, ist ein neuer *Parser* nötig, welcher die Daten aufbereitet. Ein Administrator ist in der Lage, im laufenden Betrieb ein neues Objekt vom Typ *Parser* anzulegen, welches direkt nach seiner Erstellung zum Einlesen der Energiedaten verwendet werden kann. In einem solchen *Parser*-Objekt wird beispielsweise neben der Identifikationsnummer und einer Beschreibung auch eine entsprechende *Python*-Klasse gespeichert, welche die Parser-Funktionalität bereitstellt. Für die Objekte *Ausgabeformat*, *Qualitätsreduktionsmaßnahme* und *Bewertungsmethode* wird analog vorgegangen. Sobald eine neue Funktionalität hinzugefügt werden soll, kann ein Administrator das entsprechende Objekt anlegen und die erstellte *Python*-Klasse bereitstellen, welche ab diesem Zeitpunkt zur Verfügung steht.

Selbstverständlich müssen vor einem Produktiveinsatz alle neu hinzuzufügenden Klassen überprüft und getestet werden. Auch an dieser Stelle zeigt sich, wie wichtig ein quelloffener Ansatz ist. Darüber hinaus muss es sich bei dem Administrator um den *Nutzer* selbst oder um eine vertrauenswürdige Person handeln, da über das Hinzufügen neuer Funktionalitäten ein weitreichender Zugriff auf das System möglich ist. Dennoch eröffnet diese Vorgehensweise die Möglichkeit, den *Data Custodian Service* flexibel anzupassen und zu erweitern, so dass auf veränderte Anforderungen schnell reagiert werden kann.

Vorlagensystem zur Aufbereitung und Darstellung von Webseiten

Django arbeitet nach einer Abwandlung des *Model-View-Controller*-Prinzips. Die *Django Software Foundation* verwendet hierfür den verwandten Begriff *Model-Template-View* [Dja16a]. Unabhängig von den gewählten Begriffen geht es im Kern jedoch darum, eine situationsabhängige Sicht auf die im System vor gehaltenen Daten beziehungsweise Modelle zu ermöglichen. Die Umsetzung der Modellierung wurde bereits in Abschnitt 5.3.1 beschrieben. In diesem Abschnitt wird die Sicht auf die Daten behandelt. An dieser Stelle unterscheidet *Django* explizit zwischen sogenannten *Views* und *Templates*.

Aus der Perspektive von *Django* ist eine *View* – eine „Sicht“ – eine *Python*-Funktion, in welcher festgelegt wird, welche Daten angezeigt werden sollen. Dabei kann es sich zum Beispiel um eine Art Filter handeln: Beim Aufruf einer bestimmten URL (engl. *Uniform Resource Locator*) durch einen *Nutzer* sollen ausschließlich die ihn betreffenden *Datenanfragen* präsentiert werden. *Django* nutzt dann die objektrelationale Abbildung (vgl. Abschnitt 5.3.1), um die entsprechenden Daten aus der relationalen Datenbank abzurufen. Die *View* kann diese zwischengespeicherten Daten weiterverarbeiten und – falls nötig – zusätzliche Anfragen tätigen. Die aus den Anfragen resultierenden Daten werden im nächsten Schritt dem zur *View* gehörenden *Template* übergeben.

Ein *Template* – eine Vorlage – legt fest, wie die Daten dargestellt werden sollen. *Django* nutzt dazu eine Vorlagensprache, die die Erstellung von Webseiten mit Platzhaltern ermöglicht, welche im Betrieb gefüllt werden. Auch die Verwendung bestimmter Funktionen wie beispielsweise Schleifen und weitergehender Filter ist möglich. Bei der Verarbeitung eines *Templates* werden die entsprechenden Platzhalter ersetzt und so schrittweise eine HTML-Seite aufgebaut. Diese HTML-Seite wird an einen Webserver übergeben und kann dann im Browser des Endgerätes dargestellt werden. Moderne Webtechnologien wie *HTML5*, *CSS3* und *JavaScript* ermöglichen die Erstellung von Webseiten, welche sich gleichermaßen auf normalen Computern oder mobilen Endgeräten interaktiv im Browser und ohne dedizierte Anwendung nutzen lassen.

Templates können nicht nur für die Interaktion mit Menschen, sondern auch für die Kommunikation mit anderen Rechnern verwendet werden, wobei diesbezüglich auch eine dedizierte Lösung zur Verfügung steht: das *Django*-

*REST-Framework*⁹. Die Abkürzung *REST* steht für *Representational State Transfer* und beschreibt ein Paradigma zur Erstellung eines Systems für die Maschine-zu-Maschine-Kommunikation. Im Gegensatz zu einer Webanwendung steht nicht die grafische Darstellung von Inhalten im Vordergrund, sondern die strukturierte Übermittlung von Daten, so dass automatisierte Anfragen effizient verarbeitet werden können.

Zuordnung von URLs zu Sichten

Wird zur Beschreibung der Funktionsweise von *Django* nicht der Begriff *Model-Template-View*, sondern das stark verwandte *Model-View-Controller*-Prinzip verwendet, dann nimmt *Django* selbst die Rolle des *Controllers* ein. Eine der Aufgaben, die *Django* erfüllt, ist das Zuordnen von URLs zu *Views*.

Wie im vorhergehenden Abschnitt bereits erklärt, können über den Aufruf verschiedener URLs unterschiedliche *Views* erreicht werden. Zu diesem Zweck werden reguläre Ausdrücke definiert, welche jeweils einzelnen *Views* und damit *Python*-Funktionen zugewiesen sind. So wird beispielsweise der reguläre Ausdruck `^datenanfrage_detail_nutzer/(?P<pk>[0-9]+)` der Funktion für die Ausgabe der Details zu einer Datenanfrage mit dem Primärschlüssel *pk* zugewiesen. Die regulären Ausdrücke sind in *Python*-Syntax¹⁰ notiert.

Beim Aufruf einer URL – entweder durch Personen in einem Browser oder durch Maschinen über die REST-Schnittstelle – prüft *Django* die regulären Ausdrücke und führt die entsprechende *View* aus. Das Ergebnis dieser Funktion wird anschließend wie zuvor beschrieben über ein *Template* in eine HTML-Seite überführt und kann dann über einen Webserver an den Aufrufenden übergeben werden.

Authentifizierung und Autorisierung

Django stellt diverse Funktionen zur Authentifizierung und Autorisierung seiner Clients bereit. Nutzer – im Sinne von *Nutzern* und *Dateninteressenten* im Bereich des *Data Custodian Service* (vgl. Abbildung 5.1) – sind als Objekte

⁹<http://www.djangoproject-rest-framework.org/>

¹⁰Für weitere Informationen hierzu sei erneut auf die Dokumentation zu *Django* [Dja16b] verwiesen.

in *Django* modelliert. Diese Nutzer werden jeweils unterschiedlichen Nutzergruppen zugeordnet: *Nutzer* und *Dateninteressenten*. Zusätzlich existiert noch mindestens ein Administrator, welcher für die Verwaltung des Systems zuständig ist, im Rahmen des *Data Custodian Service* aber keine weitere Rolle spielt. Dabei kann es sich wie oben beschrieben um einen *Nutzer* oder eine externe, vertrauenswürdige Person handeln.

Erst nach einer Authentifizierung der Nutzer können diese das System nutzen. Abhängig von der Nutzergruppe, welcher sie angehören, können die Nutzer unterschiedliche Sichten aufrufen. Für diese Autorisierung nutzt *Django* ein Rechtemanagement, welches es unterschiedlichen Nutzern beziehungsweise Nutzergruppen gestattet oder verbietet, bestimmte URLs zu öffnen. Da die URLs direkt mit den entsprechenden *Views* verknüpft sind, umfasst die Autorisierung auch diese. Einerseits wird auf diese Weise dafür gesorgt, dass lediglich sinnvolle Aktionen durchgeführt werden können: Ein *Nutzer* darf beispielsweise keine *Datenanfragen* aufgeben, dafür müsste er ein *Dateninteressent* sein beziehungsweise über ein solches Nutzerkonto verfügen. Andererseits verhindert *Django* damit auch nicht gestattete Zugriffe auf Daten: Ein *Dateninteressent* wird daran gehindert, andere als die eigenen *Datenanfragen* zu betrachten, ebenso wie ein *Nutzer* lediglich die eigenen *Energiedaten* betrachten kann.

Webserver

Django verfügt über einen integrierten Webserver, welcher allerdings ausdrücklich nicht für den Produktiveinsatz, sondern für die Entwicklung und das Debugging vorgesehen ist. Der hier beschriebene und als *Proof-of-Concept*-Umsetzung gedachte Demonstrator nutzt zwar den integrierten *Django*-Webserver, ist aber auch mit einem proprietären Webserver einsetzbar. *Django* ermöglicht hierfür die Nutzung der *WSGI* (*Python Web Server Gateway Interface*) genannten Schnittstelle [PJ 10]. *WSGI* ist ein in *Python* verankerter Standard, welcher eine Schnittstelle zwischen *Python*-Applikationen und Webservern definiert. Hierdurch können etablierte Produkte wie der *Apache HTTP Server*¹¹

¹¹<https://httpd.apache.org/>

oder *nginx*¹² eingesetzt werden. Für weitere Informationen hierzu sei auf die entsprechende *Python*-Richtlinie verwiesen [PJ 10].

Der Webserver ist die Schnittstelle des *Data Custodian Service* zur Außenwelt. Würde die Kommunikation mit den Interaktionspartnern an dieser Stelle ungeschützt erfolgen, wäre es möglich, die übertragenen Daten abzugreifen. Aus diesem Grund muss die gesamte Kommunikation des *Data Custodian Service* mit den jeweiligen Interaktionspartnern verschlüsselt werden. Eine etablierte Art der Transportverschlüsselung stellt hierfür *HTTPS* (*HyperText Transfer Protocol Secure*) dar. Neben der Verschlüsselung ermöglicht *HTTPS* auch eine Authentifizierung, so dass Nutzer die Identität des Webservers prüfen können. Dies ermöglicht es, potenzielle *Man-in-the-Middle*- oder *Phishing*-Attacken auf diesem Weg zu verhindern oder zumindest zu erschweren.

Schutz vor Angriffen

Obwohl das System durch ein Authentifizierungs- und Autorisierungssystem geschützt ist, könnten potenzielle Angreifer versuchen, dieses zu umgehen oder auszuhebeln, um somit – unbefugten – Zugriff auf fremde Daten zu erhalten. Es gibt zahlreiche verschiedene Angriffszenarien und Methoden, welche dabei zum Einsatz kommen können, wie beispielsweise *Cross-Site-Scripting*. Diese Art von Angriff kann unter anderem dafür verwendet werden, die Session angemeldeter Nutzer zu übernehmen und somit deren Identität anzunehmen. Darüber hinaus ist in manchen Fällen auch das Abgreifen oder Manipulieren fremder Daten denkbar.

Django bietet diverse Sicherheitsvorkehrungen, welche diese Angriffe sowie weitere erschweren oder verhindern sollen [Dja16a]¹³. Für gewöhnlich ist davon auszugehen, dass Sicherheitslücken und -probleme nie vollkommen ausgeschlossen werden können. Dennoch ist es in der Regel von Vorteil, wenn der Sicherheitsaspekt bei der Entwicklung einer Software im Vordergrund steht und der Quellcode der verwendeten Programme offenliegt – beide Aspekte werden von *Django* erfüllt. Trotzdem muss in einem Produktiveinsatz auf übliche Sicherheitspraktiken und deren gewissenhafte Anwendung geachtet

¹²<https://nginx.org/>

¹³Technische Details sind unter anderem unter <https://docs.djangoproject.com/en/1.10/topics/security/> zu finden.

werden. Darüber hinaus muss die verwendete Software stets auf einem aktuellen Stand gehalten werden, um entsprechende Sicherheitsaktualisierungen frühzeitig zum Einsatz zu bringen¹⁴. In Summe scheint *Django* in Kombination mit einem geeigneten, etablierten Webserver vergleichsweise sicher zu sein. Dies gilt nicht zuletzt deshalb, weil eine aktive Weiterentwicklung des Frameworks erfolgt und einige stark frequentiere Webseiten wie *Instagram*, *Disqus* und *Mozilla* die Software *Django* einsetzen, so dass vermutlich auch zukünftig Interesse an einer aktiven Weiterentwicklung besteht. Für eine abschließende Bewertung wären jedoch Sicherheitstest durch entsprechende Experten erforderlich.

Zusammenfassung Softwaregerüst

Das *Django*-Framework übernimmt als Grundgerüst des Demonstrators eine Reihe von Standardaufgaben, die in einer Webanwendung anfallen. Darüber hinaus ermöglicht es die Umsetzung des Datenmodells wie in Abbildung 5.2 beschrieben. Von den in Abbildung 5.1 gezeigten Komponenten werden einige von *Django* gestellt beziehungsweise unterstützt. Zunächst sind der *Data Custodian Server* und der *Data Custodian Anwendungskern* zu nennen. Darüber hinaus stellt *Django* die *Grafische Benutzeroberfläche* und die *Maschine-zu-Maschine-Schnittstelle* zur Verfügung. Schließlich dient die objektrelationale Abbildung von *Django* noch als *Datenbankzugriff*.

Wie der Ausdruck Framework jedoch andeutet, handelt es sich bei *Django* lediglich um ein Rahmenwerk – ein Grundgerüst –, welches als Basis für die zu entwickelnde Anwendung dienen kann. Alles über die Grundfunktionalität Hinausgehende und die eigentliche Programmlogik müssen selbstverständlich noch implementiert werden. Dies umfasst insbesondere den Umgang mit den Zeitreihen und deren Bearbeitung sowie die Bewertung eingehender Datenanfragen. Diese Aspekte werden in den folgenden Abschnitten dieses Kapitels besprochen.

¹⁴Der Umgang mit Sicherheitslücken in *Django* ist unter <https://docs.djangoproject.com/en/dev/internals/security/#how-django-discloses-security-issues> beschrieben.

5.3.2. Verarbeitung von Zeitreihen

Aus Abbildung 5.1 wird ersichtlich, dass für den Betrieb des *Data Custodian Service* ein Modul zur Verarbeitung beziehungsweise Vorverarbeitung von Zeitreihen nötig ist. Für den Demonstrator wird mit *Pandas*¹⁵ eine solche Bibliothek verwendet. Die Hauptanwendungsfälle sind hier die Unterstützung beim Parsen von Energiedaten, deren Aufbereitung, das Reduzieren der Ausagekraft beziehungsweise der Datenqualität und das Umwandeln für eine Datenausgabe. Diese Aspekte werden in den folgenden Unterabschnitten näher besprochen.

Pandas wurde ursprünglich für Anwendungen im Finanzsektor geschaffen, hat sich mittlerweile aber zu einem mächtigen Werkzeug für verschiedenste Arten von tabellenartigen Daten entwickelt: Es unterstützt Entwickler beim Einlesen der Daten sowie bei deren Aufbereitung, Speicherung und der Analyse. Die folgenden Ausführungen basieren größtenteils auf den Veröffentlichungen [McK+10; McK12] von Wes McKinney, dem ursprünglichen Autor von *Pandas*, sowie auf der dazugehörigen Onlinedokumentation¹⁶.

Erklärtes Ziel von *Pandas* ist es, das mächtigste, *quelloffene* Werkzeug für die Analyse und die Verarbeitung von Daten aus Realweltanwendungen zu werden. Dies spiegelt sich in der Anwendung insofern wider, als *Pandas* einerseits einen umfangreichen Satz an Befehlen aufweist und andererseits die zugrundeliegenden Routinen in hochoptimiertem C beziehungsweise C++ vorliegen. Für den Demonstrator bedeutet diese Situation wiederum, dass flexible und umfangreiche Datenmanipulationen möglich sind, welche jedoch auch effizient und schnell durchgeführt werden können.

Im Kern nutzt *Pandas* vor allem zwei Arten von Datenstrukturen: eindimensionale *Series*- und zweidimensionale *DataFrame*-Objekte. *Series* sind – wie der Name schon andeutet – für Reihen von Daten vorgesehen, die sich in einer einzelnen Tabellenspalte abbilden lassen. *DataFrames* hingegen entsprechen einer Tabelle mit beliebig vielen Spalten beziehungsweise einer Matrix; darüber hinaus sind auch mehrdimensionale Datenstrukturen darstellbar. Das Ziel, *Pandas* für Realweltdaten nutzbar zu machen, schlägt sich bereits an dieser Stelle nieder. Es ist vorgesehen, nicht vorhandene Daten explizit mit

¹⁵<http://pandas.pydata.org/>

¹⁶<http://pandas.pydata.org/pandas-docs/stable/>

5.3. Umsetzung des Konzepts in einem Demonstrator

dem Kürzel „NaN“ (von engl. *not a number*) zu kennzeichnen. Insbesondere bei Energiedaten ist dieser Aspekt – je nach Datenqualität – von besonderem Wert. Es ist nicht nötig, fehlende Daten durch „spezielle“ Gleitkommazahlen zu codieren, sondern es kann explizit ausgedrückt werden, dass an einer bestimmten Stelle keine Daten vorhanden sind. Abgesehen von Gleitkommazahlen können auch andere *Python*-Datentypen in den oben genannten Datenstrukturen gespeichert werden.

Pandas bietet eine Reihe von Funktionen, um mit den Datenstrukturen *Series* und *DataFrame* zu arbeiten. Es ist unter anderem möglich, Daten zusammenzuführen, zu extrahieren oder zu filtern. Des Weiteren sind einige auf die Manipulation von Zeitreihen spezialisierte Methoden verfügbar, welche für die Verarbeitung von Energiedaten geeignet sind. Die für den Demonstrator relevanten Funktionen erlauben es beispielsweise, Daten nach angegebenen Zeiträumen auszuwählen oder eine Umrechnung in andere Frequenzen – ein *Resampling* – vorzunehmen. Auch die Berechnung verschiedener statistischer Kennzahlen auf Basis der Energiedatenzeitreihen ist mit diesen Werkzeugen möglich. Konkrete Anwendung finden diese Funktionalitäten im Demonstrator vor allem im Bereich des Einlesens und der Aufbereitung der oft lückigen oder fehlerhaften Daten, der Bewertung von Datenanfragen und der Aufbereitung der Daten vor der Datenausgabe. Alle Zeitreihendaten innerhalb des Demonstrators liegen in Form von *DataFrames* vor. Aus diesem Grund erscheint es sinnvoll, beim Erstellen von Parsern für neue Arten von Energiedaten ebenfalls *Pandas* als Grundlage zu nutzen, auch wenn dies nicht zwangsläufig nötig ist.

Es existieren abgesehen von *Pandas* weitere Werkzeuge zur Bearbeitung von Zeitreihendaten. Für den *Data Custodian Service* beziehungsweise den Demonstrator weist *Pandas* jedoch einige Vorteile auf. Zunächst ist die Software quelloffen und wird aktiv von einer großen Zahl aktiver Entwickler gepflegt¹⁷. Darüber hinaus wurde bei der Entwicklung von *Pandas* von Anfang an großen Wert auf die Eignung für Realweltdaten gelegt – Daten wie Energiedaten, welche oft fehlerbehaftet oder beispielsweise schlecht formatiert sein können. Schließlich spielt auch die Leistung bei der Datenmanipulation eine große Rolle. Auch wenn die in Haushalten üblicherweise anfallenden Mengen von

¹⁷Siehe unter anderem <https://github.com/pandas-dev/pandas/commits/master>.

Energiedaten diesbezüglich selten Probleme aufwerfen, ist im Hinblick auf die Skalierbarkeit eine leistungsfähige Lösung vorzuziehen. In Summe ist *Pandas* eine geeignete Bibliothek zur Unterstützung bei der Verarbeitung von Energiedaten zu sein und findet deshalb Anwendung im Demonstrator des *Data Custodian Service*.

Parsen der Zeitreihen

In Abschnitt 4.1.2 wurde die Erfassungsphase des Datenlebenszyklus untersucht. Aus diesen Untersuchungen geht hervor, dass Energiedaten in zum Teil sehr unterschiedlichen Arten von Zeitreihen vorliegen können. Die Rohdaten unterscheiden sich hinsichtlich der verwendeten Dateiformate ebenso wie in den zugrundeliegenden Einheiten. Weitere Erschwernisse wie beispielsweise fehlende Angaben zu Zeitzonen beeinträchtigen die korrekte Erfassung der Energiedaten zusätzlich. Dementsprechend ist für jede Art von Energiedatenquelle ein geeigneter Parser zu erstellen, welcher die Rohdaten jeweils mit ihren individuellen Eigenschaften korrekt einlesen kann.

Das Datenmodell des *Data Custodian Service* sieht vor, dass bei jedem Hinzufügen neuer Energiedaten ein *Parser* angegeben wird (vgl. Abbildung 5.2). Ein jedes *Parser*-Objekt enthält eine Referenz auf eine *Python*-Klasse, welche die nötigen Funktionen enthält, um Energiedaten eines bestimmten Typs einzulesen und in ein *DataFrame*-Objekt zu überführen. Dementsprechend muss für jede Art von Energiedaten eine entsprechende Implementierung vorliegen, sowie ein darauf verweisendes *Parser*-Objekt angelegt werden. Auch wenn es nicht unbedingt erforderlich ist, ist es dennoch sinnvoll, auch an dieser Stelle *Pandas* als Basis für die Erstellung eines solchen Energiedatenparsers zu verwenden. Wie bereits zuvor beschrieben, bietet *Pandas* eine Reihe von grundlegenden Operationen an, die das Einlesen und die Verarbeitung von Realweltdaten vereinfachen. Auf diesen Grundfunktionalitäten aufbauend können Parser implementiert werden, welche auch mit lückenhaften oder schlecht formatierten Daten arbeiten können. Ziel ist es, dass am Ende der Bearbeitungsschritte sauber aufbereitete Energiedaten in der *DataFrame*-Datenstruktur vorliegen, welche im Demonstrator verwendet wird. Damit können alle weiteren Operationen transparent auf den Daten ausgeführt werden, unabhängig von der Herkunft der Daten beziehungsweise von der Art der Energiedatenquelle. Bei

der im nächsten Unterabschnitt angesprochenen Reduktion der Aussagekraft der Daten handelt es sich um eine solche Operation.

Reduktion der Aussagekraft

Der *Data Custodian Anwendungskern* macht intensiven Gebrauch vom Modul zur *Zeitreihen(vor)verarbeitung*, um den Informationsgehalt und die Aussagekraft der Energiedaten beziehungsweise deren Datenqualität gemäß den Präferenzen der Nutzer zu senken. Erfolgen kann dies beispielsweise durch die Reduktion der zeitlichen Auflösung, das Hinzufügen von Rauschen oder ähnliche Methoden. Dieser zentrale Aspekt wird gesondert in Abschnitt 5.3.5 behandelt. Dort werden die entsprechenden Methoden auch detailliert vorgestellt.

Ausgabe der Energiedaten

Beim Einreichen einer *Datenanfrage* muss ein *Dateninteressent* das gewünschte *AusgabefORMAT* angeben (vgl. Abbildung 5.2). Bei solchen Dateiformaten kann es sich beispielsweise um Formate wie *CSV*, *XML*, *JSON* und ähnliche handeln. Theoretisch wären auch *Usage-Control-Container* denkbar (vgl. Abschnitt 3.2.2.7), mithilfe derer ein digitales Rechtemanagement durchgesetzt werden kann. Aus technischer Sicht ist das ausgabeseitige *AusgabefORMAT* das Gegenstück zum *Parser*: Auf der Eingangsseite wandelt der *Parser* Rohdaten in eine *DataFrame*-Struktur um und auf der Ausgabeseite werden die Daten von einem *DataFrame* in das gewünschte *AusgabefORMAT* übersetzt.

Analog zu den *Parser*-Objekten enthält jedes *AusgabefORMAT*-Objekt ebenfalls eine Referenz auf eine *Python*-Klasse. Innerhalb dieser Klasse wird die Funktionalität bereitgestellt, um eine in der *DataFrame*-Struktur vorliegende Zeitreihe in ein bestimmtes AusgabefORMAT wie *CSV*, *XML* oder *JSON* zu übersetzen. Die Ausgabe ist aber, wie oben angedeutet, nicht auf solche klassischen Dateiaustauschformate beschränkt. Es ist ebenfalls denkbar, mit *DRM*-Technologie ausgestattete Ausgabemodule zu verwenden. Zum Einsatz kommen könnte beispielsweise *Usage Control* wie in Abschnitt 3.2.2.7 beschrieben.

Wie auch bei den *ParseRN* ist es nicht unbedingt nötig, aber durchaus sinnvoll, die von *Pandas* gebotenen Funktionalitäten zu nutzen, um die Übersetzung

in spezielle *Ausgabeformate* zu implementieren. *Pandas* bietet darüber hinaus aber auch Exportmöglichkeiten für einige häufig verwendete Formate.

5.3.3. Speicherung der Zeitreihen

Abbildung 5.2 zeigt die Trennung nach unterschiedlichen Datenbanken, welche für die Speicherung der jeweiligen Objekte verwendet werden. Bis auf eine Entität werden alle anderen in einer Metadatendatenbank vorgehalten. Im Falle des Demonstrators handelt es sich bei der Metadatendatenbank um die relationale Datenbank *SQLite* (vgl. Abschnitt 5.3.1). Für die Speicherung der Zeitreihendaten ist folglich eine weitere Lösung nötig (vgl. auch Abbildung 5.1). Analog zu den anderen Komponenten soll auch für die Speicherung der Zeitreihen ein Produkt zum Einsatz kommen, welches etabliert sowie quelloffen ist und weiter aktiv gepflegt wird. Darüber hinaus kann der Verwaltungsaufwand für das System reduziert werden, wenn kein weiterer Server für den Betrieb einer Datenbank nötig ist. Nicht zuletzt aus diesen Gründen fiel die Wahl auf *HDF5*¹⁸. Die folgenden Ausführungen beruhen größtenteils auf einer Veröffentlichung [Fol+11] der für *HDF5* verantwortlichen *HDF Group* und der Dokumentation zu *HDF5*¹⁹.

Die Abkürzung *HDF5* steht für *Hierarchical Data Format 5* und beschreibt nicht nur ein Datenformat, sondern eine Technologie zur Speicherung großer Datenmengen. Die *HDF5 technology suite* umfasst laut *HDF Group* die folgenden Aspekte:

- Ein Datenmodell
- Ein portables Datenformat
- Softwarebibliotheken zur Nutzung von *HDF5* in diversen Programmiersprachen
- Maßnahmen zur Optimierung der Datenhaltung und -zugriffe
- Werkzeuge und Anwendungen für die Verwaltung, Manipulation, Sicherung und Analyse von Daten

¹⁸<https://www.hdfgroup.org/hdf5/>

¹⁹<https://support.hdfgroup.org/HDF5/>

Durch die Verwendung von *B-Trees* für die Indizierung der Daten eignet sich *HDF5* besonders gut für Daten, welche in tabellarischer Form dargestellt werden können [Nat16]. Dies betrifft nicht nur die Speicherung, sondern auch das Auslesen und Bearbeiten der Daten. Für weitere technische Details sei an dieser Stelle auf die ausführliche Dokumentation verwiesen. Im weiteren Verlauf stehen die Anwendungsmöglichkeiten des Systems im Vordergrund.

Der Vorgänger *HDF4* und insbesondere die aktuelle Version *HDF5* werden bereits seit vielen Jahren in zahlreichen wissenschaftlichen Anwendungsbereichen eingesetzt. Beispielhaft soll an dieser Stelle die *NASA Earth Science Community* genannt werden, welche den Einsatz von *HDF5* empfiehlt und befürwortet [Nat16]. In den Ausführungen werden aus Sicht der Community einige Vorteile genannt, die ebenfalls bei der Verwendung von *HDF5* für Energiedaten zutreffend sind:

- Unterstützung für ausschnittweises Lesen von Dateien
- Nutzbar auf annähernd allen Plattformen
- Quellcode liegt offen
- Unterstützung durch Drittanbietersoftware wie beispielsweise *Matlab*

Bei großen Datenmengen ist es – mangels Ressourcen – häufig nicht möglich, alle Daten in den Hauptspeicher zu laden. Dementsprechend ist es oft nötig, nur die tatsächlich gewünschten Ausschnitte der Daten zu lesen. Aber auch bei kleineren Datenmengen, welche eigentlich noch in den Hauptspeicher passen würden, kann dies zu einer schnelleren Ausführung und größerer Effizienz führen. Angewandt auf Energiedaten bringt ein ausschnittweises Lesen der Daten beispielsweise den Vorteil mit sich, dass nur die Zeitbereiche gelesen werden müssen, über welche ein Nutzer Informationen haben möchte. Alle anderen Bereiche der Energiedaten bleiben unangetastet.

Bei der Entwicklung von *HDF5* wurde auf eine Unterstützung möglichst vieler unterschiedlicher Plattformen geachtet. Dieser Aspekt betrifft sowohl Betriebssysteme als auch die Art der verwendeten Rechner: Die Technologie soll auf kleinen Rechnern wie etwa Laptops ebenso nutzbar sein wie auf Großrechnern – weitgehend unabhängig vom verwendeten Betriebssystem.

Für den Demonstrator bedeutet dies, dass *HDF5* einerseits auf sehr kleinen Systemen eingesetzt werden kann – beispielsweise auf einem kleinen Server in einem Haushalt. Andererseits ist *HDF5* kein Hinderungsgrund für eine eventuelle Portierung auf deutlich leistungsfähigere Systeme.

Wie bereits an anderen Stellen betont, ist ein offener Quellcode einer der Wege, um das Vertrauen der Nutzer in eine Software zu verbessern. Darüber hinaus ist eine große Gemeinschaft an der Entwicklung von *HDF5* beteiligt beziehungsweise setzt dieses System seit vielen Jahren ein. Dementsprechend ist zu vermuten, dass auch zukünftig eine Weiterentwicklung der Technologie erfolgen wird.

Der letzte oben aufgezählte Punkt betrifft die Unterstützung durch Drittanbietersoftware. Dieser Aspekt darf nicht unterschätzt werden, da die *NASA Earth Science Community* deutlich den Nachteil hervorhebt, dass die Verwendung von *HDF5* durch dessen Komplexität mit einer signifikanten Lernkurve versehen ist. Wären entsprechende Module zur Verwendung von *HDF5* nicht verfügbar, wäre die Anwendung dieser Technologie im Rahmen des Demonstrators äußerst unwahrscheinlich. Das Softwarepaket *Pandas* (vgl. Abschnitt 5.3.2) bringt jedoch eine Schnittstelle für das Lesen und Schreiben von *HDF5*-Dateien mit, welche auch durch den Demonstrator genutzt wird. Der Demonstrator profitiert damit von einer seit vielen Jahren etablierten Technologie, welche serverlos funktioniert und gleichzeitig sehr performant ist.

Zunächst sollte davon ausgegangen werden, dass alle Energiedaten lokal vorgehalten werden. Für den Fall, dass die *HDF5*-Dateien jedoch nicht lokal, sondern auf einem externen Server gespeichert und verwaltet werden sollten, müssen die Daten vor unbefugtem Zugriff geschützt werden. *HDF5* bietet mit sogenannten Filtern eine native Lösung dafür an. Solche Filter werden normalerweise für die transparente Komprimierung der zu speichernden Daten verwendet, können jedoch auch für eine Verschlüsselung eingesetzt werden. Diese Funktion befindet sich jedoch noch im Betastadium und ist deshalb zurzeit nicht zu empfehlen²⁰.

²⁰<https://support.hdfgroup.org/projects/boeing/>

Für die Speicherung von Zeitreihen existieren selbstverständlich alternative Lösungen zu *HDF5*. Beispielhaft seien *OpenTSDB*²¹ und *InfluxDB*²² genannt. Beide Lösungen sind jedoch im Gegensatz zu *HDF5* serverbasiert und unterlagen in jüngerer Vergangenheit zum Teil starken Veränderungen. Falls im Rahmen des *Data Custodian Service* eine der zuvor genannten Technologien zum Einsatz kommen soll, ist diese Änderung dank der modularen Architektur (vgl. Abbildung 5.1) recht einfach möglich: Die Daten werden über einen Schlüssel in der *Datenbeschreibung* referenziert und damit aus der Zeitreihendatenbank geladen. Für die Implementierung bedeutet dies schlicht, dass mithilfe eines gegebenen Schlüssels die korrekten Daten aus der Zeitreihendatenbank abgerufen und dann in die oben beschriebene *DataFrame*-Datenstruktur überführt werden müssen, mit welcher der *Data Custodian Anwendungskern* arbeitet.

5.3.4. Bewertung der Datenanfragen

Der *Data Custodian Anwendungskern* nutzt ein Modul zur *Datenanfragenbewertung*, um die potenziellen Auswirkungen einer *Datenanfrage* auf die Privatsphäre der betroffenen *Nutzer* zu beurteilen (vgl. Abbildung 5.1). Aus Nutzersicht spielt die Komponente zur Bewertung von Datenanfragen eine besondere Rolle: Mithilfe ihrer Einschätzung sollen die *Nutzer* in der Lage sein, zu beurteilen, ob eine Weitergabe der in einer *Datenanfrage* angegebenen Daten gestattet oder verweigert werden soll. Eine Fehleinschätzung könnte für die Betroffenen zu einem signifikanten Verlust von Privatsphäre und damit einhergehenden weiteren Beeinträchtigungen führen.

In Abschnitt 4.2.2 wurden zahlreiche verschiedene Verfahren zur Extraktion von Informationen aus Energiedaten vorgestellt. Des Weiteren wurden in Abschnitt 4.2.3 die möglichen Auswirkungen auf die Privatsphäre der Betroffenen besprochen. Die Analyse der verschiedenen Methoden zeigt, dass abhängig von den jeweiligen Verfahren sehr unterschiedliche Ziele verfolgt werden. Darüber hinaus unterscheiden sich die Vorgehensweisen und die benötigten Daten meist grundlegend. Aus diesem Grund ist eine exakte Abschätzung des Privatsphärenverlusts kaum möglich. Deshalb ist es sinnvoll, eine allgemeine, pessimistische Abschätzung der potenziellen Auswirkungen

²¹<http://opentsdb.net/>

²²<https://www.influxdata.com/time-series-platform/influxdb/>

einer Energiedatenanalyse auf die Privatsphäre durchzuführen. Dies erfolgt ohne Kenntnis der tatsächlich zum Einsatz kommenden Verfahren zur Informationsextraktion aus Energiedaten. Im Folgenden wird die Umsetzung des Bewertungsmoduls im Demonstrator beschrieben.

Das Bewertungsmodul stellt die Auswirkungen auf die Privatsphäre mithilfe mehrerer einheitenloser Gefährdungskennzahlen dar, welche angeben, wie kritisch der jeweils betrachtete Aspekt ist. Der mögliche Wertebereich ist auf das Intervall [0; 100] normiert und eine höhere Bewertung entspricht einer größeren potenziellen Gefährdung der Privatsphäre. Die Gefährdungskennzahl wird darüber hinaus auf eine Farbskala in Form einer Ampel abgebildet, so dass besonders hohe Werte mit einer roten Markierung versehen werden, mittlere Werte mit einer gelben und unkritische Werte mit einer grünen (vgl. Abschnitt 5.4).

Die Kennzahlen beziehen sich ausschließlich auf die *Datenanfrage* des *Dateninteressenten*. Die betroffenen Energiedaten sowie deren tatsächlicher Informationsgehalt spielen zu diesem Zeitpunkt keine Rolle, da es sich um eine Abschätzung des potenziellen Risikos der *Datenanfrage* handelt und nicht um eine Beurteilung des tatsächlichen Informationsgehalts der Daten. Es handelt sich um einen heuristischen Ansatz, bei welchem vor allem Eigenschaften betrachtet werden, die gemäß Literatur die Aussagekraft von Energiedaten beeinflussen können (vgl. Abschnitt 4.2.2). Die im Folgenden beschriebenen Parameter wurden so gewählt, dass in der Literatur (vgl. Abschnitt 4.2.2) qualitativ beschriebene Zusammenhänge von Eigenschaften und Aussagekraft von Energiedaten abgebildet werden können.

In den folgenden Abschnitten 5.3.4.1 bis 5.3.4.4 werden Bewertungsfunktionen für einzelne Aspekte einer Datenanfrage definiert: für die zeitliche Auflösung, den angefragten Zeitraum, das Alter der Daten sowie für die Kategorie. Wie diese jeweiligen Kennzahlen auf eine einzelne reduziert werden können, wird in Abschnitt 5.3.4.5 gezeigt.

5.3.4.1. Zeitliche Auflösung

Die zeitliche Auflösung von Energiedaten ist durch zwei miteinander verknüpfte Eigenschaften definiert: die Frequenz und die Periodenlänge. Die Frequenz gibt an, wie viele Werte pro Zeiteinheit vorhanden sind. Die Peri-

odenlänge gibt wiederum an, welcher Zeitraum jeweils zwischen zwei Messwerten liegt. Damit handelt es sich bei der Frequenz um den Kehrwert der Periodenlänge und umgekehrt: Eine hohe zeitliche Auflösung geht mit einer hohen Frequenz und einer niedrigen Periodenlänge einher. Die Periodenlänge und die Frequenz können somit austauschbar verwendet werden, wobei im Folgenden die Periodenlänge präferiert wird.

Die zeitliche Auflösung von Energiedaten hat einen großen Einfluss auf deren Informationsgehalt. Je höher die Auflösung ist (je kleiner die Periodenlänge ist), desto mehr Informationen lassen sich potenziell aus den Daten extrahieren. Dieser Zusammenhang erscheint auf den ersten Blick trivial. Beachtenswert ist jedoch, dass sich nicht nur die Ergebnisse der eingesetzten Verfahren verbessern können, sondern dass mit steigender Auflösung auch zusätzliche Verfahren zur Verfügung stehen. Solche Verfahren nutzen beispielsweise die charakteristischen Profile von Einschalt- und Einschwingvorgängen, um elektrische Geräte zu erkennen. Diese Effekte sind jedoch in der Regel nur bei sehr hohen Auflösungen – zum Teil im Kilohertzbereich – sichtbar. Aber auch bei einer Periodenlänge von wenigen Sekunden sind oft noch erstaunlich zuverlässige Aussagen über verwendete Geräte oder deren Eigenschaften möglich. Das Verhalten von Bewohnern sowie An- oder Abwesenheitszeiten lassen sich auch aus deutlich niedriger aufgelösten Daten noch ableiten (vgl. Abschnitt 4.2.2.3).

Aus diesen Gründen sollte die zeitliche Auflösung der herauszugebenden Energiedaten so gering wie möglich gehalten werden, um einen Privatsphärenverlust zu verringern. Folglich müssen Datenanfragen mit höherer zeitlicher Auflösung (geringerer Periodenlänge) strenger bewertet werden als solche mit einer niedrigeren Auflösung (größerer Periodenlänge). Die folgende Formel zeigt, wie die Gefährdungskennzahl für die zeitliche Auflösung G_r berechnet wird (r von Englisch *resolution*).

$$G_r = \max \left\{ 0; 100 - \left(\frac{r}{r_0} \right)^{x_r} \cdot 100 \right\}$$

mit $0 < r, \quad 0 < x_r \leq 1, \quad 0 < r_0$

Die Variable r steht für die in der Datenanfrage gewünschte Auflösung in Form der Periodenlänge in Sekunden (beziehungsweise des Kehrwerts der Frequenz in Hertz, sollte die Frequenz verwendet werden). Bei x_r und r_0 handelt es sich um Parameter zur Beeinflussung der Bewertungsfunktion.

Der Parameter r_0 dient der Normierung der Auflösung beziehungsweise der Periodenlänge r . Im Demonstrator wird für r_0 der Wert $(365 \cdot 24 \cdot 60 \cdot 60)$ verwendet, was der Anzahl an Sekunden in einem Jahr entspricht. Die Bewertungsfunktion erreicht mit diesem Parameter bei einer angefragten Periodenlänge von einem Jahr oder mehr ihr Minimum von $G_r = 0$. Dies scheint insofern sinnvoll, als dass mit einer Veröffentlichung des Jahresverbrauchs oder noch geringer aufgelösten Daten vermutlich keine oder nur eine geringe Einschränkung der Privatsphäre einhergeht. Falls gewünscht, kann r_0 weiter erhöht werden, um auch gröbere Auflösungen noch negativ zu bewerten: Mit steigendem r_0 erreicht G_r auch erst mit steigenden Auflösungen (Periodenlängen) den Wert 0.

Eine sehr hohe Auflösung in Form einer gegen null gehenden Periodenlänge (beziehungsweise einer gegen unendlich gehenden Messfrequenz) führt zu einer Gefährdungskennzahl von maximal $G_r = 100$. Somit ist diese Bewertungsfunktion auf das Intervall $[0; 100]$ beschränkt.

Der Parameter x_r beeinflusst das Gewicht, welches auf höhere Auflösungen gelegt wird. Ein kleineres x_r belohnt eine Reduktion der Auflösung im kritischen Bereich stärker: Es macht einen größeren Unterschied, wenn die Periodenlänge von einer Sekunde auf einen Tag erhöht wird, als wenn sie von einer Woche ebenfalls um einen Tag auf acht Tage erhöht wird. Im Rahmen des Demonstrators wird $x_r = 0.2$ verwendet. Der Verlauf der Gefährdungskennzahl G_r über Periodenlängen bis zu 12 Monaten ist für diese Parameterwahl in Abbildung 5.3 dargestellt. Die Achsenbeschriftung ist aus Gründen der Lesbarkeit in Monaten ausgeführt. Gerechnet wird nach wie vor mit der Periodenlänge in Sekunden. Abbildung 5.4 zeigt den Verlauf für Periodenlängen r bis maximal 24 Stunden, da dieser kritische Bereich in Abbildung 5.3 nicht erkennbar ist. Das relativ niedrige x_r hat beispielsweise für eine Auflösung von 10 Hz eine Bewertung von 98 zur Folge, wobei bei einer Periodenlänge von 24 Stunden die Kennzahl auf nur noch rund 69 Punkte fällt. Somit werden hohe Auflösungen besonders stark bestraft und ein signifikanter Unterschied auch zwischen relativ hohen Auflösungen ist gut abgebildet.

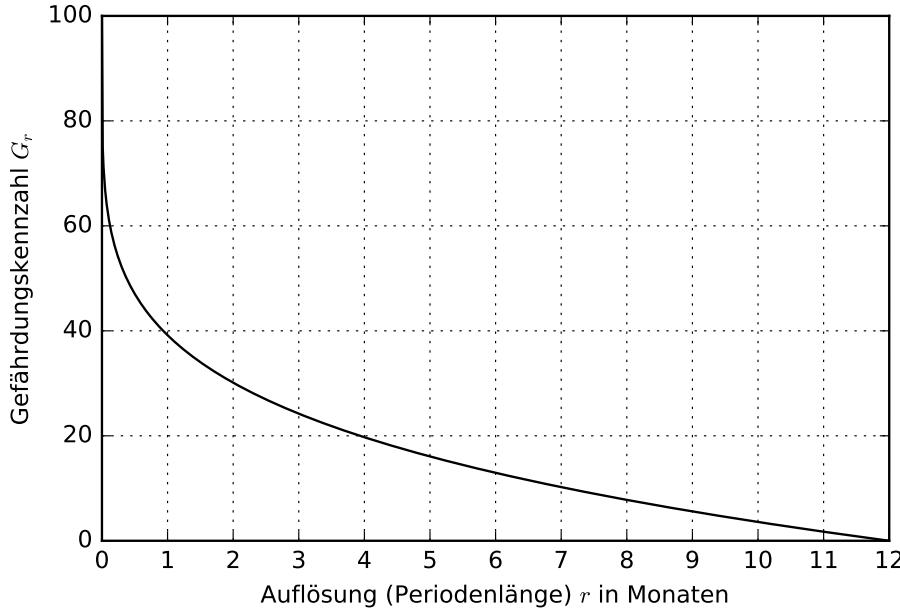


Abbildung 5.3.: Diese Grafik zeigt den Verlauf der Gefährdungskennzahl G_r in Abhängigkeit der zeitlichen Auflösung (Periodenlänge) r im Bereich von null Sekunden bis zu zwölf Monaten. Für die Berechnung der Gefährdungskennzahl G_r wird die Periodenlänge in Sekunden verwendet, die Achsenbeschriftung ist für eine bessere Lesbarkeit jedoch in Monaten ausgeführt. Es kommen die Parameter $x_r = 0.2$ und $r_0 = (365 \cdot 24 \cdot 60 \cdot 60)$ zum Einsatz. Hierdurch ergibt sich ein sehr steiler Gradient im linken Bereich der Kurve, weshalb hohe Auflösungen (kleine Periodenlängen) im Vergleich zu niedrigeren besonders stark benachteiligt werden. Da die Kennzahl auf das Intervall $[0; 100]$ beschränkt ist, erhalten alle Periodenlängen von mehr als zwölf Monaten den Wert 0 (dies ist durch r_0 eingestellt).

In die Kennzahl fließt lediglich die *zeitliche* Auflösung der Energiedatenmesswerte ein. Der Begriff Auflösung bezieht sich somit ausschließlich auf den zeitlichen Abstand zwischen zwei Zeitpunkten, an denen Messwerte erfasst wurden. Für die Bewertung einer Datenanfrage könnte auch die Auflösung der Messungen an sich mit einbezogen werden. Dies erscheint jedoch nur wenig sinnvoll, da es zwischen unterschiedlichen Datenquellen keine großen Unterschiede in der Auflösung gibt. Die Reduktion der Messwertauflösung ist jedoch eine gut geeignete Möglichkeit, um die Aussagekraft von Daten signifikant zu verschlechtern. Daher wird dieses Vorgehen in Abschnitt 5.3.5 besprochen.

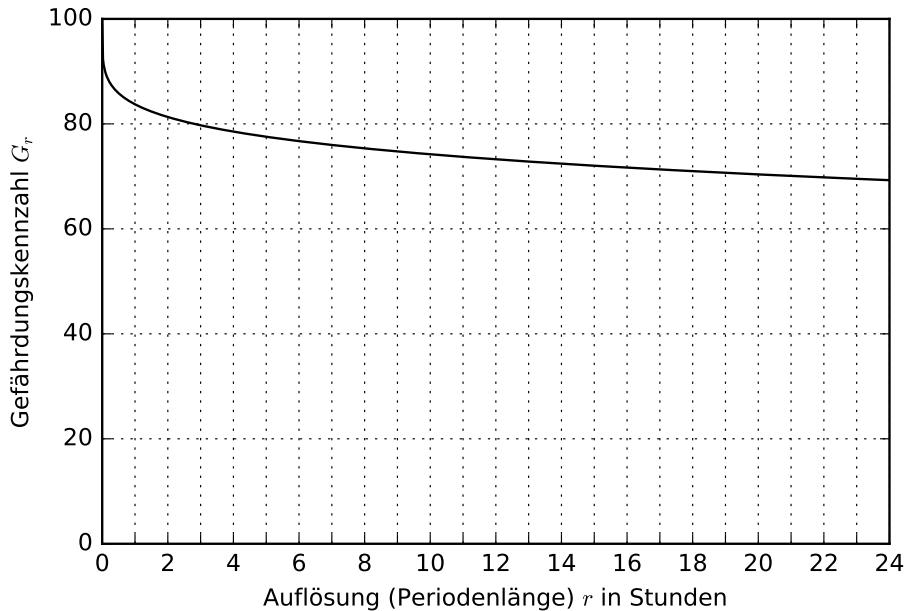


Abbildung 5.4.: Diese Grafik zeigt einen Ausschnitt aus Abbildung 5.3 für hohe Auflösungen (kleine Periodenlängen) bis maximal 24 Stunden. Die Parameter bleiben unverändert: $x_r = 0.2$ und $r_0 = (365 \cdot 24 \cdot 60 \cdot 60)$. Auch hier wurde aus Gründen der Übersichtlichkeit die Achsenbeschriftung nicht in Sekunden, sondern in Stunden ausgeführt. Die Berechnung erfolgt jedoch selbstverständlich nach wie vor mit der Periodenlänge in Sekunden.

5.3.4.2. Abgedeckter Zeitraum

Neben der zeitlichen Auflösung spielt auch der abgedeckte Zeitraum einer Datenanfrage eine wichtige Rolle im Hinblick auf die potenziellen Auswirkungen einer Datenanfrage auf die Privatsphäre der betroffenen Nutzer. Aus einem kürzeren Ausschnitt von Energiedaten lassen sich in der Regel deutlich weniger Informationen ziehen als aus einem längeren.

Einerseits ist es bei größeren Zeiträumen wahrscheinlicher, dass für den Dateninteressenten relevante Geräte tatsächlich genutzt werden. Andererseits lässt die Beobachtung eines längeren Zeitraums auch Rückschlüsse auf die Gewohnheiten und wiederkehrenden Verhaltensweisen der Bewohner zu. Dementsprechend sollte der abgedeckte Zeitraum aus Sicht der von der Energiedatenerfassung betroffenen Personen so klein wie möglich sein und große

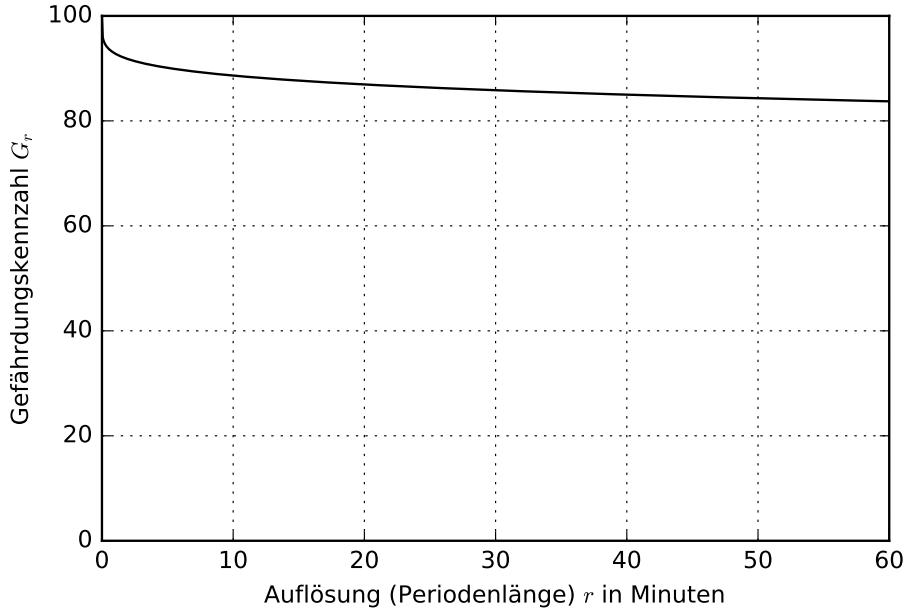


Abbildung 5.5.: Diese Grafik zeigt einen noch kleineren Ausschnitt aus Abbildung 5.3 für sehr hohe Auflösungen (sehr kleine Periodenlängen) bis maximal 60 Minuten. Die Parameter bleiben auch hier unverändert: $x_r = 0.2$ und $r_0 = (365 \cdot 24 \cdot 60 \cdot 60)$. Zu erkennen ist, dass gerade im besonders kritischen Bereich bei Periodenlängen von weniger als einer Minute deutlich zwischen unterschiedlichen Auflösungen differenziert wird.

Zeitspannen sollten dementsprechend negativ bewertet werden. Zum Einsatz kommt deshalb für die Berechnung der Gefährdungskennzahl G_z folgende Formel in Abhängigkeit des angefragten Zeitraums z in Sekunden.

$$G_z = \min \left\{ \left(\frac{z}{z_{100}} \right)^{x_z} \cdot 100; 100 \right\}$$

mit $0 < z, \quad 0 < x_z \leq 1, \quad 0 < z_{100}$

Mithilfe zweier Parameter kann auch bei dieser Gefährdungskennzahl die Bewertung beeinflusst werden. z_{100} legt den abgefragten Zeitraum fest, ab welchem die Gefährdungskennzahl ihr Maximum von $G_z = 100$ erreicht und hält. Größere Zeitspannen führen dann zu keiner weiteren Erhöhung von

G_z . Im Demonstrator wurde z_{100} auf den Wert $(2 \cdot 365 \cdot 24 \cdot 60 \cdot 60)$ gesetzt. Dies entspricht der Anzahl an Sekunden in zwei Jahren, somit erreicht G_z das Maximum von 100, wenn ein Zeitraum von zwei Jahren oder mehr abgefragt wird. Hintergrund dieser Wahl ist, dass ein so großer Zeitraum möglicherweise wiederkehrende, lange Abwesenheitszeiten wie Urlaube oder ähnliches offenlegen kann und nicht zu erwarten ist, dass mit noch größeren Zeiträumen ein signifikant größerer Verlust von Privatsphäre einhergeht. Für den Fall, dass auch bei größeren Zeitspannen noch eine Differenzierung stattfinden soll, müsste z_{100} entsprechend erhöht werden.

Der Parameter x_z stellt analog zur vorigen Gefährdungskennzahl die Empfindlichkeit im Hinblick auf Änderungen des angefragten Zeitraums ein. Mit $x_z = 0.1$ ist diese erneut hoch. Auch hier gilt, dass kleinere x_z einen steileren Gradienten im linken Bereich der Kurve zur Folge haben. Dementsprechend werden Erhöhungen der Zeitspanne bei kurzen Zeiträumen stärker bestraft als bei langen Zeiträumen. Werden beispielsweise Daten für einen Tag angefragt, hat dies eine Kennzahl von $G_z = 51.7$ zur Folge und ein Zeitraum von einer Woche führt zu einer Kennzahl von $G_z = 62.8$. Diese hohe Empfindlichkeit liegt darin begründet, dass in manchen Fällen bereits sehr kurze Zeiträume ausreichen, um die Anzahl an Personen in einem Haushalt oder deren Anziehungsweise Abwesenheitszeiten zu erkennen. Der Verlauf der Bewertungsfunktion wird in Abbildung 5.6 für Zeiträume bis zu zwei Jahren gezeigt. In Abbildung 5.7 ist ein Ausschnitt für kurze Zeiträume bis zu einem Tag abgebildet. Die Grafik zeigt, dass insbesondere zu Beginn eine Vergrößerung der Zeitspannen streng bewertet wird. In beiden Fällen wird mit dem abgefragten Zeitraum in Sekunden gerechnet, auch wenn aus Gründen der Lesbarkeit andere Skalen in den Abbildungen verwendet werden.

5.3.4.3. Alter der Daten

Nach der zeitlichen Auflösung und dem abgedeckten Zeitraum ist das Alter der Energiedaten, welche in einer Datenanfrage angefordert werden, der dritte Aspekt, der bewertet werden soll. In einer Datenanfrage wird von den Dateninteressenten zunächst der *Zeitraum* festgelegt, für welchen die Daten zur Verfügung gestellt werden sollen. Der Zeitraum ist definiert durch einen Start- und einen Endzeitpunkt. Die Differenz zwischen Endzeitpunkt und

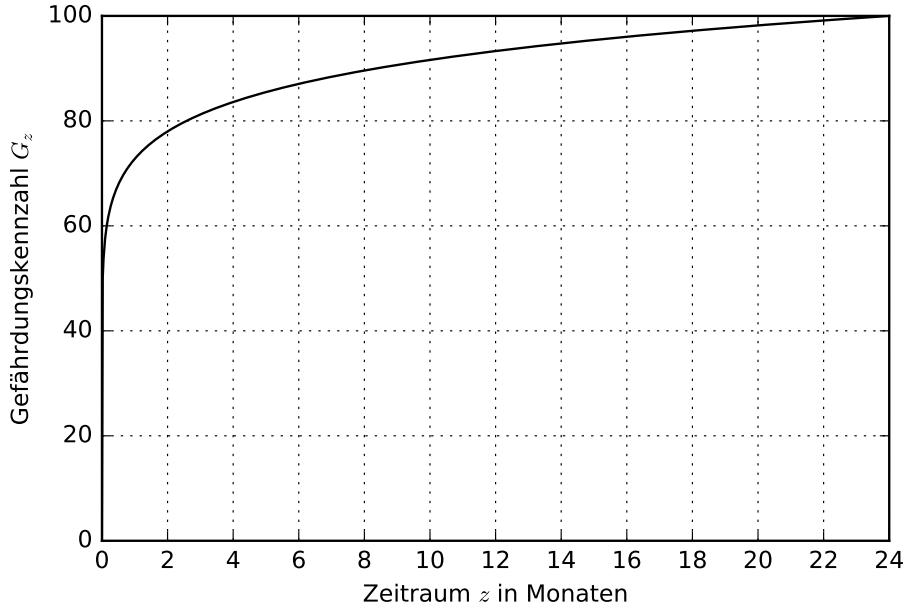


Abbildung 5.6.: Diese Grafik zeigt den Verlauf der Gefährdungskennzahl G_z für angefragte Zeiträume z bis zu 24 Monaten. Für die Berechnung der Kennzahl wurde $x_z = 0.1$ und $z_{100} = (2 \cdot 365 \cdot 24 \cdot 60 \cdot 60)$ gewählt. Dadurch ergibt sich ein extrem steiler Gradient im linken Bereich der Kurve, wodurch eine Vergrößerung der angefragten Zeitspanne frühzeitig bestraft wird. Die Kennzahl ist auf das Intervall $[0; 100]$ beschränkt, alle Zeitspannen von mehr als 24 Monaten erhalten den Wert 100. Im Minimum nimmt G_z den Wert 0 an.

dem Zeitpunkt, an welchem die Datenanfrage erstellt wird, ist das Alter der Daten.

Tendenziell sind aktuellere – jüngere – Daten kritischer als Daten, welche weit in der Vergangenheit liegen: Informationen, die sich aus alten Daten extrahieren lassen, müssen heute nicht mehr unbedingt zutreffen. Möglicherweise werden andere Geräte genutzt oder die Zusammensetzung des Haushalts hat sich geändert. Dementsprechend werden im Rahmen der Gefährdungsbewertung Anfragen, welche ältere Daten betreffen, weniger streng bewertet als solche, welche aktuellere Daten erfragen. Wie bei den vorigen Kennzahlen kommt auch in diesem Fall eine Formel zur Berechnung der Gefährdungskennzahl G_a in Abhängigkeit vom Datenalter a in Sekunden zum Einsatz.

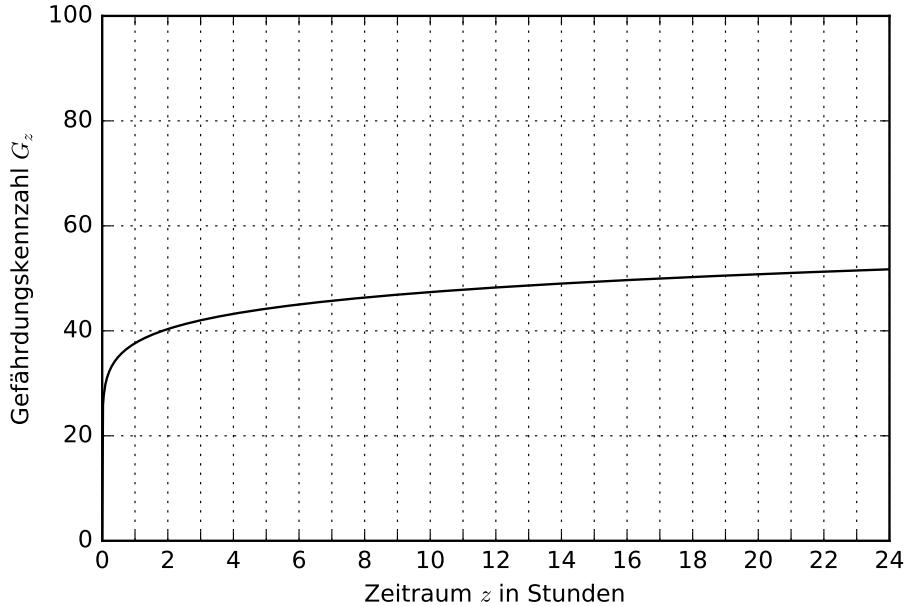


Abbildung 5.7.: Diese Grafik zeigt den Verlauf der Gefährdungskennzahl G_z für einen angefragten Zeitraum bis maximal 24 Stunden. Wie in Abbildung 5.6 betragen auch hier die Parameter $x_z = 0.1$ und $z_{100} = (2 \cdot 365 \cdot 24 \cdot 60 \cdot 60)$. Diese Wahl der Parameter sorgt dafür, dass bereits kurze Zeiträume überproportional streng bewertet werden, da auch kurze Zeitspannen Rückschlüsse beispielsweise auf die Anzahl an Bewohnern in einem Haus oder deren typische An- beziehungsweise Abwesenheitszeiten geben können. In der Grafik äußert sich dies durch einen extremen Gradienten für kurze Zeiträume bis zu einer Stunde.

$$G_a = \max \left\{ 0; 100 - \left(\frac{a}{a_0} \right)^{x_a} \cdot 100 \right\}$$

mit $0 < a, \quad 0 \leq x_a, \quad 0 < a_0$

Auch diese Funktion verfügt über zwei Parameter, mit welchen die Bewertungsfunktion angepasst werden kann. Der Parameter a_0 dient der Anpassung des Alters, ab welchem die Gefährdung mit dem Wert 0 belegt wird. Im Demonstrator wird mit $a_0 = (3 \cdot 365 \cdot 24 \cdot 60 \cdot 60)$ gerechnet, was der Anzahl an Sekunden in drei Jahren entspricht. Dies bedeutet, dass Daten, welche drei Jahre oder älter sind, eine Gefährdungskennzahl für das Datenalter von

$G_a = 0$ erhalten. Hintergrund ist hier, dass ältere Daten vermutlich kaum noch relevant sind in Bezug auf heutiges Verhalten und heutige Gewohnheiten. Sollten dennoch auch ältere Daten noch kritisch bewertet werden, müsste a_0 entsprechend erhöht werden.

Der Parameter x_a wird zur Einstellung der Empfindlichkeit genutzt. Im Demonstrator wird $x_a = 0.6$ verwendet. Dies führt zu einem moderaten Gradienten im Bereich von kleinen Alterswerten. Das Alter der Daten muss in einem deutlich größeren Zeithorizont betrachtet werden als beispielsweise die zeitliche Auflösung. Deshalb führt ein steigendes Alter nur zu einem vergleichsweise langsamem Absinken der Gefährdungskennzahl G_a . Diese beträgt für ein Alter von einer Woche $G_a = 95$. Eine Erhöhung des Alters auf einen Monat senkt den Wert lediglich auf $G_a = 89$. Dies ist insofern sinnvoll, als es sich nach wie vor um (relativ) aktuelle Daten handelt, die gegebenenfalls auch Rückschlüsse auf die Bewohner und ihr Verhalten zulassen, welche noch nicht überholt sind. Werden Daten mit einem Alter von einem Jahr angefragt, beträgt G_a nur noch 48. Der Verlauf der Bewertungsfunktion ist in Abbildung 5.8 für Datenalter bis zu drei Jahren dargestellt. Abbildung 5.9 zeigt einen Ausschnitt für Datenalter bis zu vier Wochen. An dieser Stelle ist gut zu erkennen, dass die Gefährdungskennzahl für das Datenalter nur sehr langsam sinkt, da auch einige Wochen alte Daten noch sehr aktuelle Informationen enthalten können. Auch für diese Abbildungen gilt, dass lediglich aus Gründen der Lesbarkeit das Datenalter in Monaten beziehungsweise Tagen aufgeführt wird. Gerechnet wird jedoch nach wie vor mit dem Alter in Sekunden.

Im Gegensatz zu den vorherigen Kennzahlen ist der Parameter für die Empfindlichkeit x_a nicht nach oben beschränkt, so dass die Funktion für $x_a > 1$ auch eine andere Wölbung annehmen kann. In diesem Fall würde G_a erst für große Alterswerte im Bereich von a_0 signifikant sinken und somit würden auch weniger alte Daten noch kritisch bewertet.

5.3.4.4. Kategorie der Datenquelle

Die vierte und letzte hier betrachtete Gefährdungskennzahl G_k ist abhängig von der Kategorie der angefragten Datenquelle (vgl. Abbildung 5.2). Für die Bestimmung von G_k ist folglich keine Funktion nötig, stattdessen wird eine

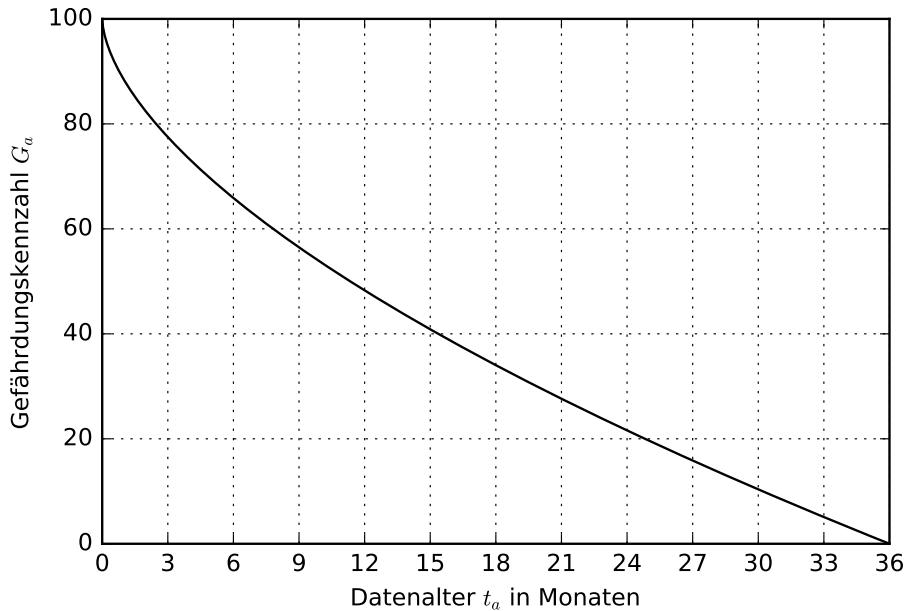


Abbildung 5.8.: Diese Grafik zeigt den Verlauf der Gefährdungskennzahl G_a für Datenalter a bis zu 36 Monaten. Für die Berechnung der Kennzahl wurde $x_a = 0.6$ und $a_0 = (3 \cdot 365 \cdot 24 \cdot 60 \cdot 60)$ gewählt. Dadurch ergibt sich ein fast linearer Verlauf für größere Alterswerte und ein Erreichen des Minimums bei einem Datenalter von drei Jahren.

Tabelle verwendet. In Tabelle 5.1 sind unterschiedliche Gerätekategorien mit ihren jeweiligen Gefährdungskennzahlen aufgelistet.

Besonders kritisch sind Verbraucher wie Herde, Backofen, Mikrowellen oder auch Kaffeemaschinen, da diese in der Regel nur in direkter Interaktion mit einem Bewohner genutzt werden. Es ist unüblich, einen angeschalteten Herd unbeaufsichtigt zu lassen. Somit ist diese Kategorie sehr gut geeignet, um An- beziehungsweise Abwesenheitszeiten zu erkennen. Gleiches gilt für Fernseher, Stereoanlagen und weitere Unterhaltungselektronik, wobei diese zum Teil auch in Abwesenheit der Bewohner im Einsatz sind, beispielsweise zur Abschreckung von Einbrechern. Dennoch lassen auch diese Geräte in der Regel umfangreiche Rückschlüsse auf die Bewohner und ihr Verhalten zu.

Geräte wie Waschmaschinen, Trockner oder Geschirrspülmaschinen können ebenfalls verwendet werden, um Anwesenheitszeiten zu erkennen. Allerdings bieten diese Geräte meist die Möglichkeit einer Startvorwahl, so dass eine

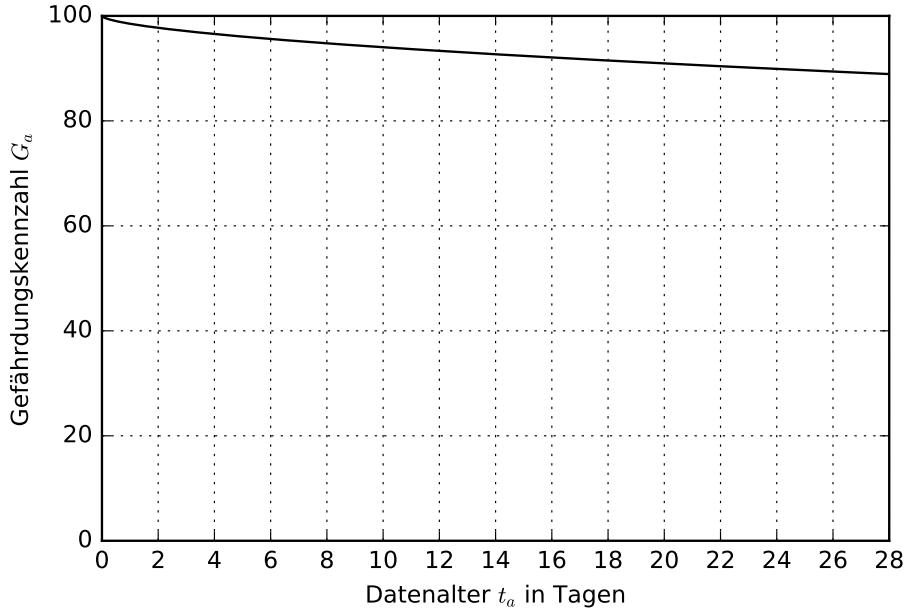


Abbildung 5.9.: Diese Grafik zeigt den Verlauf der Gefährdungskennzahl G_a für Datenalter bis maximal vier Wochen. Wie in Abbildung 5.8 beträgt auch hier der Parameter $x_a = 0.6$ und $a_0 = (3 \cdot 365 \cdot 24 \cdot 60 \cdot 60)$. Diese Parameterwahl sorgt für einen relativ flachen Verlauf der Bewertungsfunktion. Folglich müssen deutlich ältere Daten angefragt werden, um die Gefährdungskennzahl signifikant zu senken.

unmittelbare Anwesenheit nicht erforderlich ist. Mobile Haushaltsgeräte wie beispielsweise Staubsauger oder Bohrmaschinen bedingen jedoch eine Anwesenheit. Das von der Beleuchtung verursachte Lastprofil ist ebenfalls kritisch, da mithilfe dieser Daten beispielsweise die Schlafdauer und -qualität der Bewohner abgeschätzt werden kann. Andererseits sind Lampen oft dauerhaft zur Abschreckung im Einsatz oder werden von Bewegungsmeldern gesteuert.

Die Verbrauchsprofile von Kühlschränken und Tiefkühltruhen lassen nur bedingt Rückschlüsse auf das Verhalten von Bewohnern zu, da sie normalerweise durchgehend in Betrieb sind und damit periodisch Strom verbrauchen sobald der Kompressor betrieben wird. Erhöhter Kühlaufwand, wie er beispielsweise durch das Einlegen noch warmer Lebensmittel entsteht, kann jedoch zum Teil detektiert werden. Ähnliches gilt für Klimageräte, welche entweder permanent oder zeitgesteuert in Betrieb sind. Manche Verbrau-

Tabelle 5.1.: Die Gefährdungskennzahl G_k ist abhängig von der Kategorie der Datenquelle. Die in dieser Tabelle angegebenen Werte sind eine quantitative Wiedergabe einer qualitativen Abschätzung. Je nach Anwendungsfall kann daher eine Anpassung oder Erweiterung notwendig werden.

Kategorie der Datenquelle	Gefährdungskennzahl (G_k)
Herd, Backofen, Mikrowelle, Kaffeemaschine, Wasserkocher, ...	100
Fernseher, Stereoanlage, Unterhaltungselektronik, ...	95
Waschmaschine, Trockner, Geschirrspülmaschine, ...	90
Staubsauger, Bohrmaschine, ...	100
Beleuchtung	80
Kühlschrank, Tiefkühltruhe, ...	50
Klimagerät	50
Umwälzpumpe, Steuergerät für Solarthermie bzw. Heizung, ...	30
Aggrierter Haushaltsverbrauch	75

cher sind fast unabhängig vom Verhalten der Hausbewohner im Einsatz wie beispielsweise Umwälzpumpen oder auch Steuergeräte von Heizungen. Dementsprechend sind solche Daten vergleichsweise unkritisch.

Die aggregierten Verbrauchsdaten eines Haushalts können von sehr unterschiedlichem Nutzen sein. Es kommt stark darauf an, ob charakteristische Geräteprofile gut erkannt werden können oder im Rauschen untergehen. Darauf wird auch solchen Verbrauchsdaten eine hohe Bewertung zugewiesen.

Bei allen Kennzahlen handelt es sich lediglich um eine quantitative Wiedergabe einer qualitativen Abschätzung. Dementsprechend kann es notwendig sein, individuelle Anpassungen an den Gefährdungskennzahlen G_k vorzunehmen. Die Umsetzung im Demonstrator erfolgt dahingehend, dass Objekte vom Typ *Kategorie* (vgl. Abbildung 5.2) ein Attribut für die Gefährdungskennzahl aufweisen und dieses mit den entsprechenden Werten belegt wird. Ein

Administrator ist in der Lage, diese Werte nachträglich anzupassen oder neue *Kategorien* anzulegen. Da diese Möglichkeit auch zum Nachteil der Nutzer missbraucht werden kann, muss es sich bei dem Administrator selbstverständlich um eine vertrauenswürdige Person handeln (vgl. Abschnitt 5.3.1).

5.3.4.5. Reduktion auf eine Kennzahl

Die vier zuvor beschriebenen Gefährdungskennzahlen bezüglich einer Datenanfrage beziehen sich jeweils auf einen einzelnen der folgenden Aspekte: die zeitliche Auflösung (G_r), den abgefragten Zeitraum (G_z), das Alter der Daten (G_a) und die Kategorie (G_k). Mit Werten aus dem Intervall [0; 100] spiegeln sie wider, wie kritisch die Datenanfrage hinsichtlich der bewerteten Eigenschaft ist. Größere Werte entsprechen dabei einer kritischeren Einstufung. Für eine erste Präsentation gegenüber dem *Nutzer* werden die Kennzahlen über eine Gewichtung miteinander verrechnet:

$$G = w_r \cdot G_r + w_z \cdot G_z + w_a \cdot G_a + w_k \cdot G_k$$

mit $0 \leq w_r, w_z, w_a, w_k \leq 1$, $w_r + w_z + w_a + w_k = 1$

Durch die Wahl der Gewichte können einzelne Gefährdungskennzahlen stärker in die endgültige Bewertung G einfließen als andere. Aufgrund der Einschränkungen der Wertebereiche für die Gewichte ergibt sich für die Gesamtbewertung G erneut ein Intervall von [0; 100]. Im Demonstrator werden die Gefährdungskennzahlen für die zeitliche Auflösung (G_r) und für den abgefragten Zeitraum (G_z) stärker bewertet als die beiden übrigen (vgl. Tabelle 5.2).

Die Reduktion auf eine einzelne Kennzahl verschleiert naturgemäß die Details, welche zu dieser endgültigen Bewertung geführt haben. Andererseits verringert sich dadurch auch die Komplexität für die *Nutzer*. Es ist sinnvoll, die Kennzahl für eine erste Präsentation in Form einer Ampel zu codieren. Niedrige Werte führen zu einer grünen Ampel, mittlere Werte zu einer gelben und hohe Werte zu einer roten Ampel. Sollte dem Nutzer die Gesamtbewertung sowie die Farbcodierung nicht ausreichen, stehen auf Wunsch die Einzelkennzahlen zur Verfügung. Darüber hinaus ist denkbar, die Gewichte nicht fest vorzugeben, sondern den Nutzern eine interaktive Auswertung der Bewer-

Tabelle 5.2.: Die Gesamtgefährdungskennzahl G dient der Einschätzung einer Gefährdung für die Privatsphäre, welche durch eine Datenanfrage potenziell entstehen kann. G setzt sich zusammen aus den Gefährdungskennzahlen für einzelne Aspekte einer Datenanfrage, welche mit den angegebenen Gewichten addiert werden.

Gefährdungskennzahl für	Gewichtung
Zeitliche Auflösung (G_r)	$w_r = 0.4$
Abgedeckter Zeitraum (G_z)	$w_z = 0.4$
Alter der Daten (G_a)	$w_a = 0.1$
Kategorie (G_k)	$w_k = 0.1$

tung zu ermöglichen. Hierfür könnte auf der grafischen Benutzeroberfläche eine Reihe von Schieberegbern abgebildet werden, welche es ermöglichen, die Gewichte gemäß den eigenen Präferenzen dynamisch anzupassen.

Darüber hinaus können Warnschwellen für die einzelnen Gefährdungskennzahlen zu einem besseren Verständnis beitragen. Sobald eine der Kennzahlen eine Bewertung von 80 Punkten oder mehr erhält, wird ein kennzahlenspezifischer Warntext angezeigt. Bei einer besonders hohen Auflösung wird beispielsweise darauf hingewiesen, dass diese möglicherweise die Erkennung einzelner Geräte zulässt. Im Falle von besonders großen Zeiträumen werden die Nutzer darauf hingewiesen, dass eine Erkennung ihrer Gewohnheiten und Verhaltensweisen möglich sein könnte. Somit sind die Gefährdungskennzahlen nicht mehr nur abstrakte Bewertungen, sondern werden für die Nutzer greifbar.

Zusätzlich zur Bewertung einer einzelnen Datenanfrage werden zwei weitere Kennzahlen $G_{interessent}$ und G_{alle} berechnet. Diese dienen der Abschätzung der Gefährdung unter Einbeziehen *bereits eingegangener* Datenanfragen. Die Berechnung der Gesamtkennzahlen erfolgt analog zu dem oben beschriebenen Verfahren, mit dem Unterschied, dass historische Anfragen ebenfalls mit einbezogen werden. Hierfür wird jeweils der Durchschnitt der zeitlichen Auflösungen, der Datenalter sowie der Kategoriebewertungen berechnet. Darüber hinaus werden die abgedeckten Zeiträume zusammengelegt und das Ergebnis als Ganzes bewertet.

5.3. Umsetzung des Konzepts in einem Demonstrator

Im Falle von $G_{interessent}$ enthält die Bewertung alle Datenanfragen, welche der Nutzer bereits von dem Dateninteressenten erhalten hat, von welchem die neue Datenanfrage stammt. Dieses Vorgehen soll verhindern, dass ein Dateninteressent eine umfangreiche Datenanfrage in viele kleine Datenanfragen aufspaltet, um so unauffällige Bewertungen zu erhalten. Im Anschluss könnte er – sollte den vielen kleinen Anfragen stattgegeben werden sein – die Daten kombinieren und hätte damit den Schutzmechanismus umgangen. Die rückblickende Kombination und Bewertung der Datenanfragen erschwert ein solches Vorgehen.

Für die Kennzahl G_{alle} werden nicht nur die Datenanfragen des Dateninteressenten, sondern alle bereits eingegangenen Datenanfragen kombiniert. Mit dieser Bewertung soll der Fall abgefangen werden, dass ein Dateninteressent unterschiedliche Identitäten nutzt, um seine Anfragen abzugeben. Dies könnte er mit dem Ziel tun, die – jeweils nicht kritisch bewerteten – Einzelanfragen zu kombinieren und so umfangreiche Einblicke zu erhalten. Die Kennzahlen $G_{interessent}$ und insbesondere G_{alle} weisen den Nachteil auf, dass recht schnell hohe Gefährdungskennzahlen erreicht werden. Dies liegt in der Natur der fast paranoiden Sichtweise, dass alle Informationen, die den Haushalt verlassen, kombiniert werden können. Andererseits handelt es sich bei den Gefährdungskennzahlen auch um pessimistische Abschätzungen des potenziellen Risikos, welches durchaus hoch sein kann.

5.3.4.6. Anmerkungen zur Datenanfragenbewertung

Im Konzept des *Data Custodian Service* ist ein Modul zur Bewertung der Datenanfragen vorgesehen (vgl. Abbildung 5.1). Das Konzept macht jedoch keine Vorgaben bezüglich der Ausprägung eines solchen Moduls und der Vorgehensweise.

Der hier beschriebene und im Demonstrator verwendete Ansatz ist eine Heuristik, welche die potenzielle Gefährdung lediglich aus einer Datenanfrage abschätzt und sich nicht auf die tatsächlich in den Energiedaten enthaltenen Informationen bezieht. Über die beschriebenen Parameter lassen sich die einzelnen Bewertungsfunktionen nach individuellen Präferenzen anpassen und beeinflussen. Darüber hinaus ist es denkbar, die Bewertung zu erweitern, indem beispielsweise eine Community-basierte Einstufung der

Dateninteressenten mit einbezogen wird. Als zuverlässig und seriös bekannte Dateninteressenten könnten so einen Vertrauensvorschuss erhalten und auf eine weniger strenge Bewertung hoffen.

Alternativ wäre es denkbar, einzelne oder eine Kombination der vielen in Abschnitt 4.2.2 vorgestellten Verfahren anzuwenden und die Ergebnisse auszuwerten. Der *Data Custodian Service* würde die eigenen Energiedaten mit Methoden zur Informationsextraktion analysieren, um zu erkennen, welche Aussagen mit den Daten tatsächlich möglich sind. Es besteht allerdings die Gefahr, dass der Informationsgehalt unterschätzt wird, da ein Dateninteressent möglicherweise über leistungsfähigere Werkzeuge zur Analyse verfügt. Dementsprechend sollte auch hier der Fokus auf einer pessimistischen Einschätzung liegen.

Eine Integration dieser und weiterer Vorgehensweisen ist durch den modularen Aufbau des *Data Custodian Service* wenig aufwendig. Neue *Bewertungsmethoden* können durch einen Administrator des *Data Custodian Service* ebenso hinzugefügt werden wie es bereits bei den *Parseern* beziehungsweise *Ausgabeformaten* beschrieben wurde.

5.3.5. Reduktion der Aussagekraft der Daten

Datenanfragen werden, wie im vorigen Abschnitt beschrieben, im Hinblick auf ihre potenziellen Auswirkungen auf die Privatsphäre bewertet. Aus Sicht der Dateninteressenten entsteht ein Konflikt: Einerseits möchten die Dateninteressenten Daten mit möglichst hoher Aussagekraft, andererseits wird dadurch die Datenanfrage durch die Nutzer möglicherweise eher abgelehnt. Die mit einer hohen Aussagekraft der Daten einhergehende hohe Gefährdungskennzahl kann die Nutzer nämlich dazu bewegen, die Datenanfrage nicht zu genehmigen. Dementsprechend ist es für Dateninteressenten durchaus sinnvoll, eine möglichst kleine Gefährdungsbewertung zu erreichen, wobei der Informationsgehalt der Daten natürlich noch ausreichen muss, um die von den Interessenten gewünschten Analysen mit den Daten durchführen zu können.

Eine Reduktion der Aussagekraft ist möglich, indem beispielsweise nur kürzere Zeitspannen, ältere Energiedaten oder weniger kritische Kategorien abgefragt werden. Diese Aspekte haben eine direkte Auswirkung auf die Gefährdungskennzahl. Andererseits ist eine solche Einschränkung zum Teil

nicht mit den gewünschten Analysen in Einklang zu bringen, so dass die Aussagekraft auf andere Weise reduziert werden muss. Einige mögliche Methoden zur Reduktion der Aussagekraft beziehungsweise zur Reduktion der Datenqualität werden im Folgenden vorgestellt.

5.3.5.1. Reduktion der zeitlichen Auflösung

Eine naheliegende Möglichkeit zur Reduktion der Aussagekraft ist die Verringerung der zeitlichen Auflösung. Abhängig von ihrer Funktionsweise erfordern viele Verfahren zur Informationsextraktion aus Energiedaten eine Mindestauflösung, um zuverlässige Aussagen zu ermöglichen, da sie beispielsweise auf einer Kantenerkennung beruhen. Eine schlechtere zeitliche Auflösung verringert daher meist direkt die Ergebnisqualität oder verhindert sogar die Anwendung mancher Verfahren gänzlich (vgl. Abschnitt 4.2.2.3). Im Rahmen einer Datenanfrage muss der Dateninteressent die gewünschte zeitliche Auflösung angeben. Seine Wahl sollte einen Kompromiss zwischen für ihn benötigter Auflösung und möglichst geringer Gefährdungskennzahl darstellen: Ist die Datenqualität zu schlecht, sind die Daten für den Interessenten nicht von Nutzen; ist die Gefährdungskennzahl zu hoch, lehnt der Nutzer gegebenenfalls die Anfrage ab. Zur Wahl stehen den Dateninteressenten für eine Datenanfrage die Werte aus Tabelle 5.3. Für den Fall, dass so hohe Auflösungen in den Rohdaten nicht (durchgehend) zur Verfügung stehen, erfolgt eine Interpolation der Werte.

Die Abstufung der verfügbaren zeitlichen Auflösungen beziehungsweise Periodenlängen wurde derart gewählt, dass ein ganzzahliges Vielfaches kleinerer Periodenlängen – bis auf einige Ausnahmen – eine größere Periodenlänge ergibt. So lassen sich beispielsweise alle Auflösungen im Bereich von Sekunden durch Multiplikation mit einer ganzen Zahl in eine Periodenlänge von einer Minute überführen. Für diese Größenordnung gilt wiederum dasselbe in Bezug auf eine Stunde. Einen ersten deutlicheren Bruch gibt es bei dem Übergang von Tagen auf Wochen, da beispielsweise zwei beziehungsweise vier Tage nicht ganzzahlig in eine Woche überführt werden können.

Der Grund für die Wahl dieser Abstufungen liegt in der gewünschten gleichbleibenden Aufteilung der Zeiträume in gleich große Abschnitte. Die Aufteilung in einzelne Blöcke erfolgt stets vom Beginn eines Jahres ausgehend.

Kapitel 5. Konzept und Umsetzung

Tabelle 5.3.: Für eine Datenanfrage steht den Dateninteressenten eine Reihe von verschiedenen zeitlichen Auflösungen in Form von unterschiedlichen Periodenlängen zur Verfügung. Sie sind nach verschiedenen Größenordnungen gegliedert unten aufgeführt. Die Abstufung erfolgt derart, dass größere Periodenlängen nach Möglichkeit ein ganzzahliges Vielfaches von kleineren Periodenlängen darstellen.

Größenordnung	Ausprägung
Millisekunden	1, 2, 5, 10, 50, 100, 500
Sekunden	1, 2, 5, 10, 15, 30
Minuten	1, 2, 5, 10, 15, 30
Stunden	1, 2, 4, 6, 12
Tags	1, 2, 4
Wochen	1, 2, 4, 8, 12, 26
Jahre	1, 2, 3, 4

So starten die Blöcke der fünfsekündigen Periodenlänge beispielsweise jeweils an den Sekunden null, fünf, zehn und so weiter. Wird die Auflösung auf zehnsekündige Periodenlängen verringert, starten die Blöcke bei den Sekunden null, zehn, zwanzig und so weiter, das heißt, sie setzen sich aus Blöcken der vorhergehenden Auflösung zusammen.

Eine derartige Gestaltung der einzelnen Zeitblöcke hat zwei Vorteile. Zunächst ermöglicht dies eine effiziente Vorberechnung verschiedener Auflösungen, da diese aufeinander aufbauen und so beispielsweise durch Mittelwertbildung der Gruppen eine Überführung in gröbere Auflösungen möglich ist. Darüber hinaus verhindert die Festlegung auf feste Startpunkte für die jeweiligen Auflösungen einen unbeabsichtigten Informationsverlust: Ein Dateninteressent könnte andernfalls versuchen, durch zeitliche Verschiebung der Blöcke zueinander eine höhere Abtastrate zu erreichen. Falls ein Dateninteressent beispielsweise fünf Datenanfragen mit einer Auflösung von fünf Sekunden, welche jeweils zu den Sekunden null, eins, zwei, drei und vier beginnen, abgäbe, wäre dies der Fall. Die erste Anfrage würde die Blöcke von den Sekunden null bis vier, fünf bis neun und so weiter abdecken. Die zweite jedoch die Blöcke von den Sekunden eins bis fünf, sechs bis zehn und so weiter.

Eine Differenzenbildung der einzelnen Datenanfragen offenbart dann unter anderem Lastspitzen, die andernfalls nicht sichtbar gewesen wären. Dies wird durch die feste Vorgabe der Startzeitpunkte der einzelnen Blöcke verhindert.

Über die zuvor genannten Aspekte hinaus ist ein weiteres Detail zu beachten: Die zeitliche Auflösung in Form der Periodenlänge darf nicht größer sein als der angefragte Zeitraum. Auch dies würde potenziell ein Umgehen des Schutzsystems ermöglichen. Würde ein Dateninteressent beispielsweise viele, nur wenige Sekunden große Zeitspannen jeweils mit einer zeitlichen Auflösung von einem Jahr anfordern, so erhielten diese Anfragen alle geringe Bewertungen. Schließlich ist die Zeitspanne sehr klein und die Auflösung sehr grob. Zusammengesetzt ergäben die Datenanfragen jedoch eine Zeitreihe mit einer Auflösung im Bereich weniger Sekunden. Dementsprechend wird auch dieser Fall verhindert, indem sichergestellt wird, dass die Periodenlänge den angefragten Zeitraum nicht übersteigt.

5.3.5.2. Hinzugefügtes Rauschen

Die Qualität der Energiedaten kann durch das Hinzufügen von Rauschen künstlich verschlechtert werden. Für Verfahren, welche auf der Erkennung von Kanten im Profilverlauf beruhen, kann das Rauschen beispielsweise das Vorhandensein von Kanten verschleiern oder nicht vorhandene Kanten erscheinen lassen. Aber auch beim Erkennen von Anwesenheitszeiten kann künstliches Rauschen hinderlich sein, da die Grundlast beziehungsweise Abweichungen davon schwerer zu identifizieren sind.

Die technische Umsetzung des Verrauschens erfolgt dergestalt, dass auf die Werte der ursprünglichen Energiedatenzeitreihe zufällige Werte addiert werden. Die Wahl dieser Werte erfolgt in Abhängigkeit der durch den Dateninteressenten gewählten Stärke des Verrauschens. Die verschiedenen verfügbaren Stufen sind in Tabelle 5.4 aufgeführt.

Die absoluten Intervallgrenzen werden auf Basis des Durchschnittswerts der gesamten angefragten Zeitreihe berechnet. Falls beispielsweise ein Lastprofil durchschnittlich 120 W aufweist und der Dateninteressent ein Verrauschen der Stufe 3 wählt, werden auf die jeweiligen Werte der Zeitreihe zufällige, gleichverteilte Zahlen aus dem Intervall $[-24 \text{ W}; 24 \text{ W}]$ addiert. Darüber hinaus wird die Gefährdungsbewertung der Datenanfrage um den angegebenen

Tabelle 5.4.: Ein Dateninteressent kann aus einer Reihe verschiedener Stärken künstlichen Rauschens wählen, um die Datenqualität und damit die Gefährdungskennzahl zu senken. Jede Stufe entspricht einem Intervall, aus welchem zufällige, gleichverteilte Werte bestimmt werden, welche wiederum auf die ursprünglichen Daten der Zeitreihe addiert werden. Als Basis für die relativen Werte dient der Durchschnittswert der betrachteten Zeitreihe. Jeder Stufe ist darüber hinaus ein Reduktionsfaktor zugeordnet, welcher angibt, wie stark die Gefährdungskennzahl sinkt.

Stärke des Verrauschens	Intervall in Prozent	Reduktionsfaktor
1	[-5; 5]	0.95
2	[-10; 10]	0.9
3	[-20; 20]	0.8
4	[-50; 50]	0.7
5	[-100; 100]	0.6
6	[-200; 200]	0.5

Reduktionsfaktor – im Beispiel 0.8 – gesenkt. Logischerweise bedingt ein höheres Rauschen eine schlechtere Datenqualität, weshalb die Reduktionsfaktoren mit steigendem Rauschen für eine sinkende Gefährdungsbewertung sorgen.

Für den Nutzer ergibt sich der Vorteil, dass die an den Dateninteressenten weitergegebenen Informationen durch das künstliche Rauschen verschleiert werden. Für den Dateninteressenten ist dies offensichtlich ein Nachteil. Allerdings hat ein Interessent so die Möglichkeit, die Gefährdungskennzahl zu senken, so dass ein Nutzer die Datenanfrage nicht mehr ablehnt, sondern eventuell annimmt.

5.3.5.3. Reduktion der Messwertauflösung

Die letzte hier vorgestellte Variante zur Reduktion der Aussagekraft von Daten betrifft die Auflösung der Messwerte. Gemeint ist an dieser Stelle nicht die zeitliche Auflösung in Form der Periodenlänge, sondern die Auflösung der Messdaten an sich. Je nach verwendetem Messgerät bewegt sich die Auflösung beispielsweise in der Größenordnung von einer Wattstunde für Energiemessgeräte beziehungsweise einem Watt für Leistungsmessgeräte. Die

5.3. Umsetzung des Konzepts in einem Demonstrator

zur Reduktion der Datenqualität angestrebte Reduktion ist jedoch deutlich drastischer.

Die einzelnen Datenpunkte der Zeitreihe werden anhand des Durchschnittswertes der betrachteten Zeitreihe beurteilt. Messwerte, die so groß wie oder größer als der Mittelwert sind, gelten als *hoch* und Messwerte, welche darunterliegen, gelten als *niedrig*. Die zuvor fein aufgelösten Messwerte werden somit in lediglich zwei Kategorien aufgeteilt, abhängig davon, ob sie unter dem als Durchschnittswert der betrachteten Zeitreihe definierten Schwellwert liegen oder nicht. Somit ist nach Anwendung dieses Verfahrens nicht mehr erkennbar, welche absoluten Werte der Zeitreihe zugrunde lagen. Dementsprechend geht mit der Verwendung des Schwellwertverfahrens auch eine deutliche Verringerung der Gesamtgefährdungskennzahl einher und der Reduktionsfaktor beträgt folglich 0.5.

5.3.5.4. Anmerkungen zur Reduktion der Aussagekraft

In den vorhergehenden Abschnitten wurden drei Möglichkeiten zur Reduktion der Aussagekraft der Daten (Reduktion der zeitlichen Auflösung, hinzugefügtes Rauschen, Reduktion der Messwertauflösung) vorgestellt. Ein Dateninteressent muss im Rahmen einer Datenanfrage allerdings lediglich die zeitliche Auflösung festlegen. Weitere Maßnahmen zur Reduktion der Datenqualität sind optional. Der *Data Custodian Service* nutzt die Angabe zur zeitlichen Auflösung, um die gespeicherten Energiedaten in die gewünschte Periodenlänge umzurechnen. Anschließend wird die Datenqualität, falls vom Dateninteressenten gewünscht, mit den gewählten Verfahren weiter reduziert.

Die beiden vorgestellten Maßnahmen – Hinzufügen von Rauschen beziehungsweise Reduktion der Messwertauflösung – können gegebenenfalls vom Dateninteressenten gleichzeitig ausgewählt werden. Dies ist jedoch nicht erforderlich. Der *Data Custodian Service* bietet den Dateninteressenten lediglich die Möglichkeit, die Datenqualität weiter zu senken, um so die Gefährdungskennzahl zu verkleinern und damit die Chance zu erhöhen, dass ein Nutzer der Datenanfrage zustimmt.

Da nicht bekannt ist, mit welchen Verfahren ein Dateninteressent die angefragten Energiedaten auswerten wird, kann die Wirksamkeit der einzelnen Maßnahmen durch den *Data Custodian Service* nur abgeschätzt werden. Analog

zur Bewertung der Datenanfragen (vgl. Abschnitt 5.3.4) handelt es sich somit um einen heuristischen Ansatz, welcher abhängig vom Anwendungsfall erweitert oder angepasst werden kann. Möglicherweise verlieren die vorgeschlagenen Verfahren auch mit der Einführung neuer Auswertungsmechanismen an Wirkung, so dass ein Anpassen der Reduktionsfaktoren nötig wird.

Wie im Entity-Relationship-Diagramm in Abbildung 5.2 dargestellt, ist eine *Qualitätsreduktionsmaßnahme* eine gewöhnliche Entität des Datenmodells. Dementsprechend können dem *Data Custodian Service* flexibel neue Maßnahmen zur Qualitätsreduktion hinzugefügt werden. Der Prozess verläuft analog wie bereits zuvor bei den *Bewertungsmethoden* beschrieben: Ein Administrator erzeugt ein neues Objekt und gibt in diesem einen Verweis auf die *Python-Klasse* mit der entsprechenden Funktionalität an. Sobald die neue Maßnahme im Rahmen einer Datenanfrage ausgewählt wird, nutzt der *Data Custodian Anwendungskern* die Funktionalitäten des neuen Objekts, um die Energiedatenqualität zu senken.

5.3.6. Logbuchfunktion

Alle Informationen über die Energiedaten eines Nutzers als auch über die dazu gehörenden Datenanfragen werden gemäß dem Datenmodell aus Abbildung 5.2 in einer relationalen Datenbank abgelegt. Hierdurch kann der *Data Custodian Service* auch eine Logbuchfunktion erfüllen.

Der *Data Custodian Service* präsentiert den Nutzern auf Abruf eine Übersicht der jeweils von ihnen stammenden Energiedaten sowie der sie betreffenden Datenanfragen. Dies umfasst nicht nur die angenommenen Anfragen, sondern auch die abgelehnten. Der *Data Custodian Service* schafft somit einen Überblick über alle, die Energiedaten betreffenden, Aktionen. Sollte im Nachhinein ein Missbrauch der Energiedaten festgestellt werden, hat ein Nutzer zumindest die Möglichkeit, festzustellen, welcher Dateninteressent die jeweiligen Daten angefragt hat und ist so gegebenenfalls in der Lage, Sanktionen einzuleiten. Darauf hinaus werden historische Datenanfragen transparent in die Beurteilung aktueller Datenanfragen einbezogen, so dass ein Umgehen der Schutzmechanismen des *Data Custodian Service* erschwert wird (vgl. Abschnitt 5.3.4).

In seiner Funktion als Katalog für Energiedaten ermöglicht es der *Data Custodian Service* den Nutzern, alle sie betreffenden Energiedaten im Blick zu haben. Somit können Nutzer stets nachvollziehen, welche Energiedaten erfasst und vorgehalten werden. Darüber hinaus ergibt sich dadurch die Möglichkeit, Daten gezielt aus dem System entfernen zu lassen.

5.3.7. Löschung historischer Daten

Ein Nutzer kann mithilfe des *Data Custodian Service* einen Überblick über seine Energiedaten erlangen. Falls ein Nutzer entscheiden sollte, dass bestimmte Energiedaten nicht mehr benötigt werden, kann er sie aus dem System löschen lassen. Eine solche Löschung ist zu jeder Zeit möglich. Für Dateninteressenten bedeutet dies jedoch, dass sie die freigegebenen Daten aus genehmigten Datenanfragen selbstständig verwalten und vorhalten müssen, da ein Nutzer jederzeit entscheiden kann, diese aus dem System löschen zu lassen und somit ein nachträglicher Zugriff nicht garantiert werden kann.

Eine angestoßene Löschung umfasst lediglich die Energiedaten in Form der Zeitreihen, nicht aber die dazu gehörenden Metadaten. Diese Abgrenzung ist erforderlich, um die Logbuchfunktion des *Data Custodian Service* aufrecht zu erhalten und um entsprechende Datenanfragen auch weiterhin korrekt zuordnen zu können. Auf diese Weise bleibt für die Nutzer nach wie vor nachvollziehbar, welche Daten angefragt und gegebenenfalls weitergegeben wurden, ohne dass die eigentlichen Zeitreihendaten weiter vorgehalten werden müssten.

Es ist den Nutzern allerdings nicht ohne Weiteres möglich, eine Verwendung bereits weitergegebener Daten zu untersagen beziehungsweise gar eine Löschung auf Seiten der Dateninteressenten zu erzwingen. Sollte tatsächlich eine Nutzung der Daten auch nachträglich technisch verhindert werden können, müsste ein System für digitales Rechtemanagement wie beispielsweise *Usage Control* zum Einsatz kommen (vgl. Abschnitt 3.2.2.7). Andernfalls sind einmal weitergegebene Daten unwiderruflich außerhalb der Kontrolle der Nutzer – zumindest aus technischer Sicht.

Dementsprechend sollten auch – wie bereits zuvor beschrieben – nur Daten der geringstmöglichen Aussagekraft weitergegeben werden, um den Schaden einer missbräuchlichen Verwendung der Daten zu minimieren. Dies gilt selbst

für den Fall, dass eine Löschung vertraglich zugesichert wird, da es sich hierbei lediglich um eine regulatorische Maßnahme handelt (vgl. Abschnitt 3.2.1).

Abgesehen von manuell anzustoßenden Löschungen unterstützt der *Data Custodian Service* auch regelmäßige Prüfungen für den Nutzer, ob eine Entfernung mancher Daten sinnvoll ist. Zu diesem Zweck wird der Datenbestand eines Nutzers auf für eine lange Zeit – im Normalfall zwei Jahre – nicht angefragte Energiedaten überprüft. Für den Fall, dass solche Energiedaten vorhanden sind, wird dem Nutzer vorgeschlagen, diese Daten zu löschen. Auch hier erfolgt selbstverständlich lediglich die Löschung der Zeitreihendaten, nicht aber der Metadaten.

5.3.8. Zusammenfassung Demonstrator

Teile des modularen Konzepts des *Data Custodian Service* für ein nutzerorientiertes Energiedatenmanagementsystem wurden im Rahmen dieser Arbeit implementiert und in einem Demonstrator umgesetzt. Zum Einsatz kommen dabei ausschließlich quelloffene Werkzeuge und Bibliotheken, welche die einzelnen Komponenten bei der Erbringung ihrer jeweiligen Funktionalitäten unterstützen. Dies sind unter anderem *Pandas* für die Verarbeitung von Zeitreihen, *Django* als Framework für Webanwendungen, *SQLite* als relationale Datenbank und *HDF5* als Speichertechnologie für die Speicherung von Zeitreihen. Abbildung 5.10 zeigt, wo die einzelnen Werkzeuge verwendet werden. An dieser Stelle sei erneut betont, dass es sich bei dem Demonstrator um eine *Proof-of-Concept*-Implementierung handelt, welche nicht für den Produktiveinsatz vorgesehen ist. Es ist jedoch durchaus denkbar, unter Verwendung der vorgestellten Werkzeuge ein für den Produktiveinsatz geeignetes Produkt auf Basis des Demonstrators zu erstellen.

Der implementierte Demonstrator nutzt eine Reihe von Bewertungsfunktionen, um eingehende Datenanfragen hinsichtlich ihrer potenziellen Auswirkungen auf die Privatsphäre der Nutzer zu beurteilen. Einfluss auf die Gefährdungsbewertung haben die zeitliche Auflösung, der angefragte Zeitraum, das Alter der Daten und die Kategorie, aus welcher die Energiedaten stammen. Dateninteressenten haben die Möglichkeit, aus verschiedenen Maßnahmen zu wählen, um die Aussagekraft der angefragten Daten zu reduzieren.

5.3. Umsetzung des Konzepts in einem Demonstrator

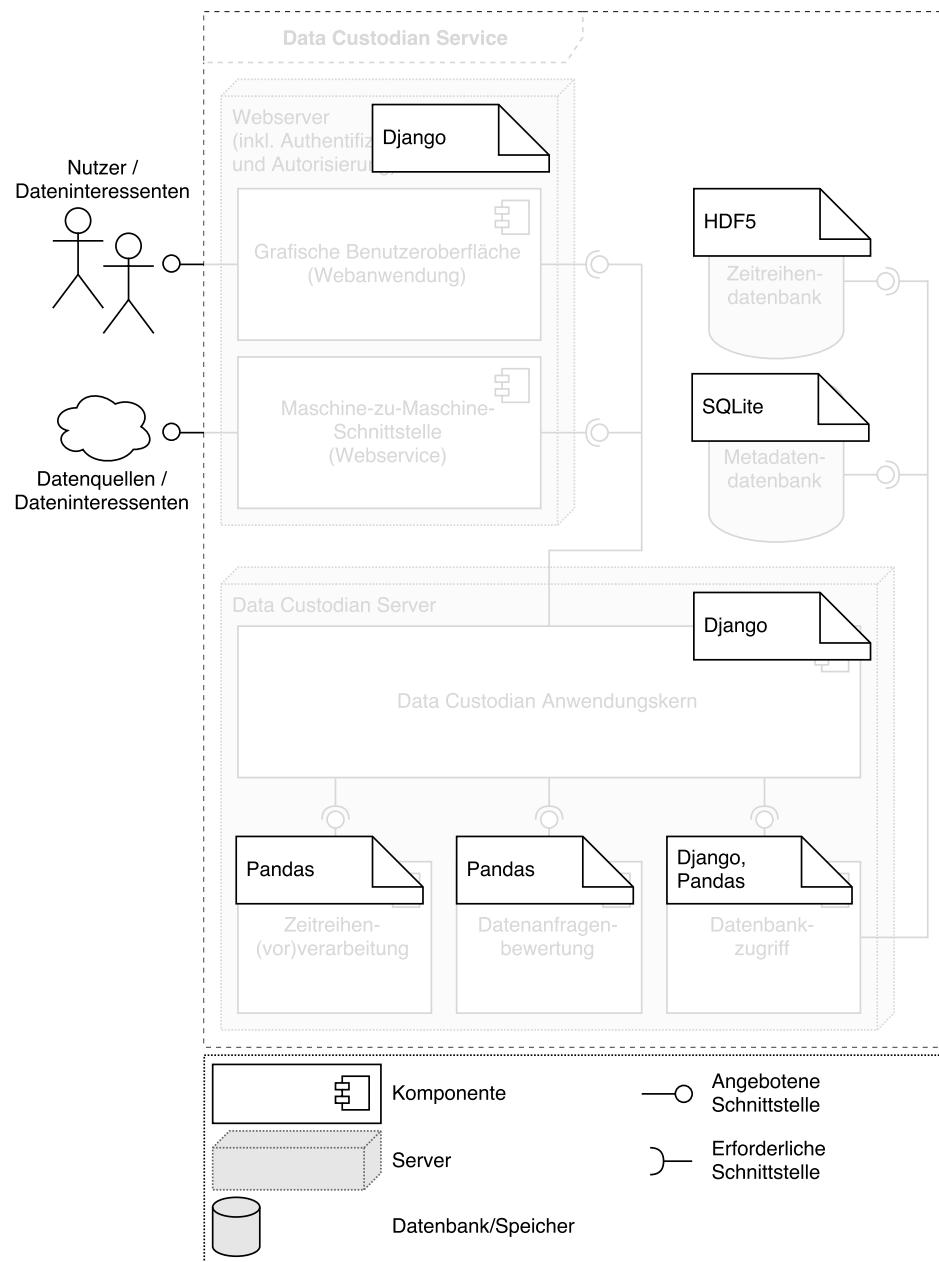


Abbildung 5.10.: Für die Implementierung des Data Custodian Service wird eine Reihe von quelloffenen Bibliotheken und Werkzeugen für die Realisierung einzelner Komponenten genutzt. Dies sind unter anderem Pandas für die Verarbeitung von Zeitreihen, Django als Framework für Webanwendungen, SQLite als relationale Datenbank und HDF5 als Spechertechnologie für die Speicherung von Zeitreihen. Diese Abbildung zeigt, an welcher Stelle das jeweilige Werkzeug Anwendung findet.

Dies resultiert im Gegenzug in einer niedrigeren Bewertung der Gefährdung, so dass Nutzer womöglich eher geneigt sind, der Datenanfrage zuzustimmen.

Des Weiteren präsentiert der Demonstrator seinen Nutzern einen Überblick über deren Energiedaten sowie über alle eingegangenen Datenanfragen. Mit Hilfe dieser Übersichten können die Nutzer zu jedem Zeitpunkt erkennen, welche sie betreffenden Energiedaten dem System vorliegen, welche Interessenten ihre Daten angefragt haben und wie über diese Anfragen entschieden wurde.

5.4. Typische Arbeitsabläufe

In diesem Abschnitt werden ausgewählte Arbeitsabläufe vorgestellt, wie sie typischerweise im *Data Custodian Service* beziehungsweise im Demonstrator vorkommen können. Die Betrachtung erfolgt dabei sowohl aus der Perspektive der Nutzer als auch aus Sicht der Dateninteressenten. Anhand des untenstehenden Beispiels sollen die ausgewählten Arbeitsabläufe dargestellt werden.

Bei den Abbildungen im Folgenden handelt es sich um Screenshots der grafischen Benutzeroberfläche des Demonstrators wie sie in einem aktuellen Desktop-Browser dargestellt wird. Da es sich dabei um eine auf *HTML5* und *CSS3* basierende Webseite handelt, kann der Demonstrator auch auf mobilen Endgeräten wie Smartphones genutzt werden, jedoch würde sich die Darstellung aufgrund der geringeren Auflösung in der Regel unterscheiden.

Beispiel Ein Nutzer mit Namen *Nutzer1* stellt über das Energiedatenmanagementsystem einige Energiedaten bereit. Teile dieser Energiedaten werden von einem Dateninteressenten mit Namen *Dateninteressent1* angefragt. *Nutzer1* verwendet die vom Demonstrator durchgeführte Bewertung, um zu entscheiden, ob den Datenanfragen stattgegeben werden soll oder nicht. Schließlich kann *Dateninteressent1* die entsprechenden Daten herunterladen und verwenden.

5.4.1. Verfügbare Aktionen

Für die Verwendung des Demonstrators ist ein Benutzerkonto erforderlich, da nicht authentifizierten Personen der Zugriff verweigert wird. Es werden drei Benutzerrollen unterschieden: Nutzer, Dateninteressenten und Administratoren. Abhängig von der jeweiligen Rolle stehen den Verwendern unterschiedliche Aktionen zur Verfügung und wie in Abbildung 5.11 zu sehen ist, können Nutzer zunächst aus drei Aktionen wählen:

- Verwaltung der Energiedaten
- Bereitstellung neuer Energiedaten
- Verwaltung eingegangener Datenanfragen

Diese Aktionen werden in den kommenden Unterabschnitten detaillierter erläutert. Den Dateninteressenten stehen, wie in Abbildung 5.12 zu sehen ist, hingegen nur zwei Aktionen zur Verfügung:

- Erstellung neuer Datenanfragen
- Verwaltung eingereichter Datenanfragen

5.4.2. Verwaltung der eigenen Energiedaten

Den Nutzern soll es zu jedem Zeitpunkt möglich sein, zu überblicken, welche Energiedaten sie dem Energiedatenmanagementsystem zur Verwaltung übergeben haben. Der Demonstrator präsentiert hierzu, wie in Abbildung 5.13 zu sehen, eine Liste mit grundlegenden Informationen zu den Energiedaten des angemeldeten Nutzers. Daten anderer Nutzer sind in dieser Ansicht selbstverständlich nicht zu sehen. Ebenso wenig ist diese Ansicht den Dateninteressenten zugänglich.

Die einzelnen Einträge der Liste sind mit der jeweiligen Energiedatendetailansicht verknüpft, so dass Nutzer weitere Informationen zu den hinterlegten Daten erhalten können. Wie in Abbildung 5.14 zu sehen, enthält die Detailansicht eine Reihe technischer Informationen zu den Energiedaten. Des Weiteren

Kapitel 5. Konzept und Umsetzung

The screenshot shows the Data Custodian Service homepage. At the top, there is a header with the text "Data Custodian Service" on the left and "LOGOUT" on the right. The main title "Aktion wählen" is centered above a welcome message "Herzlich willkommen, Nutzer1!". Below this, a text block says: "Als Nutzer können Sie hier Ihre Energiedaten verwalten, neue Energiedaten zur Verfügung stellen oder eingegangene Datenanfragen betrachten." At the bottom, there are three buttons: "MEINE ENERGIEDATEN VERWALTEN" (blue), "NEUE ENERGIEDATEN HOCHLAGEN" (blue), and "DATENANFRAGEN VERWALTEN" (blue). Each button has a corresponding text description below it.

Abbildung 5.11.: Nutzern des Demonstrators stehen nach einer erfolgreichen Authentifizierung drei verschiedene Aktionen auf der Startseite zur Verfügung: Verwaltung der Energiedaten, Hinzufügen neuer Energiedaten und Verwaltung eingegangener Datenanfragen. Dateninteressenten können auf diese Funktionen nicht zugreifen.

The screenshot shows the Data Custodian Service homepage for a data requester. At the top, there is a header with the text "Data Custodian Service" on the left and "LOGOUT" on the right. The main title "Aktion wählen" is centered above a welcome message "Herzlich willkommen, Dateninteressent1!". Below this, a text block says: "Als Dateninteressent können Sie hier neue Datenanfragen aufgeben sowie ihre eingereichten Datenanfragen betrachten." At the bottom, there are two buttons: "NEUE DATENANFRAGE AUFGEBEN" (red) and "DATENANFRAGEN VERWALTEN" (red). Each button has a corresponding text description below it.

Abbildung 5.12.: Dateninteressenten können nach einer erfolgreichen Anmeldung aus zwei verschiedenen Aktionen wählen: Erstellung einer neuen Datenanfrage, Verwaltung existierender Datenanfragen. Für Nutzer ist ein Zugriff auf diese Funktionen nicht möglich.

5.4. Typische Arbeitsabläufe

Datensatz 1: Hochgeladen von Ihnen am 2. November 2016 mit Kategorie(n) Waschmaschine

Datensatz 2: Hochgeladen von Ihnen am 2. November 2016 mit Kategorie(n) Kühlschrank

Datensatz 3: Hochgeladen von Ihnen am 2. November 2016 mit Kategorie(n) Trockner

Datensatz 4: Hochgeladen von Ihnen am 2. November 2016 mit Kategorie(n) Geschirrspülmaschine

Abbildung 5.13.: Der Demonstrator präsentiert seinen Nutzern eine Übersicht über alle Energiedaten, welche ein Nutzer hochgeladen hat. Über die entsprechenden Listeneinträge können Detailansichten aufgerufen werden.

werden alle Datenanfragen aufgelistet, welche die betrachteten Energiedaten betreffen – im Beispiel existieren bisher keine solchen Anfragen. Zuletzt wird dem Nutzer noch die Möglichkeit geboten, die Energiedaten zu löschen, wobei wie bereits zuvor beschrieben, nur eine Löschung der Zeitreihendaten, nicht aber der Metadaten vorgesehen ist.

5.4.3. Bereitstellung neuer Energiedaten

Die zweite Hauptfunktion, die der Demonstrator für Nutzer bietet, ist ein Formular für das Bereitstellen neuer Energiedaten. Nachdem die entsprechende Funktion (vgl. Abbildung 5.11) ausgewählt wurde, wird dem Nutzer ein Webformular präsentiert. Mithilfe der Eingabemaske müssen die Datei mit den Rohdaten, die entsprechenden Datenkategorien, die zugrundeliegende Einheit und der passende Parser ausgewählt werden. Abbildung 5.15 zeigt die verwendete Webseite mit den auszufüllenden Feldern.

Nach dem Hochladen der Rohdaten versucht der Demonstrator mithilfe des Parsers, die Rohdaten einzulesen, und legt im Erfolgsfall ein neues Datenbeschreibungsobjekt in der relationalen Datenbank an. Darüber hinaus werden die eingelesenen und in das *HDF5*-Format überführten Daten auf dem Server

Kapitel 5. Konzept und Umsetzung

The screenshot shows a web-based energy data management system. At the top, there are navigation links for 'Data Custodian Service' and 'LOGOUT'. The main title is 'Energiedaten Detailansicht'. Below the title, a message states: 'Diese Energiedaten wurden am 2. November 2016 um 14:20 Uhr von Ihnen zur Verfügung gestellt.' The 'Technische Informationen' section includes details like category (Waschmaschine), unit (kWh), parser (CSV Typ Plugwise), and data owner (Nutzer1). The 'Datenanfragen' section indicates no requests have been made. The 'Löschen' section contains a note about不可逆性 and a red 'ENERGIEDATEN LÖSCHEN' button.

Technische Informationen

Kategorie(n): Waschmaschine
Einheit: kWh (Kilowattstunden)
Parser: CSV Typ Plugwise (Kommaseparierte Dateien)
Dateneigner: Nutzer1 (Sie selbst)

Diese Energiedaten betreffende Datenanfragen

Diese Energiedaten wurden bisher noch nicht angefragt.

Diese Energiedaten löschen

Sie können diese Energiedaten löschen. Achtung: Der Vorgang kann nicht rückgängig gemacht werden! Gelöscht werden nur die eigentlichen Zeitreihendaten, nicht aber die Informationen über die Daten (Metadaten).

ENERGIEDATEN LÖSCHEN

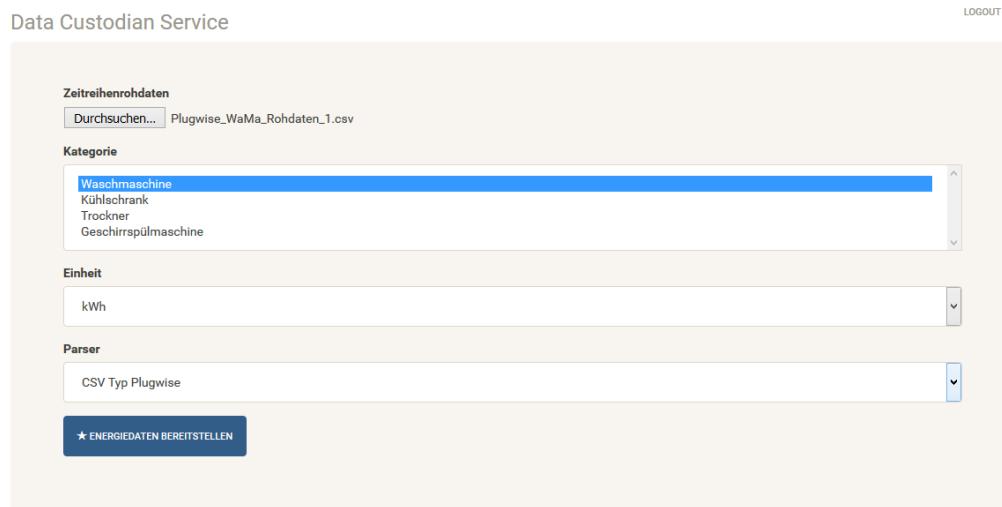
Abbildung 5.14.: Zu jedem Eintrag der Energiedatenübersichtsliste existiert eine Detailansicht. In dieser werden diverse technische Informationen dargestellt. Darüber hinaus werden die entsprechenden Datenanfragen aufgezeigt, welche den Datensatz angefragt haben. Zuletzt wird die Möglichkeit geboten, die Daten zu löschen.

abgespeichert. Fortan stehen die Daten im Energiedatenverwaltungssystem bereit und können damit potenziell von Dateninteressenten angefragt werden.

5.4.4. Erstellung neuer Datenanfragen

Sobald einer oder mehrere Nutzer Energiedaten zur Verfügung gestellt haben, können Dateninteressenten Anfragen für diese Daten einreichen. Wie bei den vorherigen Aktionen erfolgt die Erstellung einer Datenanfrage über ein Webformular, wie es in Abbildung 5.16 dargestellt ist. Diese Funktionalität steht ausschließlich Dateninteressenten zur Verfügung und Nutzer haben keinen Zugriff darauf. Für den Fall, dass ein Nutzer andere Energiedaten anfragen möchte, muss ein Benutzerkonto mit der Rolle eines Dateninteressenten eröffnet werden.

5.4. Typische Arbeitsabläufe



The screenshot shows a web-based data submission form titled 'Data Custodian Service'. At the top right is a 'LOGOUT' link. The form fields include:

- Zeitreihenrohdaten:** A button labeled 'Durchsuchen...' with the file path 'Plugwise_WaMa_Rohdaten_1.csv'.
- Kategorie:** A dropdown menu showing 'Waschmaschine' as the selected item, with other options like 'Kühlschrank', 'Trockner', and 'Geschirrspülmaschine' listed below.
- Einheit:** A dropdown menu showing 'kWh'.
- Parser:** A dropdown menu showing 'CSV Typ Plugwise'.
- Buttons:** A blue button at the bottom left labeled '★ ENERGIEDATEN BEREITSTELLEN'.

Abbildung 5.15.: Über ein Webformular kann ein Nutzer neue Energiedaten zur Verfügung stellen. Nach dem Absenden des Formulars werden die Daten durch den Demonstrator in das HDF5-Format überführt und abgespeichert.

Zunächst muss der Dateninteressent einen Nutzer auswählen, von dem die Energiedaten stammen sollen. An dieser Stelle stehen alle Benutzerkonten mit der Rolle eines Nutzers zur Verfügung, unabhängig davon, ob tatsächlich Energiedaten hinterlegt wurden oder nicht. Nach der Eingrenzung auf einen Nutzer ist der gewünschte Zeitraum in Form eines Start- sowie Endzeitpunktes anzugeben. Mit der Angabe einer Kategorie ist die Eingrenzung abgeschlossen. Auch bei den letzten genannten Punkten kann ein Dateninteressent beliebige Werte auswählen, welche zunächst nicht in Bezug auf die tatsächlich vorhandenen Energiedaten geprüft werden. Dies verhindert, dass auf einem Umweg Informationen über die von einem Nutzer zur Verfügung gestellten Daten abgefragt werden können. Würde dem Dateninteressenten direkt Rückmeldung bezüglich Zeiträumen und Kategorien gegeben, könnte dieser prüfen, ob Daten einer bestimmten Kategorie vorliegen und damit Rückschlüsse auf vorhandene Geräte ziehen.

Ein Dateninteressent muss außerdem aus einer Reihe von Ausgabeformaten eines auswählen, welches verwendet wird, um die angefragten Energiedaten im Falle einer Zustimmung durch den Nutzer zu exportieren. Dies vereinfacht die Weiterverwendung der Daten in anderen Werkzeugen.

Kapitel 5. Konzept und Umsetzung

Abbildung 5.16.: Über ein Webformular kann ein Dateninteressent eine Datenanfrage einreichen. Dabei können vom Dateninteressenten beliebige Zeiträume und Kategorien angefragt werden, unabhängig davon, ob ein Nutzer überhaupt entsprechende Daten zur Verfügung stellt. Dies dient dem Schutz der Privatsphäre der Nutzer, da Dateninteressenten so nicht erkennen können, ob bestimmte Daten tatsächlich zur Verfügung gestellt wurden oder nicht.

Wie in Abschnitt 5.3.5.1 erläutert, spielt die zeitliche Auflösung bei der Reduktion der Aussagekraft eine entscheidende Rolle. Einerseits benötigen die Dateninteressenten eine gewisse Datenqualität und Aussagekraft der Daten, um die gewünschten Analysen durchführen zu können. Andererseits geht mit einer hohen Aussagekraft auch eine hohe Gefährdungskennzahl einher, die die Nutzer dazu bewegen könnte, die Datenanfrage abzulehnen. Somit sollte ein Dateninteressent die geringstmögliche Datenqualität festlegen, die ausreicht, um die angestrebten Analysen durchzuführen.

Im Rahmen einer Datenanfrage muss ein Dateninteressent deshalb die Auflösung angeben, welche die angefragten Energiedaten aufweisen sollen.

Darüber hinaus kann aus einer Auswahl verschiedener Qualitätsreduktionsmaßnahmen gewählt werden, die die Datenqualität weiter verringern (vgl. Abschnitt 5.3.5.2 und Abschnitt 5.3.5.3). Dies führt zwar zu einer Verschlechterung der Aussagekraft der Daten, jedoch gleichzeitig zu einer geringeren Gefährdungskennzahl und damit zu einer höheren Wahrscheinlichkeit, dass die Nutzer die Datenanfrage annehmen. Eine Mehrfachauswahl von Qualitätsreduktionsmaßnahmen wird nur berücksichtigt, falls sie sinnvoll ist. Dies ist beispielsweise der Fall, wenn hinzugefügtes Rauschen in Verbindung mit einer Reduktion der Messwertauflösung gewählt wird. Andernfalls wird aus der gewählten Kombination die strogste Variante ausgewählt und angewendet.

Mit dem Erstellen einer Datenanfrage wird diese dem betroffenen Nutzer zur Prüfung vorgelegt. Ein Datenzugriff ist für die Dateninteressenten zu diesem Zeitpunkt jedoch selbstverständlich nicht möglich, da zuvor der betroffene Nutzer eine Entscheidung über die Annahme der Datenanfrage treffen muss.

5.4.5. Verwaltung eingegangener Datenanfragen

Analog zur Übersicht über die eigenen Energiedaten bietet der Demonstrator den Nutzern einen Überblick über eingegangene Datenanfragen. Diese Ansicht steht ausschließlich Nutzern und nicht Dateninteressenten zur Verfügung, wobei ein Nutzer lediglich die ihn betreffenden Datenanfragen sehen kann und keine weiteren.

Im Überblick wird zwischen drei verschiedenen Kategorien von Datenanfragen unterschieden: noch unbearbeitete, angenommene und abgelehnte Anfragen. Auf diese Weise haben die Nutzer stets die Möglichkeit, nachzuvollziehen, welche Datenanfragen noch von ihnen bearbeitet werden sollten und welche Energiedaten bereits angefragt wurden. Der Screenshot in Abbildung 5.17 zeigt die vom Demonstrator erstellte Ansicht für einen Nutzer, welcher zwei neue Datenanfragen von einem Dateninteressenten erhalten hat.

Über die entsprechenden Einträge in der Liste kann der Nutzer die Datenanfragen bearbeiten, das heißt, die vom Demonstrator geschätzte Gefährdung erfahren und die Anfrage annehmen oder ablehnen. Die Abbildungen 5.18 und 5.19 zeigen die Detailansichten für die beiden Datenanfragen, wie sie dem Nutzer durch den Demonstrator präsentiert werden.

Kapitel 5. Konzept und Umsetzung

The screenshot shows a web-based application for managing data requests. At the top, there are navigation links for 'Data Custodian Service' and 'LOGOUT'. The main content area is titled 'Übersicht zu eingegangenen Datenanfragen' (Overview of incoming data requests). It displays a message: 'Folgende Datenanfragen wurden bisher an Sie gerichtet, Nutzer1:'. Below this, there are two entries in a table:

Übersicht zu eingegangenen Datenanfragen
Zu bearbeitende Datenanfrage 1: Erstellt am 7. November 2016 durch Dateninteressent1 bezüglich Waschmaschine
Zu bearbeitende Datenanfrage 2: Erstellt am 7. November 2016 durch Dateninteressent1 bezüglich Kühlschrank

Below this table, there are three sections: 'Noch nicht bearbeitete Anfragen' (Not yet processed requests), 'Angenommene Anfragen' (Accepted requests), and 'Abgelehnte Anfragen' (Rejected requests). Each section contains a message: 'Sie haben bisher keine Datenanfragen angenommen.' (You have not yet accepted any data requests.), 'Sie haben bisher keine Datenanfragen abgelehnt.' (You have not yet rejected any data requests.).

Abbildung 5.17.: Dieser Screenshot zeigt die Datenanfragenübersicht für einen Nutzer, der lediglich zwei Datenanfragen erhalten und bisher noch keine davon bearbeitet hat. Über die entsprechenden Listeneinträge können die Datenanfragen betrachtet und bearbeitet werden.

Der Demonstrator erstellt mithilfe der Daten aus der relationalen Datenbank eine Beschreibung der Anfrage und präsentiert diese dem Nutzer. Zunächst werden jeweils allgemeine Informationen über die Datenanfrage aufgelistet. Diese umfassen den Zeitpunkt, zu dem die Datenanfrage eingereicht wurde und den Dateninteressenten, welcher die Daten erhalten möchte. Darüber hinaus werden dem Nutzer auch der abgedeckte Zeitraum und die gewünschte Kategorie übermittelt.

Danach kommt der für den Nutzer wichtigste Part: die Bewertung der Datenanfrage. Hierfür analysiert der Demonstrator die Datenanfrage mithilfe der hinterlegten Bewertungsmethode und berechnet diverse Gefährdungskennzahlen (vgl. Abschnitt 5.3.4). Neben der errechneten Gesamtbewertung in Form einer Kennzahl von 0 bis 100 sind auch die detaillierten Kennzahlen für die Auflösung, den Zeitraum, das Datenalter und die Kategorie von Relevanz (vgl. Abschnitt 5.3.4). Dass es durchaus sinnvoll ist, auch diese Kennzahlen mit aufzuführen, zeigt ein Vergleich der Abbildungen 5.18 und 5.19.

5.4. Typische Arbeitsabläufe

Data Custodian Service LOGOUT

Datenanfrage Detailansicht

Diese Datenanfrage wurde am 7. November 2016 um 09:57 Uhr von Dateninteressent1 eingereicht.

Allgemeine Informationen

Zeitraum: Von 1. Januar 2016 00:00 bis 30. Juni 2016 23:59
Betroffene Kategorie: Waschmaschine

Datenanfragenbewertung

Der Data Custodian Service hat diese Datenanfrage mit der Standardbewertungsmethode für Sie bewertet ([Bewertungsmethode ändern](#)). Folgendes Ergebnis wurde festgestellt:

Die Anfrage erhält die Gesamtbewertung 73 von 100

Diese Einzelbewertungen wurden berücksichtigt:
Auflösung 55 **Zeitraum 87** **Datenalter 71** **Kategorie 90**

Hinweis: Der angefragte Zeitraum ist sehr groß! Möglicherweise sind detaillierte Rückschlüsse über Ihre Gewohnheiten oder Ihr Verhalten möglich.
Hinweis: Die Daten stammen aus einer kritischen Kategorie.
Hinweis: Es wurden keine weiteren Qualitätsreduktionsmaßnahmen gewählt.
Sie haben die Möglichkeit, auch [historische Anfragen in die Bewertung mit einzubinden](#). Achtung: Dies kann zu sehr pessimistischen/paranoiden Bewertungen führen.

Ihre Entscheidung

ANFRAGE ANNEHMEN ANFRAGE MIT HINWEIS AN DATENINTERESSENTEN ABLEHNNEN NICHTS TUN

Abbildung 5.18.: Diese Abbildung zeigt eine Datenanfrage mit einer Gefährdungskennzahl von 73 und dementsprechend rot hinterlegter Fläche. Zwar sollen die Daten nur mit einer geringen Auflösung weitergegeben werden (gelbe Fläche), jedoch ist der angefragte Zeitraum recht groß und die Daten entstammen einer kritischen Kategorie (jeweils rote Fläche). Über die entsprechenden Schaltflächen kann ein Nutzer diese Datenanfrage mit einem Hinweis an den Dateninteressenten ablehnen beziehungsweise annehmen und so dem Dateninteressenten Zugriff auf die in der Datenanfrage festgelegte Datenqualität erlauben.

Beide Anfragen weisen die gleiche Gesamtbewertungskennzahl auf, unterscheiden sich jedoch signifikant in den Einzelbewertungen. In der ersten Anfrage werden Daten einer Waschmaschine mit einer niedrigen Auflösung über einen langen Zeitraum angefragt. In der zweiten handelt es sich hingegen nur um Daten eines Kühlschranks über einen kurzen Zeitraum, dafür aber mit einer hohen Auflösung.

Im Anschluss an die einzelnen Gefährdungskennzahlen werden nach Bedarf kurze textuelle Hinweise eingeblendet, die die Nutzer auf besonders hohe Kennzahlen und die damit potenziell verbundenen Folgen hinweisen sollen. Ein Einbeziehen historischer Anfragen in die Bewertung ist in der Stan-

Kapitel 5. Konzept und Umsetzung

Data Custodian Service

LOGOUT

Datenanfrage Detailansicht

Diese Datenanfrage wurde am 7. November 2016 um 09:58 Uhr von Dateninteressent1 eingereicht.

Allgemeine Informationen

Zeitraum: Von 1. Oktober 2016 00:00 bis 1. Oktober 2016 23:59
Betroffene Kategorie: Kühlschrank

Datenanfragenbewertung

Der Data Custodian Service hat diese Datenanfrage mit der Standardbewertungsmethode für Sie bewertet ([Bewertungsmethode ändern](#)). Folgendes Ergebnis wurde festgestellt:

Die Anfrage erhält die Gesamtbewertung 73 von 100

Diese Einzelbewertungen wurden berücksichtigt:

Auflösung 97 **Zeitraum 52** **Datenalter 85** **Kategorie 50**

Hinweis: Die Auflösung ist sehr hoch! Möglicherweise sind detaillierte Rückschlüsse auf Ihnen verwendete Geräte möglich.
Hinweis: Es wurden vergleichsweise aktuelle Daten angefragt.
Hinweis: Es wurden keine weiteren Qualitätsreduktionsmaßnahmen gewählt.

Sie haben die Möglichkeit, auch [historische Anfragen in die Bewertung mit einzubiehen](#). Achtung: Dies kann zu sehr pessimistischen/paranoiden Bewertungen führen.

Ihre Entscheidung

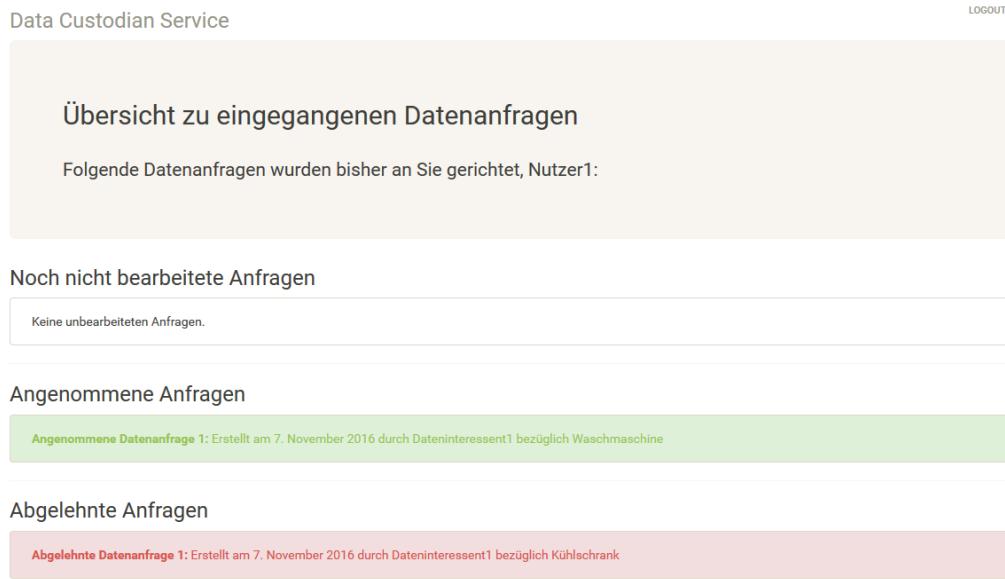
ANFRAGE ANNEHMEN **ANFRAGE MIT HINWEIS AN DATENINTERESSENTEN ABLEHNNEN** **NICHTS TUN**

Abbildung 5.19.: Diese Abbildung zeigt ebenfalls eine Datenanfrage mit einer Gefährdungskennzahl von 73 (rot hinterlegt). Im Vergleich zu Abbildung 5.18 ist hier die gewünschte Auflösung deutlich höher (rot hinterlegt), dafür aber der angefragte Zeitraum kürzer (gelb hinterlegt). Des Weiteren entstammen die Daten einer weniger kritischen Kategorie (gelb hinterlegt). In Summe ergibt sich für den Nutzer jedoch dieselbe Gesamtgefährdungskennzahl, auch wenn sie sich anders zusammensetzt.

dardansicht zunächst nicht vorgesehen, um die Übersichtlichkeit der Seite nicht zu sehr einzuschränken.

Schließlich bietet der Demonstrator den Nutzern noch die Möglichkeit, über die Datenanfrage zu entscheiden. Mit den entsprechenden Schaltflächen kann die Datenanfrage angenommen, mit einem Hinweis an den Dateninteressenten abgelehnt werden oder die Entscheidung vertagt werden. Der Screenshot in Abbildung 5.20 zeigt die Übersicht über die eingegangenen Datenanfragen, nachdem der Nutzer alle Anfragen abgearbeitet hat. Hierdurch ist es den Nutzern möglich, zu jedem Zeitpunkt nachzuvollziehen, welche Datenanfragen genehmigt wurden und welche Daten damit potenziell im Umlauf sind. Darauf hinaus können zunächst angenommenen Datenanfragen auch nachträglich abgelehnt werden, so dass ein erneuter Zugriff auf die Daten

5.4. Typische Arbeitsabläufe



The screenshot shows the 'Data Custodian Service' interface. At the top, there are links for 'Data Custodian Service' and 'LOGOUT'. The main content area is titled 'Übersicht zu eingegangenen Datenanfragen' and displays a message: 'Folgende Datenanfragen wurden bisher an Sie gerichtet, Nutzer1:'. Below this, there are three sections: 'Noch nicht bearbeitete Anfragen' (No pending requests), 'Angenommene Anfragen' (Accepted requests, with one entry: 'Angenommene Datenanfrage 1: Erstellt am 7. November 2016 durch Dateninteressent1 bezüglich Waschmaschine'), and 'Abgelehnte Anfragen' (Rejected requests, with one entry: 'Abgelehnte Datenanfrage 1: Erstellt am 7. November 2016 durch Dateninteressent1 bezüglich Kühlschrank').

Abbildung 5.20.: Diese Abbildung zeigt die Übersicht über die eingegangenen Datenanfragen, nachdem der betroffene Nutzer die ursprünglich noch offenen Anfragen bearbeitet hat (vgl. Abbildung 5.17).

unterbunden wird. Damit obliegt es der Verantwortung der Dateninteressenten, die freigegebenen Daten selbstständig zu verwalten und vorzuhalten, da die Möglichkeit eines nachträglichen Herunterladens nicht garantiert werden kann.

5.4.6. Verwaltung eingereichter Anfragen

Nicht nur die Nutzer, sondern auch die Dateninteressenten haben eine Möglichkeit zur Verwaltung der Datenanfragen. Im Falle der Dateninteressenten ist der Zweck hierbei jedoch nicht die Beurteilung und Bewertung der Datenanfragen, sondern ein Überblick über die eingereichten beziehungsweise abgelehnten Datenanfragen sowie die damit verbundenen Energiedaten. Abbildung 5.21 zeigt die Ansicht, wie sie dem Dateninteressenten präsentiert wird, nachdem der betroffene Nutzer die an ihn gerichteten Anfragen bearbeitet hat. Analog zur Übersicht für die Nutzer erfolgt auch hier eine Einteilung in drei Kategorien: zu bearbeitende, angenommene und abgelehnte Anfragen.

Kapitel 5. Konzept und Umsetzung

The screenshot shows a web-based application for managing data requests. At the top, there are navigation links for 'Data Custodian Service' and 'LOGOUT'. The main content area is titled 'Übersicht zu abgegebenen Datenanfragen' (Overview of submitted data requests). Below this, a message states: 'Folgende Datenanfragen haben Sie bisher eingereicht, Dateninteressent1:'. The interface is divided into several sections:

- Noch nicht durch Nutzer bearbeitete Anfragen:** A box containing the message 'Keine unbearbeiteten Anfragen.'
- Von Nutzern angenommene Anfragen:** A box containing the message 'Angenommene Datenanfrage 1: Erstellt am 7. November 2016 für Nutzer1 bezüglich Waschmaschine'
- Von Nutzern abgelehnte Anfragen:** A box containing the message 'Abgelehnte Datenanfrage 1: Erstellt am 7. November 2016 für Nutzer1 bezüglich Kühlschrank'

Abbildung 5.21.: Dateninteressenten verfügen – wie auch die Nutzer – über einen Überblick bezüglich ihrer eingereichten Datenanfragen. Analog zur Darstellung für die Nutzer erfolgt auch hier eine Unterteilung nach noch zu bearbeitenden, angenommenen und abgelehnten Anfragen.

Über die Einträge in den Listen kann ein Dateninteressent auf die Detailsichten zugreifen. In den Abbildungen 5.22 und 5.23 sind die Ansichten für eine angenommene und eine abgelehnte Datenanfrage dargestellt. Die hier angezeigten Informationen unterscheiden sich geringfügig von denen, welche die Nutzer präsentiert bekommen (vgl. Abbildung 5.18). Die Dateninteressenten erfahren beispielsweise mehr technische Details zu ihren Datenanfragen wie die gewünschte Auflösung und das eingestellte Ausgabeformat.

Abhängig von der Entscheidung der betroffenen Nutzern haben die Dateninteressenten die Möglichkeit, an dieser Stelle die angefragten Daten in der vereinbarten Qualität herunterzuladen: Für den Fall, dass ein Nutzer eine Datenanfrage genehmigt, wird eine entsprechende Schaltfläche angezeigt, mit der die Daten abgerufen können. Sobald die Daten abgerufen werden sollen, sucht der *Data Custodian Service* nach passenden Energiedaten des Nutzers und selektiert den entsprechenden Ausschnitt. Anschließend wird die Datenqualität, wie in der Datenanfrage festgelegt, verringert und die Daten in das gewünschte Ausgabeformat überführt. Für den Fall, dass keine Daten beziehungsweise nur lückenhafte Daten zur Verfügung gestellt werden

5.4. Typische Arbeitsabläufe



The screenshot shows a web interface for 'Data Custodian Service'. At the top right is a 'LOGOUT' link. The main content area has a light gray header 'Datenanfrage Detailansicht'. Below it, a message says 'Diese Datenanfrage haben Sie am 7. November 2016 um 09:57 Uhr bei Nutzer1 eingereicht.' (This data request was submitted by user1 on November 7, 2016, at 09:57 AM).

Allgemeine Informationen

Zeitraum: Von 1. Januar 2016 00:00 bis 30. Juni 2016 23:59
Zeitliche Auflösung: 1w
Zusätzliche Qualitätsreduktionsmaßnahmen: Keine Qualitätsreduktion
Betroffene Kategorie: Waschmaschine
Ausgabeformat: JSON

Entscheidung des betroffenen Nutzers

Die Energiedatenanfrage wurde durch den Nutzer genehmigt. Sie können die Daten herunterladen.

[DATEN IN ANGEFRAGTER QUALITÄT HERUNTERLADEN](#)

Abbildung 5.22.: Diese Abbildung zeigt die Detailansicht einer eingereichten Datenanfrage, welche durch den betroffenen Nutzer genehmigt wurde. Über die Schaltfläche kann der Dateninteressent die angegebenen Energiedaten in der vereinbarten Qualität herunterladen.



The screenshot shows a web interface for 'Data Custodian Service'. At the top right is a 'LOGOUT' link. The main content area has a light gray header 'Datenanfrage Detailansicht'. Below it, a message says 'Diese Datenanfrage haben Sie am 7. November 2016 um 09:58 Uhr bei Nutzer1 eingereicht.' (This data request was submitted by user1 on November 7, 2016, at 09:58 AM).

Allgemeine Informationen

Zeitraum: Von 1. Oktober 2016 00:00 bis 1. Oktober 2016 23:59
Zeitliche Auflösung: 1s
Zusätzliche Qualitätsreduktionsmaßnahmen: Keine Qualitätsreduktion
Betroffene Kategorie: Kühlschrank
Ausgabeformat: JSON

Entscheidung des betroffenen Nutzers

Die Datenanfrage wurde durch den Nutzer abgelehnt! Sie können jedoch versuchen, eine neue Datenanfrage mit geringerer Qualität einzureichen.

[KEIN DOWNLOAD MÖGLICH](#)

Abbildung 5.23.: Diese Abbildung zeigt die Detailansicht einer eingereichten Datenanfrage, welche durch den Nutzer ohne weitere Hinweise abgelehnt wurde. Im Gegensatz zur genehmigten Datenanfrage (vgl. Abbildung 5.22) können die Daten hier nicht heruntergeladen werden.

können, wird eine Warnung ausgelöst. Die Daten werden jedoch trotzdem generiert, wobei mögliche Lücken markiert werden.

5.4.7. Überführung in ein Produkt für den Realwelteinsatz

Wie bereits zuvor erläutert, handelt es sich bei dem Demonstrator um eine *Proof-of-Concept*-Implementierung, die hauptsächlich dazu dient, eine mögliche Funktionsweise und typische Arbeitsabläufe eines nutzerorientierten Energiedatenmanagementsystems zu demonstrieren. Dementsprechend ist die gezeigte Anwendung nicht unmittelbar für einen produktiven Realwelteinsatz vorgesehen.

Dies äußert sich vor allem darin, dass zum Teil Software-Komponenten verwendet werden, die für eine Produktivumgebung nicht vorgesehen sind, wie beispielsweise der in *Django* integrierte Webserver. Vor einem tatsächlichen Produktiveinsatz müssten die entsprechenden Komponenten durch geeignete ersetzt werden. Darüber hinaus sollte das System von entsprechenden Experten auf seine Sicherheit geprüft werden.

Davon abgesehen wären einige Erweiterungen des Systems denkbar, die die Nutzbarkeit und den Komfort, aber gegebenenfalls auch die Sicherheit verbessern. Einige Anregungen hierzu werden im Folgenden kurz skizziert.

Grafische Aufarbeitung der Daten

Es ist naheliegend, die von einem Dateninteressenten angefragten Daten grafisch für die Nutzer aufzubereiten. Mit einem entsprechenden Werkzeug könnte den Nutzern die Möglichkeit geboten werden, die Daten zu betrachten und zu erkunden und damit möglicherweise Hinweise darauf zu entdecken, dass bestimmte Daten nicht freigegeben werden sollten. Darüber hinaus würde eine solche Funktionalität vermutlich das Verständnis für die Daten verbessern und gegebenenfalls das Interesse am eigenen Energieverbrauch fördern. Mit einer solchen Funktionalität könnte der *Data Custodian Service* auch dazu beitragen, die Transparenz des eigenen Energieverbrauchs zu erhöhen und damit einen Einspareffekt herbeiführen.

Aushandeln der Aussagekraft

Die oben vorgestellten Arbeitsabläufe zeigen, dass Nutzer und Dateninteressenten nur über ein begrenztes Interaktionspotential verfügen. So ist es den Nutzern lediglich möglich, eine Datenfrage anzunehmen oder abzulehnen – gegebenenfalls mit einem textuellen Hinweis, warum dies geschehen ist. Es ist aber auch vorstellbar, diesen Mechanismus dahingehend zu erweitern, dass ein interaktives Aushandeln der Aussagekraft der Daten möglich wird. Anstatt eine – aus Nutzersicht – zu kritische Datenanfrage lediglich abzulehnen, könnten die Nutzer stattdessen ein Gegenangebot vorlegen, welches Daten mit geringerer Aussagekraft enthält. Somit wäre es einfacher, den Interessenkonflikt der Dateninteressenten und Nutzer im Hinblick auf die Aussagekraft der Daten und den damit einhergehenden Verlust an Privatsphäre zu entschärfen. Dieser Aspekt steht in engem Zusammenhang zum nächsten vorgestellten Punkt, dem Anreizsystem.

Anreizsystem

Im Demonstrator ist zum gegenwärtigen Zeitpunkt kein Anreizsystem für die Teilnahme beziehungsweise das Freigeben der Energiedaten vorgesehen. Zwar ist es durchaus denkbar, dass Nutzer an einem wie dem hier vorgestellten System teilnehmen: Die Beispiele *OpenStreetMap*²³ und *Wikipedia*²⁴ zeigen eindrucksvoll, dass auch ohne monetäre Gegenleistung eine enorme Zahl an Beiträgen gesammelt werden kann. Bei Energiedaten handelt es sich jedoch um deutlich kritischere und zudem personenbezogene Daten, die vermutlich von den meisten Personen nicht ohne entsprechende Anreize freigegeben würden. Dementsprechend sollte ein geeignetes Anreizsystem geschaffen werden, welches Nutzer dazu animiert, ihre Daten an Dateninteressenten weiterzugeben. Denkbar sind hier neben dem offensichtlichen monetären Ansatz auch das Anbieten von Mehrwertdiensten oder *Gamification*-Technologien. Vollkommen ohne Anreizsystem würde das hier vorgestellte Energiedatenmanagementsystem vermutlich nicht funktionieren, da die meisten Nutzer nicht bereit wären, ihre Daten zu teilen.

²³<https://www.openstreetmap.de/>

²⁴<https://de.wikipedia.org/>

Erweiterte Datenanfragemöglichkeiten

Obgleich die Anfragemöglichkeiten des Demonstrators für die meisten Arten von Datenanfragen ausreichen, könnten sie – aus Sicht der Dateninteressenten – für manche Anwendungen komfortabler sein. So ist es beispielsweise denkbar, dass ein Dateninteressent Daten zum Stromverbrauch von Waschmaschinen aus einer großen Zahl von Haushalten erhalten möchte. Anstatt nun die Anfragen an jeden Haushalt einzeln schicken zu müssen, könnte eine übergeordnete Einheit diese Aufgabe übernehmen und die Anfragen an alle potenziell geeigneten Haushalte automatisiert verteilen. Bei Erreichen der gewünschten Anzahl an Datensätzen würde die allgemeine Datenanfrage zurückgezogen und der Dateninteressent erhielte die gewünschten Daten der Haushalte, welche zugestimmt haben. Eine nochmalige Erweiterung dieses Prinzips wäre es, die Daten aller Haushalte bereits zu aggregieren, bevor sie überhaupt weitergegeben werden. Damit wären die Auswirkungen hinsichtlich der Privatsphäre für die einzelnen Haushalte nochmals signifikant geringer.

Aus Nutzersicht könnte es wiederum komfortabler sein, gewisse Datenanfragen automatisiert zu bearbeiten. Dies geht jedoch mit der Gefahr einher, dass mit der Zeit entweder alle Datenanfragen abgelehnt oder angenommen werden, da die Nutzer das Energiedatenmanagementsystem aus Bequemlichkeit nicht mehr aktiv nutzen. Dementsprechend sollte abgewogen werden, ob es sinnvoll ist, eine automatisierte Bearbeitung der Datenanfragen tatsächlich zu ermöglichen.

Community-basierte Bewertung

Im Demonstrator ist ein Standardverfahren zur Bewertung der Datenanfragen integriert. Durch die Modellierung der Bewertungsfunktionen als Entität des Datenmodells ist es jedoch einfach möglich, neue Bewertungsfunktionen zur Verfügung zu stellen. Es wäre daher denkbar, auch Bewertungsverfahren zu integrieren, welche durch andere Nutzer zur Verfügung gestellt werden. Ein ähnliches Prinzip ist im Bereich der Werbeblocker für Browser zu beobachten, für welche Filter von verschiedenen Anbietern zur Verfügung gestellt und gemeinsam verbessert werden.

Darüber hinaus wäre es denkbar, den Nutzern die Möglichkeit zu geben, die Dateninteressenten zu bewerten. Solche Bewertungen könnten wiederum in die Beurteilung der Datenanfragen mit einfließen, so dass seriöse Dateninteressenten eine bessere Bewertung erhalten können also solche, die bereits durch negatives Verhalten aufgefallen ist. Dieser Aspekt wäre insbesondere bei der Implementierung eines Anreizsystems interessant, da die Nutzer so besser beurteilen können, ob versprochene Anreize auch tatsächlich gewährt werden.

5.5. Fazit zum Konzept und zum Demonstrator

In diesem Kapitel wurde zunächst ein Konzept für ein nutzerorientiertes Energiedatenmanagementsystem vorgestellt. Dieses beschreibt eine modulare Architektur, in welcher die jeweiligen Komponenten unterschiedliche Aufgaben im Hinblick auf die Verwaltung, Verarbeitung und Präsentation von Energiedaten übernehmen. Darüber hinaus wurde ein Datenmodell präsentiert, welches es ermöglicht, Informationen zu Energiedaten, sowie diese Daten betreffende Anfragen verknüpft mit weiteren Informationen in einem Datenbanksystem vorzuhalten.

Auf diesen abstrakten Modellierungen basierend wurde ein Demonstrator entworfen, welcher eine Möglichkeit darstellt, das Konzept umzusetzen. Hierbei wurde nicht nur der Struktur des Demonstrators Beachtung geschenkt, sondern auch aufgezeigt, welche etablierten Werkzeuge geeignet sind, um den Demonstrator umzusetzen. Diese Komponenten kommen in der realisierten Implementierung auch zum Einsatz. Des Weiteren wurde eine Reihe von Bewertungsfunktionen entworfen, welche jeweils unterschiedliche Aspekte einer Datenanfrage im Hinblick auf ihre potenziellen Auswirkungen auf die Privatsphäre bewerten. Darüber hinaus wurden diverse Verfahren zur Reduktion der Aussagekraft der Daten vorgestellt, um die Privatsphäre der betroffenen Nutzer zu schützen.

Schließlich wurden typische Arbeitsabläufe vorgestellt, wie sie in einem Energiedatenmanagementsystem, welches auf dem vorgestellten Konzept beruht, vorkommen können. Die Arbeitsabläufe entstammen dem implementierten Demonstrator, wie er in diesem Kapitel vorgestellt wurde. Es ist denkbar,

dass dieser Demonstrator in Kombination mit den vorgestellten Werkzeugen als Basis für die Entwicklung eines Systems dient, welches auch unter Realweltbedingungen einsetzbar ist.

Evaluation

Dieses Kapitel widmet sich der Evaluation des in dieser Arbeit vorgestellten Konzepts für ein Energiedatenmanagementsystem. Zunächst erfolgt eine Prüfung im Hinblick auf den in Kapitel 4 beschriebenen Anforderungskatalog. Anschließend wird geprüft, ob der *Data Custodian Service* in der Lage ist, für typische Anwendungsfälle die erforderlichen Daten zu liefern und gleichzeitig die Einschränkungen der Privatsphäre im Vergleich zum Referenzfall zu reduzieren. Das Kapitel schließt mit einem Ausblick auf weitergehende Evaluationsmöglichkeiten.

6.1. Prüfung der Anforderungen

In Kapitel 4 wurde eine Reihe von Anforderungen erarbeitet, die ein nutzerorientiertes Energiedatenmanagementsystem erfüllen sollte. Bei diesen Anforderungen handelt es sich einerseits um solche, die aus einer technischen Analyse hervorgegangen sind, und andererseits um nutzergetriebene Forderungen. Hierbei standen insbesondere der Datenschutz und die Privatsphäre im Vordergrund.

Der Anforderungskatalog ist in Tabelle 6.1 erneut zusammengefasst. In den folgenden Abschnitten wird für jede Anforderung geprüft, ob das in Kapitel 5 vorgestellte Konzept diese erfüllt. Für ein besseres Verständnis wird die zu prüfende Anforderung zu Beginn eines jeden Abschnitts kurz zusammengefasst. Detailliertere Informationen bezüglich der jeweiligen Anforderung sind in Kapitel 4 auf den in Tabelle 6.1 abgedruckten Seiten zu finden.

Tabelle 6.1.: Zusammenfassung der in Kapitel 4 erarbeiteten Anforderungen.

Nr.	Anforderungsbezeichnung	Seite
1	Unterstützung unterschiedlicher Datenquellen	79
2	Robustheit gegenüber fehlenden und fehlerhaften Daten	80
3	Lokale Sammelstelle unter Kontrolle des Nutzers	86
4	Transparenz der Datenübertragungswege	86
5	Unterstützung einer geeigneten Speicherform in Abhängigkeit der Datenart	91
6	Verschlüsselung bei externer Speicherung	91
7	Lokale Vorverarbeitung der Daten	95
8	Unterstützung unterschiedlicher Datenausgabeformate	98
9	Sicherheit der Datenübertragungswege	98
10	Lösung nicht mehr benötigter Daten	101
11	Einschränkung der Datenweitergabe	115
12	Reduktion der Aussagekraft auf akzeptables Mindestmaß vor Weitergabe	116
13	Aufklärung über mögliche Auswirkungen einer Datenfreigabe auf die Privatsphäre	116
14	Unterstützung zusätzlicher, externer Schutzmechanismen	117
15	Unterstützung verschiedener Endgeräte zur Darstellung der Benutzeroberfläche	119

6.1.1. Unterstützung unterschiedlicher Datenquellen

Die Analyse der ersten Phase des Datenlebenszyklus hat gezeigt, dass Energie-daten sehr heterogen sein können. Sie entstammen unterschiedlichen Quellen (beispielsweise intelligenten Zwischensteckdosen oder intelligenten Strom-zählern), liegen in verschiedenen Dateiformaten vor (beispielsweise CSV oder XML) und beruhen auf diversen physikalischen Größen (beispielsweise elektri-

sche Leistung oder elektrische Energie). Ein Energiedatenmanagementsystem sollte dennoch Daten aus unterschiedlichsten Quellen einlesen können.

Das Konzept des *Data Custodian Service* sieht im Hinblick auf die Handhabung von Zeitreihen ein dediziertes Modul für deren Verarbeitung vor (vgl. Abschnitt 5.1). Diese Komponente wird für mehrere Aufgaben verwendet, unter anderem auch für das Einlesen der Energiedaten. Hierzu wird ein quellentypspezifisches *Parser*-Objekt im *Data Custodian Service* angelegt, welches die – für diese Art von Energiedatenquelle – nötigen Funktionalitäten mit Hilfe der Komponente zur Zeitreihenverarbeitung bereitstellt. Ein *Parser* kann folglich nicht nur Daten einer bestimmten einzelnen Quelle einlesen, sondern alle Daten eines entsprechenden Quellentyps. Hierbei kann es sich beispielsweise um intelligente Zwischensteckdosen eines bestimmten Typs handeln.

Für den Umgang mit den unterschiedlichen physikalischen Größen stehen im Datenmodell hinterlegte *Einheiten*-Objekte bereit (vgl. Abschnitte 4.1.2.2 und 5.2). Dies erlaubt es, die eigentlichen Zahlenwerte unabhängig von den dazugehörigen Einheiten zu speichern, da die zugrundeliegende physikalische Größe in den Metadaten hinterlegt wird, und nicht in den Zeitreihen selbst. Sollte eine Einheit nicht im *Data Custodian Service* verfügbar sein, kann diese als neues Objekt im Datenmodell angelegt, und fortan verwendet beziehungsweise referenziert werden. Im Falle der intelligenten Zwischensteckdosen wäre die zugrundeliegende physikalische Größe beispielsweise die elektrische Leistung und die verwendete Einheit *Watt*.

Auch für eine kontinuierliche und nicht blockweise Erfassung von Energiedaten könnten entsprechende *Parser* implementiert werden. Hierbei würden die Daten dem Energiedatenmanagementsystem nicht dateiweise zur Verfügung gestellt, sondern kontinuierlich per REST-Schnittstelle über den Webserver an den *Data Custodian Service* übertragen. Ein solches Vorgehen ist jedoch im Vergleich zu den dateibasierten *Parseern* deutlich aufwendiger, da nicht nur einmalig, sondern kontinuierlich Daten eingelesen, umgewandelt und abgespeichert werden müssen.

Zusammenfassend ist der *Data Custodian Service* zwar nicht in der Lage, automatisch beliebige Arten von Energiedaten korrekt zu interpretieren. Es ist jedoch mit geringem Aufwand möglich, für beliebige Arten von Energiedaten, welche in Form von Zeitreihen vorliegen, einen *Parser* zu entwerfen,

welcher die Daten von einem Fremdformat in das zur internen Speicherung verwendete *HDF5*-Format überführt. Sobald diese Implementierungsarbeit für ein bestimmtes EingabefORMAT geleistet wurde, kann der *Parser* für alle Daten des gleichen Typs verwendet werden. Intern werden alle Arten von Energiedaten im *HDF5*-Format gespeichert und unabhängig vom ursprünglichen Quellentyp mittels *DataFrame*-Datenstrukturen verarbeitet. Der *Data Custodian Service* ist somit in der Lage, beliebige Arten von Energiedaten im Zeitreihenformat zu verarbeiten.

6.1.2. Robustheit gegenüber fehlenden und fehlerhaften Daten

Im Rahmen der Lebenszyklusanalyse wurde festgestellt, dass Energiedaten zum Teil fehler- und lückenhaft sein können. Ein Energiedatenmanagementsystem sollte dennoch in der Lage sein, auch solche Daten einzulesen und die fehlerfreien Ausschnitte zu verwenden.

Im *Data Custodian Service* werden die Energiedaten abhängig vom Quellentyp mit dem entsprechenden *Parser* eingelesen. Im Rahmen dessen werden Lücken, wie sie beispielsweise durch Übertragungsfehler bei der Verwendung drahtloser Sensoren auftreten können, erkannt und explizit gekennzeichnet. Im Gegensatz zu dem Ansatz, solche Lücken über Zahlenwerte mit „besonderer“ Bedeutung zu codieren – beispielsweise mit der Zahl -1 –, erfolgt im *Data Custodian Service* eine explizite Markierung mit dem dafür vorgesehenen Wert *NaN* (von engl. *not a number*). Dieser Wert ist von *pandas* (vgl. Abschnitt 5.3.2) explizit für diesen Zweck vorgesehen. Seine Verwendung hat den Vorteil, dass der Wert nicht fehlinterpretiert werden kann, wie es bei dem oben genannten Beispiel -1 durchaus der Fall sein kann. Ein weiterer Vorteil ist, dass lückenhafte Bereiche nicht über Metadaten referenziert werden müssen, sondern direkt in den Zeitreihendaten markiert werden können.

Sobald *pandas* Daten, welche Lücken – Stellen mit dem Wert *NaN* – enthalten, verarbeitet, kann gezielt auf die fehlenden Daten reagiert werden. Hierdurch können diese Lücken entweder ignoriert oder aber geschlossen werden, indem die fehlenden Werte interpoliert werden. Insbesondere für den Fall, dass Energiedaten im großen Umfang zeitlich aggregiert werden sollen, spielen

einige wenige Lücken in der Regel keine Rolle. Dementsprechend ist es von Vorteil, die Daten weiterverarbeiten und nutzen zu können, auch wenn die Rohdaten stellenweise lückenhaft waren. Es ist jedoch auch möglich, fehlende Daten nicht zu interpolieren, sondern die Lücken bei der Weiterverarbeitung zu erhalten und bei der Ausgabe zu markieren.

Falls auch fehlerhafte Daten – im Sinne von Energiedaten, welche falsch gemessene Werte enthalten – vom *Data Custodian Service* verarbeitet werden sollen, ist eine ähnlich einfache Erkennung in der Regel nicht mehr möglich. Stattdessen muss der verwendete *Parser* mit einer Fehlererkennungs- oder einer Filterfunktion ausgestattet werden, die fehlerhafte Werte erkennt und korrigiert, beziehungsweise durch den Wert *Nan* ersetzt. Damit würden die entsprechenden Fehlerbereiche in den Zeitreihen explizit gekennzeichnet und könnten fortan analog zu Lücken in den Daten behandelt werden.

Zusammenfassend kann festgehalten werden, dass der *Data Custodian Service* Möglichkeiten zur Verfügung stellt, um fehlende und fehlerhafte Daten explizit zu kennzeichnen, ohne jedoch die Nutzbarkeit der betroffenen Daten einzuschränken. Da die Informationen direkt in die Zeitreihen eingebettet sind, können sie effizient bei einer Datenweitergabe mitgeführt werden. Die Erkennung der problematischen Datenbereiche obliegt jedoch den *Parseern*, welche folglich mit entsprechenden Funktionalitäten ausgestattet werden müssen. Andererseits bedeutet dies, dass gezielt auf die Eigenheiten der unterschiedlichen Arten von Energiedatenquellen eingegangen werden kann und quellentypische Fehler einfacher erkannt werden können.

6.1.3. Lokale Sammelstelle unter Kontrolle des Nutzers

Im Rahmen der Untersuchung der Übertragungsphase im Datenlebenszyklus wurde festgestellt, dass Energiedaten zum Teil auch an externe Parteien übertragen werden und dort gespeichert werden. Die Daten unterliegen damit jedoch nicht mehr der Kontrolle der Nutzer. Daher sollte ein Energiedatenmanagementsystem eine einzelne, lokale Sammelstelle für Daten darstellen, die unter der vollständigen Kontrolle der Nutzer ist.

Der *Data Custodian Service* ist nicht in der Lage, andere Anwendungen an der Übertragung von Energiedaten zu hindern. Dementsprechend kann bei Verwendung von Fremdanwendungen nicht garantiert werden, dass keine

Daten an externe Stellen weitergegeben werden. Der *Data Custodian Service* verfügt jedoch über sehr flexible Möglichkeiten zur Erfassung von Energiedaten, so dass gegebenenfalls die Verwendung von Fremdsoftware entfallen kann. Wenn für einen bestimmten Quellentyp ein passender *Parser* zur Verfügung steht, kann auf proprietäre Programme verzichtet werden und der *Data Custodian Service* kann als alleinige, lokale Datensenke dienen. Darüber hinaus ist er nicht darauf angewiesen, Daten extern zu speichern, sondern verwendet in der Regel lokale Datenbanken für Metadaten und Zeitreihen. Der *Data Custodian Service* setzt bewusst nicht auf *Cloud*-Lösungen oder vergleichbare Angebote, sondern ist ausdrücklich als lokale Instanz unter Kontrolle der Nutzer entworfen worden. So können Nutzer auch zu jeder Zeit die Übersichten über den kompletten sie betreffenden Datenbestand einsehen. Ein hierdurch entstehender Nachteil ist, dass die Nutzer selbst für die Erstellung von Datensicherungen verantwortlich sind und dies nicht durch einen externen Dienstleister übernommen wird.

6.1.4. Transparenz der Datenübertragungswege

Diese Anforderung ist eng verwandt mit der zuvor genannten. Die Analyse der Übertragungsphase ergab, dass für die Nutzer oftmals nicht einsehbar ist, welche Parteien die Energiedaten erhalten und ob die Daten die Kontrolle des Nutzers verlassen oder nicht.

In einem ersten Schritt sollte der *Data Custodian Service* zunächst die alleinige Datensenke für die Energiedaten eines Nutzers sein. Diese Aufgabe kann von dem hier beschriebenen Energiedatenmanagementsystem erfüllt werden (vgl. Abschnitt 6.1.3). Sobald sichergestellt ist, dass die Daten ausschließlich zum *Data Custodian Service* und an keine weiteren Stellen übertragen werden, unterliegen die weiteren Übertragungswege der Kontrolle der Nutzer.

Vom *Data Custodian Service* werden keinerlei Daten ohne explizite Zustimmung der betroffenen Personen weitergegeben. Für den Fall, dass externe Parteien Zugriff auf die Daten erhalten möchten, muss eine Datenanfrage eingereicht werden, über die die Nutzer entscheiden. Falls die Daten durch eine bewusste Entscheidung der Nutzer freigegeben werden, ist dies transparent, welche Daten in welcher Qualität an welchen Dateninteressenten weitergegeben werden. Eine Datenweitergabe erfolgt *direkt* an die Anfragenden und

ohne Mittelsmann. Die Datenweitergabe erfolgt stets verschlüsselt über die von Webserver des *Data Custodian Service* angebotenen Schnittstellen (vgl. Abschnitt 6.1.9). Somit werden den Nutzern die Übertragungswege der Daten bis zur anfragenden Partei offengelegt und die Daten werden nicht durch dritte Parteien verschickt.

Sobald die Daten in Abstimmung mit den Nutzern jedoch weitergegeben wurden, unterliegen die potenziellen weiteren Übertragungswege nicht mehr der Kontrolle des *Data Custodian Service* oder des Nutzers. An dieser Stelle müssten zusätzliche Schutztechnologien wie beispielsweise *Usage Control* zum Einsatz kommen (vgl. Abschnitte 3.2.2.7 und 6.1.14). Der *Data Custodian Service* gewährleistet folglich die Transparenz der Übertragungswege – allerdings nur bis zu den anfragenden Parteien. Sollen auch die weiteren Übertragungswege gesichert werden, wird der Einsatz zusätzlicher Schutzmechanismen nötig.

6.1.5. Unterstützung einer geeigneten Speicherform in Abhängigkeit der Datenart

Die Analyse der dritten Phase des Datenlebenszyklus, der Speicherung, hat gezeigt, dass es zwei unterschiedliche Arten von Daten im Bereich der Energiedaten gibt: die Zeitreihendaten und die dazugehörigen Metadaten. Ein Energiedatenmanagementsystem sollte in Abhängigkeit der Datenart die geeignete Speicherform auswählen, um die Daten jeweils effizient abzuspeichern und bestmöglich nutzen zu können.

Der *Data Custodian Service* nutzt zwei verschiedene Speicherformen beziehungsweise Datenbanken, um die Energiedaten zu speichern. Der erste Teil, die Metadaten, wird in einer relationalen Datenbank vorgehalten. Über ein entsprechendes Datenmodell (vgl. Abschnitt 5.2) werden die Beziehungen der jeweiligen Entitäten zueinander abgebildet. Dies ermöglicht eine effiziente und einfache Durchsuchbarkeit und das Ausnutzen der Verbindungen der Entitäten untereinander. So ist beispielsweise schnell erkennbar, welche Datenanfragen ein bestimmter Dateninteressent eingereicht hat und welche Nutzer dies betrifft. Eine relationale Datenbank ist jedoch nicht für die effiziente Speicherung von Zeitreihendaten geeignet, so dass dort eine andere Technologie zum Einsatz kommen sollte.

Dementsprechend werden im *Data Custodian Service* die Zeitreihen in einem *HDF5* genannten und auf tabellenartige Daten spezialisierten Dateisystem vorgehalten. Dieses erlaubt einerseits die effiziente und performante Speicherung von großen Mengen von Zeitreihendaten. Andererseits ermöglicht es eine einfache Selektion und ein effizientes Auslesen ganzer Blöcke oder einzelner Abschnitte. Somit können auch aus großen Datenmengen die gewünschten Daten einfach extrahiert und anschließend verwendet werden. Der *Data Custodian Service* verwendet intern zur Weiterverarbeitung der Zeitreihendaten eine von *pandas* bereitgestellte Datenstruktur namens *DataFrame* (vgl. Abschnitt 5.3.2). Da diese Struktur nicht von der Art der Datenspeicherung abhängig ist, können prinzipiell auch andere Arten von Zeitreihendatenbanken zum Einsatz kommen. Dies ermöglicht es, gegebenenfalls zukünftig auf neue, leistungsfähigere Zeitreihendatenbanken zu setzen, welche in der jüngeren Vergangenheit entwickelt wurden (vgl. Abschnitt 4.1.4.3).

Der *Data Custodian Service* erfüllt damit die gestellte Anforderung und nutzt die Vorteile der jeweiligen Datenbanktechnologie aus. Darüber hinaus ist er durch die Verwendung einer dedizierten Komponente für den Datenbankzugriff (vgl. Abschnitt 5.1) flexibel im Hinblick auf die Wahl einer geeigneten Speicherform und kann mit geringem Aufwand an andere Datenbanken angepasst werden.

6.1.6. Verschlüsselung bei externer Speicherung

Für den Fall, dass Daten extern abgespeichert werden sollen, müssen diese verschlüsselt werden, um einen unbefugten Zugriff zu verhindern. Dementsprechend sollte ein Energiedatenmanagementsystem bei Verwendung externer Speicher eine vorgelagerte Verschlüsselung der Daten ermöglichen.

Der *Data Custodian Service* nutzt eine dedizierte Komponente für den *Datenbankzugriff* (vgl. Abschnitt 5.1). Dieses Modul ist dafür verantwortlich, die intern verwendeten Datenstrukturen umzuwandeln und in einer relationalen Datenbank beziehungsweise in einem Zeitreihenspeicher abzulegen. Im Zuge der Umwandlung können die Daten auch verschlüsselt und erst dann in den Datenbanken gespeichert werden. Sowohl für die objektrelationale Abbildung in *Django*, als auch für das Speichersystem *HDF5* stehen entsprechende Werkzeuge zur Verschlüsselung bereit (vgl. Abschnitte 5.3.1 und 5.3.3). Die beiden

Werkzeuge wurden bisher jedoch keiner eingehenden Sicherheitsprüfung durch unabhängige Dritte unterzogen, so dass von einem produktiven Einsatz abgeraten werden sollte.

Darüber hinaus ist es generell kritisch zu beurteilen, ob der gesamte Datenbestand – selbst wenn er verschlüsselt wird – einer externen Partei anvertraut werden sollte. Es muss von Fall zu Fall abgewogen werden, ob die Vorteile das Risiko eines Datenverlustes ausgleichen. Selbst wenn aus heutiger Sicht eine sichere Verschlüsselungsmethode gewählt wird, kann nicht ausgeschlossen werden, dass zukünftig Verfahren entwickelt werden, um die Verschlüsselung zu brechen. Wäre dies der Fall, hätte die externe Partei, welche die – zuvor „sicher“ verschlüsselten – Datenbanken vorgehalten hat, möglicherweise Zugriff auf den gesamten Datenbestand. Damit wäre aber wiederum die Privatsphäre der betroffenen Personen stark gefährdet. Folglich sollte eine externe Speicherung der Metadaten und Zeitreihen nur nach Abwägung von Risiko und Nutzen verwendet werden.

Der *Data Custodian Service* verfügt prinzipiell über die Möglichkeit, seinen Datenbestand verschlüsselt in externen Speichern abzulegen. Dieser Ansatz ist aus oben genannten Gründen jedoch kritisch zu sehen und sollte nach Möglichkeit im Interesse der Dateneigner nicht zum Einsatz kommen.

6.1.7. Lokale Vorverarbeitung der Daten

Im Rahmen der Untersuchung der vierten Phase des Datenlebenszyklus (Analyse) wurde festgestellt, dass ein Bedarf für eine lokale Vorverarbeitung der Energiedaten in Energiedatenmanagementsystemen existiert. Dies betrifft einerseits die Aufbereitung der Rohdaten und andererseits die Bereitstellung grundlegender Operationen auf den Zeitreihen.

Der *Data Custodian Service* verfügt über ein auf die Verarbeitung von Zeitreihen spezialisiertes Modul namens *Zeitreihen(vor-)verarbeitung* (vgl. Abschnitt 5.1). Im Demonstrator kommt für diese Komponente die Bibliothek *Pandas* zum Einsatz (vgl. Abschnitt 5.3.2). *Pandas* wird beispielsweise auch von den *Parseern* genutzt, um die Rohdaten einzulesen und in die intern verwendete Datenstruktur zu überführen. Die Bibliothek bietet eine Vielzahl verschiedener Methoden zur Manipulation von Zeitreihen wie beispielsweise die

Selektion von Ausschnitten oder die Addition von Zeitreihen. Darüber hinaus sind aber auch aufwendige Kombinationen mehrerer Zeitreihen möglich.

Ein Zugriff auf diese Funktionalitäten wird für externe Parteien beziehungsweise Dateninteressenten jedoch nur in einem sehr begrenzten Maß und indirekt ermöglicht: Sobald eine Datenanfrage durch einen Nutzer genehmigt wird, werden mithilfe von *Pandas* die entsprechenden Daten selektiert. Im Anschluss an diese Selektion erfolgt, ebenfalls mithilfe der Bibliothek, eine zeitliche Aggregation und gegebenenfalls eine weitere Manipulation der Daten. Diese Operationen sind dem Dateninteressenten jedoch nicht direkt zugänglich, sondern werden implizit – gemäß den Vereinbarungen in der Datenanfrage – durchgeführt.

Es wäre denkbar, Dateninteressenten die Möglichkeit zu bieten, auch weitere Funktionen von *Pandas* explizit zu nutzen. Dies würde beispielsweise Anfragen wie die folgende erlauben: Ein Dateninteressent könnte Interesse an dem aggregierten Verbrauch mehrerer Haushalte haben, ohne jedoch die jeweiligen Einzelverbräuche erfahren zu wollen. Vor einer Datenfreigabe könnten im *Data Custodian Service* die Zeitreihendaten mehrerer Nutzer, welche ihre Daten in dieser Instanz des *Data Custodian Service* hinterlegt haben, durch *Pandas* kombiniert und erst im Anschluss aggregiert ausgegeben werden. Mit dem bisherigen Anfrageschema, wie in Abschnitt 5.4.4 beschrieben, ist dies nicht ohne Weiteres möglich. Dementsprechend wären neue Anfrageschemata oder gar eine eigene, auf Energiedaten zugeschnittene, Anfragesprache nötig. Damit würde allerdings das Risiko steigen, dass Dateninteressenten durch geschickte Formulierung von Datenanfragen gegebenenfalls mehr Informationen erhalten können, als auf den ersten Blick ersichtlich ist. Aus diesem Grund werden den Dateninteressenten gegenwärtig keine Operationen von *Pandas* zugänglich gemacht.

Eine alternative Lösungsmöglichkeit für die oben beschriebene Anfrage wäre die Verwendung von Verfahren, welche gezielt für solche Anwendungsfälle entwickelt wurden. Bei *SMART-ER* (vgl. Abschnitt 3.2.2.2 und [FB14b]) handelt es sich beispielsweise um eine solche Technologie, die zum Einsatz kommen könnte. Der *Data Custodian Service* würde über ein entsprechend gestaltetes Ausgabemodul die Energiedaten eines Nutzers nicht an den Dateninteressenten, sondern an das auf *SMART-ER* basierende System schicken. Dieses System würde wiederum die Daten aller betroffenen Nutzer – welche

von den jeweiligen Instanzen eines *Data Custodian Service* geliefert wurden – aggregieren und erst dann an den Dateninteressenten weitergeben.

Der *Data Custodian Service* nutzt mit *Pandas* eine umfangreiche Bibliothek zur Manipulation von Zeitreihen. Dementsprechend sind verschiedenste Operationen auf den Energiedaten möglich. Diese Möglichkeiten werden externen Parteien jedoch nicht zur Gänze offengelegt, um die Gefahr eines verschleierten Abgreifens von Informationen zu minimieren.

6.1.8. Unterstützung unterschiedlicher Datenausgabeformate

Bei der Analyse der Verteilungsphase des Datenlebenszyklus wurde festgestellt, dass abhängig vom jeweiligen Anwendungsfall unterschiedliche Datenausgabeformate bei der Ausgabe von Energiedaten zum Einsatz kommen können. Ein Energiedatenmanagementsystem sollte deshalb möglichst viele Dateiformate zur Weitergabe der Daten unterstützen.

Im *Data Custodian Service* sind die *Ausgabeformate* das Pendant zu den *Parsern*. Sie sind ebenfalls Objekte im Datenmodell (vgl. Abschnitt 5.2) und können dem *Data Custodian Service* dynamisch hinzugefügt werden, um dessen Funktionsumfang im Hinblick auf die Datenausgabe zu erweitern. Wie die *Parser*, so machen auch die Entitäten für die Datenausgabe Gebrauch von der Komponente zur Zeitreihenverarbeitung, wobei im Falle des Demonstrators hierfür *Pandas* verwendet wird.

Nativ bietet *Pandas* unter anderem die Ausgabe der internen *DataFrame*-Datenstruktur in Dateiaustauschformaten wie beispielsweise *CSV* oder *JSON*, aber auch im *Excel*-Dateiformat an. Darüber hinaus können auch die bereits zuvor beschriebenen *HDF5*-Dateien verwendet werden, um Energiedaten auszugeben. Da es sich bei den Objekten für die Dateiausgabe jedoch um normale *Python*-Klassen handelt, können beliebige Ausgabeformate unterstützt werden. Dies umfasst insbesondere auch Maßnahmen für das digitale Rechtemanagement wie beispielsweise *Usage Control* (vgl. Abschnitt 3.2.2.7 und 6.1.14).

Der *Data Custodian Service* ist folglich durch den modularen Aufbau auch ausgabeseitig in der Lage, mit beliebigen Dateiformaten zu arbeiten, wobei

durch die verwendete Bibliothek *Pandas* bereits nativ eine große Zahl verbreiteter Formate genutzt werden kann. Für den Fall, dass weitere Ausgabeformate unterstützt werden sollen, ist eine Integration dieser mit wenig Aufwand möglich.

6.1.9. Sicherheit der Datenübertragungswege

Sobald eine Datenanfrage genehmigt wurde, können die Energiedaten an die anfragende Partei übertragen werden. Wird die Übertragung vom Energiedatenmanagementsystem zum Dateninteressenten nicht geschützt, können die Daten von unbefugten Dritten möglicherweise abgegriffen werden. Deshalb sollte ein Energiedatenmanagementsystem jegliche Kommunikation mit seinen Interaktionspartnern schützen.

Das Konzept des *Data Custodian Service* sieht vor, mit den Interaktionspartnern lediglich über einen Webserver zu kommunizieren (vgl. Abschnitt 5.1). Dementsprechend muss insbesondere die Verbindung vom Webserver zu den Nutzern beziehungsweise zu den Dateninteressenten geschützt werden. In einer Produktivumgebung sollten deshalb ausschließlich etablierte Webserver wie beispielsweise der *Apache HTTP Server*¹ oder *nginx*² zum Einsatz kommen. Diese Produkte bieten eine native und erprobte Unterstützung von Transportverschlüsselungsverfahren: Sie können so konfiguriert werden, dass ausschließlich eine Kommunikation per *HTTPS (Hypertext Transfer Protocol Secure)* unter Nutzung von *TLS (Transport Layer Security)* möglich ist.

Aus Sicherheitsgründen sollten nämlich nicht nur die Übertragungen der angefragten Energiedatenzeitreihen geschützt werden, sondern die *gesamte* Kommunikation der Interaktionspartner mit dem *Data Custodian Service*. Dies erschwert beziehungsweise verhindert nicht nur das Abgreifen von Energiedaten, sondern auch das Ausspähen von Informationen über potenziell verfügbare Daten und angemeldete Nutzer: Würden eingehende Datenanfragen nicht verschlüsselt eingereicht, könnte ein Angreifer beispielsweise sehr einfach in Erfahrung bringen, welche Daten bei welchem Nutzer von welchem Dateninteressenten angefragt wurden und aus der Reaktion des Nutzers schließen, ob solche Daten vorhanden sind, oder nicht. Der Schutz

¹<https://httpd.apache.org/>

²<https://nginx.org/>

der Kommunikation nach außen obliegt damit größtenteils nicht dem *Data Custodian Service* direkt, sondern dem vorgesetzten Webserver, welcher ein Teil des Energiedatenmanagementsystems ist.

In Summe kann festgehalten werden, dass der *Data Custodian Service* die Datenübertragung gemäß dem Stand der Technik unter Verwendung etablierter Transportverschlüsselungsverfahren schützen kann. In einem Produktiveinsatz sollten diese Möglichkeiten daher zur Gänze ausgeschöpft werden und die gesamte Kommunikation verschlüsselt werden. Im Rahmen dessen könnte als Leitlinie zur konkreten Ausgestaltung auch das BSI-Schutzprofil (vgl. Abschnitt 3.2.1.2) herangezogen werden.

6.1.10. Löschung nicht mehr benötigter Daten

Die Analyse der letzten Phase des Datenlebenszyklus ergab, dass nicht mehr benötigte Energiedaten gelöscht werden sollten. Ein Energiedatenmanagementsystem sollte seine Nutzer daher regelmäßig dazu anhalten, den Datenbestand zu prüfen und gegebenenfalls auszudünnen. Darüber hinaus sollte eine Möglichkeit geboten werden, Daten auf Wunsch der Nutzer umgehend aus dem Energiedatenmanagementsystem zu entfernen.

Der *Data Custodian Service* dient als lokale und, im Idealfall, als einzige Energiedatensenke für seine Nutzer. Aus diesem Grund können die Nutzer davon ausgehen, dass aus dem *Data Custodian Service* gelöschte Daten – falls sie zuvor nicht weitergegeben wurden – endgültig gelöscht wurden und fortan nicht mehr zur Verfügung stehen. Der *Data Custodian Service* kann die Löschung der Daten auf zweierlei Weise unterstützen. Zunächst können die Nutzer darauf hingewiesen werden, dass historische Daten vorliegen, die über einen gewissen Zeitraum nicht verwendet wurden und dementsprechend gelöscht werden könnten. Darüber hinaus bietet der *Data Custodian Service* einen Katalog über die von den jeweiligen Nutzern bereitgestellten Daten (vgl. Abschnitt 5.4.2). Dieser ermöglicht es den Nutzern, zu überblicken, welche Daten dem System noch vorliegen. Die Nutzer können somit bewusste Entscheidungen über die Löschung der Daten treffen und haben zu jedem Zeitpunkt einen Überblick über alle sie betreffenden Daten.

Bei einer gewünschten Löschung der Daten bleiben die Metadaten erhalten und lediglich die Zeitreihendaten werden entfernt. Hierdurch können die

Nutzer auch zukünftig nachvollziehen, welche Daten von Dateninteressenten angefragt wurden und gegebenenfalls weitergegeben wurden.

Sind Daten aus genehmigten Datenanfragen allerdings bereits von den Dateninteressenten abgerufen worden, kann eine Löschung dieser weitergegebenen Daten technisch nicht ohne Weiteres erzwungen werden. Hierfür sollten nachgelagerte Maßnahmen für das digitale Rechtemanagement wie beispielsweise *Usage Control* zum Einsatz kommen. Dies ist im Rahmen des *Data Custodian Service* möglich, wie in den Abschnitten 6.1.8 und 6.1.14 erläutert.

Der *Data Custodian Service* unterstützt die Nutzer bei der Löschung nicht mehr benötigter oder unerwünschter Daten. Die Löschung betrifft jedoch nur die im System vorgehaltenen Daten, da eine Löschung bereits weitergegebener Daten technisch nicht ohne Weiteres erzwungen werden kann.

6.1.11. Einschränkung der Datenweitergabe

Es gibt eine Vielzahl unterschiedlicher Verfahren, um Energiedaten auszuwerten und aus ihnen Informationen zu extrahieren. Bei einer Datenweitergabe kann folglich nur schwer abgeschätzt werden, welche Verfahren möglicherweise zum Einsatz kommen werden und welche Informationen sich damit potenziell aus den Daten extrahieren lassen. Deshalb sollte ein Energiedatenmanagementsystem die Datenweitergabe so weit wie möglich einschränken.

Der *Data Custodian Service* gibt ohne explizite Zustimmung der Nutzer keine Energiedaten weiter. Darüber hinaus müssen Dateninteressenten gezielt Datenanfragen formulieren und können keine Generalerlaubnis für den Zugriff auf die Energiedaten eines Nutzers erlangen. Die Nutzer entscheiden von Fall zu Fall, ob die Energiedaten an die anfragende Partei weitergegeben werden dürfen. Die Zugriffsteuerung des *Data Custodian Service* geht damit in Summe über ein klassisches rollenbasiertes System hinaus.

Auch auf Seiten der Nutzer ist ein automatisiertes Bearbeiten eingehender Datenanfragen nicht vorgesehen. Durch diese Vorkehrung laufen Nutzer nicht Gefahr, aus Bequemlichkeit oder Unwissenheit Datenanfragen unbedacht zu erlauben. Stattdessen muss jede einzelne Datenanfrage von den betroffenen Nutzern aktiv geprüft und angenommen werden. Im Vergleich zu einer komplett automatisierten Lösung ist dies aber natürlich mit Komforteinbußen verbunden. Daher könnte als Kompromiss der Einsatz von Stellvertretern in

Erwägung gezogen werden, welche für die betroffenen Nutzer die Entscheidungen fällen.

Die Verwendung des *Data Custodian Service* ohne weitere Zwischeninstanzen bringt jedoch den Vorteil mit sich, dass es den Nutzern möglich ist, die Datenweitergabe extrem einzuschränken und bewusst für jede Anfrage zu entscheiden, ob diese genehmigt werden soll oder nicht.

6.1.12. Reduktion der Aussagekraft auf akzeptables Mindestmaß vor Weitergabe

Diese Anforderung ist eng verwandt mit der vorherigen: Da unklar ist, auf welche Art die Energiedaten potenziell ausgewertet werden, sollte nicht nur die Menge der freizugebenden Daten klein gehalten werden, sondern auch die Aussagekraft der Daten so weit wie möglich reduziert werden. Hierbei sollte ein Energiedatenmanagementsystem die Nutzer unterstützen.

Der *Data Custodian Service* nimmt eine mehrstufige Bewertung eingehender Datenanfragen vor. In die Beurteilung der Anfragen fließen unter anderem die zeitliche Auflösung, der angefragte Zeitraum, das Alter der Daten und die angefragte Kategorie ein. Die einzelnen Aspekte werden jeweils mit Gefährdungskennzahlen versehen, welche dem Nutzer wiederum über eine Ampel visualisiert werden (vgl. Abschnitt 5.4.5). Auf diese Weise können die von der Datenanfrage betroffenen Nutzer abschätzen, welche Aussagekraft die angefragten Daten möglicherweise haben könnten. Je kritischer die Bewertung des *Data Custodian Service* ausfällt, desto eher werden die Nutzer geneigt sein, die Datenanfragen abzulehnen. Falls Nutzer eine Datenanfrage ablehnen, können sie den Dateninteressenten den Grund dafür mitteilen und gegebenenfalls ein Gegenangebot machen. Aus Sicht der Dateninteressenten ist es folglich sinnvoll, abzuschätzen, welche Datenqualität beziehungsweise Aussagekraft für die Durchführung der geplanten Analyse tatsächlich erforderlich ist, und auch nur Daten in dieser Aussagekraft anzufragen. Eine übermäßig hohe Aussagekraft führt schließlich durch die damit verbunden kritischeren Bewertungen vermutlich eher zu Ablehnungen durch die betroffenen Nutzer.

Durch den Bewertungsmechanismus animiert der *Data Custodian Service* die Dateninteressenten, Daten nur in der Qualität beziehungsweise Aussagekraft

anzufragen, wie sie tatsächlich benötigt wird. Den Nutzern steht es dann frei, die Anfragen zu den festgelegten Konditionen zu akzeptieren oder eine weitere Verringerung der Aussagekraft zu fordern. Damit werden Daten stets nur mit einem für die Nutzer akzeptablen Qualitätsniveau weitergegeben.

6.1.13. Aufklärung über mögliche Auswirkungen einer Datenfreigabe auf die Privatsphäre

Auch diese Anforderung ist verwandt mit den beiden vorherigen: Ein Energie-datenmanagementsystem sollte seinen Nutzern dabei helfen, bewusste und aufgeklärte Entscheidungen über die Freigabe der Daten zu treffen.

Wie im Rahmen der Prüfung der vorherigen Anforderung erläutert, berechnet der *Data Custodian Service* eine Reihe von Gefährdungskennzahlen für eingehende Datenanfragen. Diese dienen nicht nur der Bewertung der Anfragen im Hinblick auf potenzielle Auswirkungen auf die Privatsphäre. Sie dienen vielmehr auch der Aufklärung der betroffenen Nutzer. So erlaubt beispielsweise eine hohe zeitliche Auflösung tendenziell Rückschlüsse auf verwendete Geräte, selbst wenn der angefragte Zeitraum relativ kurz ist. Umgekehrt erlaubt ein großer Zeitraum selbst bei geringer zeitlicher Auflösung potenziell die Erkennung von Verhaltensmustern (vgl. Abschnitte 5.3.4.1 und 5.3.4.2). Diese Aspekte sollten den Nutzern vor der Freigabe einer Datenanfrage bekannt und bewusst sein.

Einerseits dienen die vom *Data Custodian Service* berechneten und dem Nutzer angezeigten Kennzahlen bereits als erster Indikator für die möglichen Auswirkungen auf die Privatsphäre. Andererseits ist dies – gerade für wenig versierte Nutzer – jedoch möglicherweise nicht ausreichend. Deshalb zeigt der *Data Custodian Service* bei Überschreiten definierter Grenzwerte auch textuelle Hinweise und Erklärungen an, die den Nutzer auf die möglichen Konsequenzen hinweisen. Durch diese Mechanismen werden die Nutzer über potenzielle Auswirkungen einer Datenfreigabe auf die Privatsphäre aufgeklärt und können somit eine bewusste Entscheidung treffen.

6.1.14. Unterstützung zusätzlicher, externer Schutzmechanismen

Einmal weitergegebene Daten unterliegen nicht mehr der Kontrolle des Nutzers. Somit ist es beispielsweise nicht möglich, eine Löschung der Daten zu erzwingen oder auch nur die erneute Nutzung der Daten zu unterbinden. Mit Technologien für das digitale Rechtemanagement existieren Werkzeuge, um solche Ansprüche technisch durchzusetzen. Ein Energiedatenmanagementsystem sollte die Verwendung solcher Technologien auf Wunsch der Nutzer ermöglichen.

Der *Data Custodian Service* reduziert einerseits den Umfang und andererseits die Aussagekraft der Daten bereits auf ein von den betroffenen Nutzern akzeptiertes Maß. Es ist dennoch denkbar, dass Nutzer einen zusätzlichen Schutz der Daten wünschen, um beispielsweise eine unbefugte Weitergabe durch den Dateninteressenten an Dritte zu verhindern. Durch die flexible Wahl unterschiedlicher Ausgabeformate ermöglicht der *Data Custodian Service* auch den Einsatz von technologischen Schutzmechanismen wie beispielsweise *Usage Control* (vgl. Abschnitte 3.2.2.7 und 6.1.8). Hierfür müsste ein entsprechendes *AusgabefORMAT*-Objekt mit der in einer *Python*-Klasse hinterlegten Funktionalität im *Data Custodian Service* angelegt werden. Die freigegebenen Daten würden dann nicht direkt in eine (textbasierte) Datei geschrieben, sondern in einem entsprechenden *Usage-CONTROL*-Container hinterlegt. Dieser würde nicht nur die eigentlichen Zeitreihendaten enthalten, sondern auch Vorschriften zur Nutzung der Daten. Den Empfängern des Containers ist es dann nicht möglich, die Daten Verwendungszwecken zuzuführen, die vom betroffenen Nutzer nicht gestattet wurden. Vielmehr sind nur die in den Verwendungsvorschriften genehmigten Aktionen erlaubt, so dass hierdurch beispielsweise die Weitergabe der Daten verhindert werden kann.

Auch wenn der *Data Custodian Service* die entsprechenden Möglichkeiten für die Verwendung externer Schutzmechanismen bietet, scheint deren Einsatz nicht immer praktikabel. Vielmehr zielt dieses Energiedatenmanagementkonzept darauf ab, die Aussagekraft der Daten grundsätzlich so weit zu reduzieren, dass eine missbräuchliche Verwendung kaum noch möglich ist. Sollte dennoch auf Wunsch der Nutzer ein externer Schutzmechanismus zum Einsatz kommen, ist dies mit dem *Data Custodian Service* möglich.

6.1.15. Unterstützung verschiedener Endgeräte zur Darstellung der Benutzeroberfläche

Die Nutzung eines Energiedatenmanagementsystems sollte einfach und komfortabel sein. Dies impliziert, dass der Zugriff darauf mit den in den meisten Haushalten bereits vorhandenen Endgeräten möglich sein sollte.

Der *Data Custodian Service* ist eine Webanwendung, welche aktuelle Webtechnologien wie *HTML5* und *CSS3* verwendet. Diese ermöglichen es, Webseiten zu generieren, welche sowohl auf Geräten mit größerem Bildschirm – wie beispielsweise Laptops oder Tablets –, als auch auf Geräten mit kleinerem Bildschirm – wie Smartphones – komfortabel nutzbar sind. Hierbei handelt es sich um eine etablierte Technologie, welche bereits vielfach durch unterschiedlichste Webseiten eingesetzt wird.

Für die typischen Arbeitsabläufe des *Data Custodian Service* ist somit keine native Anwendung oder gar ein dediziertes Gerät nötig (vgl. Abschnitt 5.4). Vielmehr können die meisten modernen Browser verwendet werden, um die gesamte Webanwendung darzustellen und zu nutzen.

6.1.16. Zusammenfassung zur Erfüllung der Anforderungen

Wie in den vergangenen Abschnitten gezeigt, erfüllt der *Data Custodian Service* die in Kapitel 4 erarbeiteten Anforderungen. Einige wenige Anforderungen können jedoch nicht ohne das Einbeziehen zusätzlicher Schutzmechanismen wie beispielsweise *Usage Control* vollumfänglich erfüllt werden. Dies betrifft insbesondere die Sicherheit und den Schutz bereits weitergegebener Daten: Es ist dem *Data Custodian Service* ohne Verwendung weiterer technologischer Schutzmaßnahmen nicht möglich, eine erneute Weitergabe oder missbräuchliche Verwendung der Daten zu unterbinden. Diesem zentralen Problem begegnet das vorgestellte Energiedatenmanagementsystem auf zwei Arten. Zum einen werden Daten entweder überhaupt nicht oder nur mit einer für die betroffenen Personen akzeptablen Aussagekraft an die Dateninteressenten weitergegeben. Zum anderen ermöglicht der *Data Custodian Service* durch die flexible Unterstützung verschiedener *Ausgabeformate* eine einfache Integration zusätzlicher Schutztechnologien, die nachgelagert an das Energiedaten-

managementsystem zum Einsatz kommen können. Die Kombination dieser Maßnahmen führt zu einem effektiven Schutz der Energiedaten und ist in der Lage, den potenziellen Schaden einer missbräuchlichen Verwendung der Daten zu verringern.

6.2. Datenbereitstellung für typische Anwendungsfälle

Im Folgenden soll für eine Reihe typischer Anwendungsfälle geprüft werden, ob der *Data Custodian Service* in der Lage ist, die für die geplante Analyse nötige *Datenbasis* zu liefern. Die Analyse der Daten wird jedoch nach wie vor nicht im *Data Custodian Service*, sondern beim Dateninteressenten durchgeführt. Darüber hinaus soll auch geprüft werden, inwiefern die Privatsphäre der Nutzer geschützt werden kann. Energiedaten können, wie in Abschnitt 2.2.3 gezeigt, für viele unterschiedliche Zwecke von Nutzen sein. Die für die jeweiligen Anwendungsfälle erforderliche Aussagekraft unterscheidet sich jedoch zum Teil. Es ist deshalb keineswegs nötig, stets Daten von höchster Qualität und Aussagekraft zu liefern. Im Gegenteil können oft auch Daten von deutlich geringerer Aussagekraft für die angestrebten Analysen herangezogen werden. Dementsprechend wird neben der Eignung für typische Anwendungsfälle auch geprüft, inwieweit der *Data Custodian Service* eine Reduktion des Informationsgehalts der Daten ermöglicht. Als Referenz dienen hierbei Daten, wie sie ohne Verwendung des *Data Custodian Service* weitergegeben würden. Dies sind zum einen Zeitreihendaten aus intelligenten Zwischensteckdosen mit einer angenommenen zeitlichen Auflösung von 10 Sekunden³ und zum anderen Daten, wie sie ein gewöhnlicher intelligenter Zähler liefert. Hier beträgt die zeitliche Auflösung 15 Minuten.

³Solche Daten wurden beispielsweise mithilfe von *Plugwise*-Zwischensteckdosen im Rahmen eines Feldversuchs in rund zehn Haushalten erfasst.

6.2.1. Anwendungsfall 1: Aufzeigen der größten Verbrauchsposten

Für den ersten Anwendungsfall wird angenommen, dass ein Dienstleister Haushalten das Angebot unterbreitet, deren Stromverbrauch zu analysieren und ihnen mitzuteilen, welche Geräte den größten Anteil an der Stromrechnung haben. Insbesondere sollen die Ergebnisse mit Daten aus der Nachbarschaft in Relation gesetzt werden, um den Teilnehmern eine bessere Einschätzung geben zu können, ob und an welcher Stelle Einsparungen möglich wären. Dementsprechend ist eine rein lokale Durchführung der Analyse nicht möglich. Dieser Anwendungsfall ist dem Bereich der Transparenzsteigerung zuzuordnen (vgl. Abschnitt 2.2.3.1).

Für die Abschätzung des Stromverbrauchs einzelner Geräte sind räumlich hoch aufgelöste Daten nötig, damit möglichst viele unterschiedliche Verbraucher getrennt voneinander betrachtet werden können. Folglich müssten die Verbrauchsdaten beispielsweise aus intelligenten Zwischensteckdosen oder von den jeweiligen Geräten direkt stammen. Falls keine intelligenten Zwischensteckdosen vorliegen, wäre alternativ auch denkbar, dass es sich nicht um tatsächlich gemessene Verbrauchsprofile handelt, sondern dass im Vorfeld ein Verfahren zur Lastgangzerlegung zum Einsatz kommt. Ein solches könnte einen Gesamtlastgang in einzelne Teilprofile zerlegen, welche wiederum den vermutlichen Verursachern zugeordnet werden (vgl. Abschnitt 4.2). Unabhängig von ihrem Ursprung werden die Daten als einzelne Datensätze im *Data Custodian Service* mit den entsprechend zugewiesenen Kategorien geführt.

Die Daten liegen, im Falle von intelligenten Steckdosen beispielsweise, in einer zeitlichen Auflösung im Bereich von 10 Sekunden vor. Würden sie ohne weitere Reduktion der Aussagekraft weitergegeben, wären umfangreiche Rückschlüsse auf das Verhalten der Bewohner, die verwendeten Geräte und Ähnliches möglich. Die tatsächlich erforderlichen Daten können jedoch, wie im Folgenden erläutert, in stark reduzierter Aussagekraft durch den *Data Custodian Service* bereitgestellt werden, ohne das ursprüngliche Analyseziel zu gefährden.

Um einen geeigneten Datensatz für die geplante Analyse zu erhalten, müsste ein Dienstleister über den *Data Custodian Service* bei den jeweiligen Haushalten mehrere Datenanfragen einreichen. Diese sollten jeweils einen längeren Zeit-

6.2. Datenbereitstellung für typische Anwendungsfälle

raum – beispielsweise einen Monat – umfassen, damit die Verbrauchsdaten repräsentativ im Hinblick auf typische Nutzungsmuster sind. Darüber hinaus sollten die einzelnen Datenanfragen jeweils die entsprechenden Kategorien – beispielsweise Waschmaschinen, Trockner, Geschirrspüler und so weiter – abdecken. Anstatt nun jedoch eine hohe zeitliche Auflösung anzufragen, kann diese massiv reduziert werden. Es ist ausreichend, einen einzigen Wert pro Gerät für die gesamte Periode anzufragen: Aus diesem Durchschnittswert für den gesamten Zeitraum kann der Dateninteressent den Energieverbrauch berechnen – die Größe, welche ursprünglich untersucht werden sollte.

Durch den *Data Custodian Service* sind dessen Nutzer somit in der Lage, eine solche Dienstleistung in Anspruch zu nehmen, ohne jedoch Daten von unnötig hoher Aussagekraft weitergeben zu müssen. Der *Data Custodian Service* versieht die oben beschriebenen Anfrage mit einer deutlich kleineren Gefährdungskennzahl als beispielsweise eine Anfrage, welche Daten ohne jegliche Qualitätsreduktion umfasst. Dementsprechend können die Nutzer diejenigen Dienstleister präferieren, welche die gewünschten Analysen mit Datenanfragen mit der kleinsten Gefährdungskennzahl durchführen können. Zwar weisen die für diesen Anwendungsfall nötigen Daten durchaus eine hohe räumliche Auflösung auf, die tatsächlich erforderliche zeitliche Auflösung ist jedoch sehr gering. Würden hingegen die Rohdaten ohne Reduktion der Aussagekraft weitergegeben, wäre damit eine signifikante Gefährdung der Privatsphäre verbunden, da der Dateninteressent detaillierte Einblicke bekäme.

6.2.2. Anwendungsfall 2: Bestimmung der Grundlast in einem Haushalt

Der zweite hier vorgestellte Anwendungsfall beschäftigt sich erneut mit dem Stromverbrauch in einem Haushalt und der Steigerung der Transparenz. Ein fiktiver Dienstleister bietet Haushalten eine Prüfung deren Grundlast an, mit dem Ziel, aufzudecken, ob „versteckte“ Verbraucher für einen überdurchschnittlich hohen Stromverbrauch sorgen, auch wenn keine Geräte aktiv genutzt werden. Erneut findet ein Vergleich mit Daten aus der Nachbarschaft statt, um eine bessere Einschätzung der eigenen Ergebnisse zu ermöglichen.

Für diese Untersuchung sind zunächst keine räumlich hoch aufgelösten Energiedaten nötig. Es ist im Gegenteil sogar wahrscheinlich, dass viele der für die Grundlast verantwortlichen Geräte überhaupt nicht durch intelligente Zwischensteckdosen erfasst werden. Zwar tragen einerseits beispielsweise Kühlschränke und Gefriertruhen zur Grundlast bei, da diese durchgehend in Betrieb sind und somit in regelmäßigen Abständen Strom benötigen, um ihr Kühlaggregat zu betreiben. Diese Geräte könnten tatsächlich auch einfach einzeln gemessen werden. Andererseits gibt es jedoch eine Vielzahl anderer Geräte, die nur schwer einzeln gemessen werden können, welche aber dennoch permanent Strom verbrauchen. Dies sind zum Beispiel Umwälzpumpen, Bewegungsmelder, Überwachungskameras, Alarmanlagen, DSL-Router, Steuergeräte von Heizungen und viele weitere. Auch wenn der Energiebedarf der genannten Geräte im Vergleich zu den großen Verbrauchern in einem Haushalt gering erscheinen mag, addiert sich deren Energieverbrauch über die Zeit. Folglich kann es durchaus sinnvoll sein, zu prüfen, ob die Grundlast im Vergleich zu anderen Haushalten hoch ist und ob weitere, gezieltere Nachforschungen sinnvoll sind.

Die für die Durchführung dieser Analyse nötige zeitliche Auflösung ist ebenfalls nicht hoch. Eine Auflösung von vier Stunden dürfte in den meisten Fällen ausreichend sein: Man kann davon ausgehen, dass bei einem ausreichend großen Zeitraum – beispielsweise einer Woche – in mehreren Vierstundenblöcken keine Geräte aktiv genutzt werden und die Bewohner entweder schlafen oder außer Haus sind. Der vierstündige Zeitraum mit dem geringsten durchschnittlichen Verbrauch kann als Abschätzung für die Grundlast des Haushalts angesehen werden. Falls ein Haushalt eine auffällig hohe Grundlast aufweist, können die Bewohner darüber informiert werden und danach detailliert nachforschen, welche Geräte hierfür verantwortlich sein können.

Auch diese Datenanfrage ist mit dem *Data Custodian Service* möglich. Der Dienstleister könnte Zugriff auf das vom intelligenten Stromzähler gemessene Lastprofil für eine Woche beantragen. Dieses muss jedoch nicht die übliche zeitliche Auflösung von 15 Minuten aufweisen, sondern kann zeitlich auf vier Stunden aggregiert werden. Die an der fiktiven Untersuchung teilnehmenden Haushalte müssten somit weder zeitlich noch räumlich hoch aufgelöste Daten preisgeben. Sie können im Vergleich zu den unveränderten Rohdaten Daten mit geringerer Aussagekraft liefern und erhalten dennoch eine Abschätzung

ihrer Grundlast sowie die gewünschten Informationen über den Vergleich mit ähnlichen Haushalten.

6.2.3. Anwendungsfall 3: Auswahl des richtigen zeitvariablen Tarifs

Im dritten vorgestellten Anwendungsfall wird davon ausgegangen, dass zukünftig zahlreiche Energieversorger verschiedene zeitvariable Stromtarife anbieten. Solche Tarife sehen beispielsweise in Abhängigkeit des Wochentags und der Uhrzeit niedrigere und höhere Arbeitspreise oder gar mehrere Tarifstufen vor. Ein fiktiver Dienstleister bietet Interessenten die Möglichkeit an, ihren Stromverlauf im Hinblick auf potenziell geeignete dynamische Tarife zu prüfen und diesbezüglich Empfehlungen für einen Tarifwechsel zu geben.

Für die Abrechnung des Stromverbrauchs ist es unerheblich, welche Geräte den Strom verbraucht haben. Dementsprechend können für die in diesem Anwendungsfall durchzuführenden Untersuchung räumlich aggregierte Daten, wie beispielsweise von einem intelligenten Stromzähler, verwendet werden. Eine Aufschlüsselung nach einzelnen Verbrauchern mithilfe von intelligenten Zwischensteckdosen ist nicht erforderlich.

Die für die Analyse benötigte zeitliche Auflösung ist davon abhängig, welche zeitlichen Abstufungen die Tarife vorsehen. Gewöhnliche intelligente Stromzähler müssen beispielsweise alle 15 Minuten den Stromverbrauch übermitteln. In Zukunft könnte eine Reduktion auf 5 Minuten erfolgen, falls zum einen eine Reduktion der Handelsintervalle erfolgt und zum anderen die Messfrequenz im Hinblick darauf angepasst wird. Es kann folglich davon ausgegangen werden, dass auch zukünftig keine Tarife mit einer feineren zeitlichen Abstufung als 15 Minuten beziehungsweise 5 Minuten angeboten werden. Vielmehr ist denkbar, dass es sogar nur einige wenige Zeitblöcke von mehreren Stunden Länge gibt, in denen unterschiedliche Arbeitspreise gelten. Für diesen Fall würden auch Energiedaten mit einer zeitlichen Auflösung im Bereich von Stunden ausreichend sein.

Für eine sehr genaue Abschätzung, welcher Tarif am besten für einen Haushalt geeignet ist, sollten die Energiedaten ohne zusätzliche Reduktion der Datenqualität angefordert werden. Der Dienstleister könnte beispielsweise

das vom intelligenten Stromzähler gemessene Verbrauchsprofil für einen Zeitraum von einem Monat und mit einer zeitlichen Auflösung von einer Stunde anfragen. Mit diesen Informationen könnte sehr genau berechnet werden, welche Kosten dem Haushalt in den jeweiligen Tarifen entstanden wären. Andererseits ermöglichen diese relativ detaillierten Daten gegebenenfalls bereits Einblicke in das Verhalten der Personen des betrachteten Haushalts. Eine Reduktion der Aussagekraft ermöglicht es jedoch, die Privatsphäre der Personen zu Lasten der Güte der Stromtarifanalyse zu schützen.

Hierfür könnte der Dienstleister die Qualität der angefragten Daten vor einer Weitergabe durch den *Data Custodian Service* reduzieren lassen. Konkret kann dies durch eine Einschränkung der Messwertauflösung (vgl. Abschnitt 5.3.5.3) erfolgen. Diese Qualitätsreduktionsmaßnahme sieht vor, dass für jeden einzelnen Messwert geprüft wird, ob dieser kleiner als der durchschnittliche Wert der gesamten Zeitreihe ist, oder nicht. Kleinere Werte werden als niedrig, die anderen als hoch definiert. Der Dienstleister erhält nun keinen detaillierten Lastgang mehr, sondern nur noch eine Zeitreihe, deren Messwertauflösung massiv – auf zwei Ausprägungen – reduziert wurde. Mit diesen Daten ist offensichtlich keine zuverlässige Aussage über die tatsächlich entstehenden Kosten in den jeweiligen dynamischen Tarifen möglich. Der Dienstleister kann jedoch prüfen, ob die günstigen Zeiträume der Tarife mit den durchschnittlich hohen Bereichen des Lastprofils zusammenfallen, beziehungsweise ob Bereiche mit hoher Last mit den günstigeren Zeiträumen zusammenfallen. Aufgrund dieses Ergebnisses kann der Dienstleister zwar keinen genauen Betrag für die Kosten angeben, jedoch Tarife empfehlen, die gut zu dem typischen Stromverbrauch des Haushalts passen.

Der *Data Custodian Service* bietet seinen Nutzern damit auch für diesen Anwendungsfall die Möglichkeit, ihre Daten und ihre Privatsphäre zu schützen, und die Dienstleistung dennoch in Anspruch zu nehmen. In diesem Fall geht mit einem besseren Schutz der Privatsphäre jedoch auch eine Einschränkung der Ergebnisqualität einher, so dass die Nutzer abwägen müssen, welcher Aspekt ihnen wichtiger ist.

6.2.4. Zusammenfassung Anwendungsfälle

Die drei Anwendungsfälle, die in den vorherigen Abschnitten geprüft wurden, zeigen nur einen Ausschnitt der Möglichkeiten, die die Nutzung von Energiedaten eröffnet (vgl. Abschnitt 2.2.3). Der *Data Custodian Service* kann für viele Anwendungsfälle als Werkzeug dienen, um die Qualität und die Aussagekraft der Rohdaten gezielt auf ein angemessenes Niveau zu senken. In einigen Fällen ist dies ohne eine signifikante Reduktion der Ergebnisqualität möglich. In anderen müssen die Nutzer jedoch abwägen, ob sie eher ihre Privatsphäre schützen möchten, oder ob sie aussagekräftigere Ergebnisse präferieren. In jedem Fall bietet ein solches Energiedatenmanagementsystem den Nutzern jedoch überhaupt die Möglichkeit, selbst zu entscheiden, ob und welche Daten weitergegeben werden dürfen, falls der angestrebte Verwendungszweck dies wert ist.

6.3. Weitergehende Evaluationsmöglichkeiten

Das hier vorgestellte Werkzeug wurde bisher nur in Form eines Demonstrators implementiert. Es wäre jedoch denkbar, den Demonstrator als Ausgangsbasis für die Entwicklung eines Produkts heranzuziehen, welches im Rahmen eines Feldtests weiter ausgebaut wird (vgl. Abschnitt 5.4.7). Dies wäre insbesondere in Forschungsprojekten interessant, in welchen im großen Umfang Energiedaten aus Haushalten oder Unternehmen gesammelt werden. Im Rahmen solcher Projekte könnten über Befragungen der potenziellen Nutzer des geplanten Energiedatenmanagementsystems zunächst weitere Anforderungen erfasst werden. Diese könnten wiederum in die Erweiterung des hier vorgestellten Konzepts einfließen.

Ausgehend von dem in dieser Arbeit vorgestellten Demonstrator könnte im Rahmen eines Feldtests ein Prototyp entwickelt werden, welcher alle hier beschriebenen Funktionalitäten aufweist und insbesondere von Experten auf seine Sicherheit geprüft wurde. In diesem Prototyp könnten dann die tatsächlich gemessenen Energiedaten der Feldtestteilnehmer hinterlegt werden und verschiedenen realen Dateninteressenten die Möglichkeit geboten werden, Datenanfragen zu stellen. Die Verwendung von echten, die Nutzer betreffenden, Daten würde es ermöglichen, nicht nur das Konzept weiter zu evaluieren,

sondern auch das Verhalten der Nutzer im Hinblick auf den Umgang mit ihren Daten unter realistischen Rahmenbedingungen zu beobachten.

Darüber hinaus ließe sich mit einem solchen Produkt möglicherweise das Problem der Nutzung von Energiedaten nach Projektende für zu Beginn noch nicht bekannte Zwecke lösen. Oftmals werden umfangreiche Datenmengen in Projekten erfasst und gemäß den Projektzielen bearbeitet. Nach dem Ende des Projekts gestaltet sich deren weitere Nutzung jedoch oftmals schwierig, da die betroffenen Personen einer Weiterverwertung zustimmen müssten. Ein auf dem hier vorgestellten Konzept beruhendes Energiedatenmanagementsystem könnte dazu dienen, dieses Einverständnis bei den betroffenen Personen einzuholen und gleichzeitig die Daten zu verwalten. Im Idealfall ließe sich aus einem solchen System sogar ein Repository für Energiedaten aufbauen, welches eine umfangreiche Datensammlung enthält, wobei auch hier die betroffenen Personen oder aber ernannte Stellvertreter über die eingehenden Datenanfragen entscheiden dürfen.

Zusammenfassung und Ausblick

In dieser Arbeit wurde zunächst ein Katalog von Anforderungen an ein nutzerorientiertes Energiedatenmanagementsystem ausgearbeitet. Im Anschluss wurde ein Konzept für ein solches System vorgestellt und skizziert, wie eine mögliche Umsetzung erfolgen kann. In den nächsten Abschnitten dieses Kapitels folgt eine Zusammenfassung der wichtigsten Aspekte diesbezüglich, gefolgt von einer Einbettung der Arbeit in den aktuellen Stand der Wissenschaft. Abschließend werden in einem Ausblick Möglichkeiten aufgezeigt, die Ergebnisse dieser Arbeit im Rahmen weiterer Forschung aufzugreifen.

7.1. Zusammenfassung der Arbeit

Die Energiewende gilt als eine der größten Herausforderungen dieses Jahrhunderts. Es sind Anstrengungen in einer Vielzahl verschiedener Forschungsbereiche nötig, um eine zügige und effiziente Umsetzung zu ermöglichen. Hierzu trägt auch die Energieinformatik bei, welche das Ziel hat, durch eine geschickte Nutzung von Informations- und Kommunikationstechnologien ein nachhaltiges Energiesystem zu schaffen. In diesem Kontext können auch Energiedaten dabei helfen, das zukünftige Energiesystem intelligenter und effizienter zu machen. Sie können beispielsweise für verbesserte Verbrauchsprognosen oder als Basis für eine Optimierung des Lastprofils in Haushalten oder Unternehmen herangezogen werden. Darüber hinaus können sie den Verbrauchern helfen, ihre Energienutzung besser zu verstehen und Einsparpotentiale zu erkennen und zu nutzen. Energiedaten sind jedoch meist auch sensible Daten, die besonders geschützt werden müssen. Gleichzeitig muss bei der Handhabung und Verwaltung von Energiedaten auch eine Reihe von technischen

Herausforderungen bewältigt werden. Im nächsten Abschnitt wird daher die Ausarbeitung eines Anforderungskatalogs zusammengefasst, der die genannten Aspekte berücksichtigt. Anschließend erfolgt eine Zusammenfassung des in dieser Arbeit vorgestellten Konzepts, welches die Anforderungen erfüllt.

7.1.1. Zusammenfassung der Anforderungsanalyse

Die in dieser Arbeit durchgeführte Anforderungsanalyse umfasst sowohl eine technische Datenlebenszyklusanalyse als auch eine Untersuchung der nutzergetriebenen Anforderungen. Aus den Betrachtungen resultiert ein Katalog von fünfzehn Anforderungen, die ein nutzerorientiertes Energiedatenmanagementsystem erfüllen sollte.

Technische Datenlebenszyklusanalyse

Das im Rahmen der technischen Analyse des Datenlebenszyklus verwendete Modell sieht sechs verschiedene Phasen vor, welche jedoch nicht strikt voneinander getrennt vorliegen müssen. Hierbei handelt es sich um die Erfassung, die Übertragung, die Speicherung, die Analyse, die Verteilung und die Löschung der Daten. Für jede einzelne dieser Phasen wurden Herausforderungen im Hinblick auf die Verwaltung und Handhabung von Energiedaten identifiziert. Daraus wurden wiederum technische Anforderungen formuliert, welche für die Konzeption eines Energiedatenmanagementsystems herangezogen werden sollten.

So entstammen Energiedaten beispielsweise vielen unterschiedlichen Datenquellen, werden in verschiedenen Dateiformaten vorgehalten und sind oft fehlerbehaftet oder weisen Lücken auf. Dementsprechend müssen Energiedatenmanagementsysteme äußert flexibel im Hinblick auf einzulesende und auszugebende Energiedaten sein, um eine gute Interoperabilität mit anderen Systemen zu gewährleisten. Gleichzeitig muss eine effiziente Speicherung, sowohl der eigentlichen Zeitreihendaten, als auch der dazugehörigen Metadaten, möglich sein. Des Weiteren müssen Energiedatenmanagementsysteme auch über Funktionen zur Manipulation von Energiedaten verfügen. Hierzu zählen etwa die Selektion von Datenausschnitten, die Umrechnung in andere Abtastraten oder simple Additionen. Nicht zuletzt ist auch erforderlich,

dass das Energiedatenmanagementsystem Vorkehrungen zum Schutz der Kommunikation mit seinen Interaktionspartnern trifft.

Nutzergetriebene Anforderungen

Der technischen Lebenszyklusanalyse schloss sich eine Untersuchung der nutzergetriebenen Anforderungen an. Als Nutzer eines Energiedatenmanagementsystems werden in dieser Arbeit diejenigen Personen oder auch Unternehmen bezeichnet, die von der Erfassung von Energiedaten betroffen sind. Es sollte ein vorrangiges Ziel sein, die Privatsphäre der Dateneigner zu wahren und den Datenschutz zu gewährleisten. Dies führt zu einem Interessenkonflikt im Hinblick auf die Dateninteressenten – jene externen Parteien, die möglichst umfangreiche Daten für ihre Zwecke anfragen und auswerten möchten.

In der jüngeren Vergangenheit wurde eine große Zahl unterschiedlichster Verfahren zur Informationsextraktion aus Energiedaten entwickelt. Hierbei werden verschiedene Ansätze verfolgt: Angefangen bei der Erkennung charakteristischer Kanten im Lastprofil, über die Verwendung von *Markov*-Modellen, künstlichen neuronalen Netzen oder genetischen Algorithmen, bis hin zur Auswertung von Oberwellen im hochauflösten Verlauf der Spannung. Die einzelnen Verfahren unterscheiden sich nicht nur hinsichtlich der Ansätze, sondern auch in der benötigten Datenbasis für die Analyse, sowie dem Ziel, welches mit der Auswertung verfolgt wird. Für die Dateneigner, deren Daten an externe Parteien weitergegeben werden sollen, ist es folglich schwierig, einzuschätzen, welche Informationen sich potenziell aus den Daten extrahieren lassen. In der Literatur wird für viele unterschiedliche Beispiele aufgezeigt, welche Aussagen möglich sind: Hinweise auf die Zusammensetzung eines Haushalts, An- und Abwesenitserkennung, sogar Informationen über das konsumierte Fernsehprogramm und vieles mehr.

Dem Energiedatenmanagementsystem fällt in diesem Kontext die Aufgabe zu, die Privatsphäre und den Datenschutz seiner Nutzer zu wahren. Gleichzeitig sollten die Daten jedoch auch im Sinne der Nutzer verwendet und weitergegeben werden können, sowie dabei helfen, die Energiewende voranzutreiben. Ein Energiedatenmanagementsystem darf daher die Daten nur in einem, für seine Nutzer, akzeptablen Umfang mit möglichst geringerer Aussagekraft an externe Parteien weitergegeben – und das auch nur nach

expliziter Zustimmung durch die betroffenen Personen. Diese müssen wiederum im Vorfeld vom Energiedatenmanagementsystem über potenzielle Folgen einer Datenfreigabe aufgeklärt und damit befähigt werden, eine bewusste Entscheidung über die Weitergabe der eigenen Daten zu treffen. Gleichzeitig sollte ein solches Energiedatenmanagementsystem für seine Nutzer einfach und komfortabel zu verwenden zu sein, so dass es auch im Alltag zum Einsatz kommen kann.

7.1.2. Zusammenfassung zum entwickelten Konzept und zu dessen Umsetzung

Nach der Erstellung des Anforderungskatalogs wurde ein Konzept für ein nutzerorientiertes Energiedatenmanagementsystem entworfen. Es umfasst sowohl die Softwarearchitektur, als auch ein Datenmodell zur Abbildung der Beziehungen der einzelnen Entitäten und Akteure. Im Rahmen der Arbeit wurde des Weiteren die Implementierung des Konzepts in Form eines Demonstrators unter Verwendung etablierter, quelloffener Bibliotheken und Anwendungen beschrieben.

Architektur des Energiedatenmanagementsystems

Das *Data Custodian Service* genannte Konzept sieht einen modularen Aufbau des Energiedatenmanagementsystems vor, in welchem die einzelnen Komponenten jeweils bestimmte Aufgabenbereiche übernehmen. Ein zentrales Modul übernimmt dabei die Orchestrierung des Zusammenspiels der anderen Komponenten. Dies sind unter anderem Schnittstellen für die Kommunikation mit den Nutzern und Dateninteressenten, ein Modul für die Verarbeitung von Zeitreihen, eines für die Bewertung eingegangener Datenanfragen, eine Komponente für den Datenbankzugriff, sowie Datenbanken für die Speicherung von Zeitreihen und Metadaten. Jede Komponente übernimmt einen spezifischen Aufgabenbereich und kann von den anderen Modulen genutzt werden, um die typischerweise anfallenden Arbeitsabläufe zu bewältigen: Einlesen neuer Energiedaten, Entgegennehmen und Bearbeiten eingehender Datenanfragen, Bewertung der Datenanfragen und gegebenenfalls Freigabe der Daten an die Dateninteressenten.

Verwendetes Datenmodell

Der *Data Custodian Service* nutzt eine relationale Datenbank, um die Metadaten zu den hinterlegten Energiedaten, sowie die damit verbundenen Informationen zu speichern. Hierfür wurde ein Datenmodell entworfen, welches die Beziehungen der einzelnen Entitäten zueinander abbildet. Dadurch werden beispielsweise die im Energiedatenmanagementsystem hinterlegten Energiedaten den jeweiligen Nutzern zugeordnet. Darüber hinaus wird unter anderem vermerkt, welcher Dateninteressent eine Datenanfrage mit einer festgelegten Datenqualität an einen Nutzer gestellt hat, und wie sich dieser entschieden hat. Dies ermöglicht es den Nutzern, zu jedem Zeitpunkt einen Überblick über die sie betreffenden Energiedaten sowie alle damit zusammenhängenden Anfragen zu erhalten.

Umgang mit Datenanfragen

Dateninteressenten können nicht ohne explizite Erlaubnis der betroffenen Nutzer auf deren Energiedaten zugreifen. Stattdessen müssen sie beim *Data Custodian Service* eine Datenanfrage einreichen, in der unter anderem vermerkt ist, welcher Zeitraum angefordert wird, welche zeitliche Auflösung die Daten haben sollen und welchem Datenquellentyp die Daten entstammen sollen. Darüber hinaus können zusätzliche Maßnahmen zur Reduktion der Aussagekraft der Daten aktiviert werden, so dass der potenzielle Schaden für die Privatsphäre, der mit jeder Datenfreigabe einhergeht, reduziert werden kann.

Die eingehenden Datenanfragen werden den betroffenen Nutzern inklusive einer Bewertung der möglichen Gefährdung für die Privatsphäre vorgelegt. Für die Bewertung einer Datenanfrage beurteilt der *Data Custodian Service* die gewünschte zeitliche Auflösung, die Größe des angefragten Zeitraums, das Alter der Daten, sowie die Kategorie, aus welcher die Daten stammen. Für jeden Aspekt wird eine Kennzahl berechnet und mithilfe einer Ampel codiert: Eine rote Ampel bedeutet, dass aufgrund des beurteilten Aspekts eine deutliche Gefahr für die Privatsphäre anzunehmen ist. Eine gelbe Ampel deutet an, dass diese Gefahr zwar noch vorhanden, aber nicht sehr stark ausgeprägt ist, und eine grüne Ampel signalisiert, dass die Gefahr vermutlich gering sein dürfte. Die einzelnen Kennzahlen werden zudem auf eine Ge-

samtbewertung reduziert, an der sich der Nutzer auf einen Blick orientieren kann. Den Dateninteressenten wird beim Erstellen einer Datenanfrage die Möglichkeit geboten, zusätzliche Maßnahmen zu aktivieren, die die Aussagekraft der Daten künstlich reduzieren. Im Gegenzug wird dafür die vom *Data Custodian Service* berechnete Gefährdungskennzahl reduziert, so dass die Nutzer gegebenenfalls eher geneigt sind, der Datenanfrage zuzustimmen. Dieser Mechanismus fördert, dass Dateninteressenten vor einer Anfrage festlegen, welche Analysen sie mit den gewünschten Daten durchführen möchten und welche Aussagekraft der Daten sie dafür tatsächlich benötigen. Es sollte dann auch nur dieses Minimum an Datenqualität angefragt werden, um nicht Gefahr zu laufen, dass Nutzer durch eine für sie zu hohe Gefährdungskennzahl die Datenanfrage ablehnen. Für den Fall, dass ein Nutzer die Datenanfrage dennoch ablehnen möchte, kann er dem Dateninteressenten einen Hinweis zukommen lassen, warum dies geschehen ist – beispielsweise wegen einer zu hohen zeitlichen Auflösung. Der Dateninteressent kann daraufhin nachbessern und eine neue Datenanfrage mit reduzierter Aussagekraft der Daten stellen. Durch diesen Prozess kann der Interessenkonflikt zwischen Nutzern und Dateninteressent entschärft werden. Die beiden Parteien können auf diese Weise erreichen, dass einerseits der Dateninteressent die gewünschten Daten erhält, aber diese andererseits eine soweit reduzierte Aussagekraft aufweisen, dass es für die betroffenen Nutzer akzeptabel ist.

Umsetzung in einem Demonstrator

Im Rahmen der Arbeit wurde das entworfene Konzept in Form eines Demonstrators umgesetzt. Der *Data Custodian Service* wurde in der Programmiersprache *Python* implementiert und nutzt ausschließlich quelloffene Anwendungen und Bibliotheken. Im Kern des Demonstrators wird mit *Django* ein etabliertes *Web-Application-Framework* verwendet. *Django* erlaubt es, das in der Arbeit entworfene Datenmodell und die vorgeschlagene Softwarearchitektur umzusetzen. Für die Speicherung der Metadaten kommt eine relationale Datenbanklösung wie beispielsweise *SQLite* zum Einsatz. Die Energiedatenzeitreihen werden in einem, auf große Mengen von tabellenartigen Daten spezialisierten, Speichersystem namens *HDF5* vorgehalten. Die Verarbeitung und Manipulation der Zeitreihen wird durch die ursprünglich aus dem Finanzbereich

stammende Bibliothek *Pandas* übernommen. Die Kommunikation mit den Nutzern beziehungsweise Dateninteressenten erfolgt ausschließlich über einen dedizierten Webserver. Die unter Verwendung von *HTML5* und *CSS3* erstellte Oberfläche ist in allen modernen Browsern verwendbar.

Der Demonstrator wurde in Form einer *Proof-of-Concept*-Implementierung umgesetzt und ist daher nicht unmittelbar für einen Produktiveinsatz vorgesehen. Dennoch ist er geeignet, um als Ausgangsbasis für die Entwicklung eines entsprechenden Produktes für den Realwelteinsatz zu dienen: Im Rahmen dieser Arbeit wurde das zugrundeliegende Konzept entwickelt und beschrieben, eine prototypische Realisierung vorgenommen, sowie eine Evaluation durchgeführt. Darüber hinaus wurde beschrieben, welche Schritte nötig wären, um den Demonstrator in ein für den Realwelteinsatz geeignetes Produkt zu überführen.

Evaluation des Konzepts

Das in der Arbeit vorgestellte Konzept wurde zunächst im Hinblick auf die Erfüllung der einzelnen zuvor ausgearbeiteten Anforderungen geprüft. Die Evaluation hat ergeben, dass alle Anforderungen des Anforderungskatalogs erfüllt werden, wobei für eine vollständige Erfüllung mancher Anforderungen der Einsatz zusätzlicher, externer Schutztechnologien nötig ist. Dieser Aspekt betrifft ein Kernproblem bei der Weitergabe sensibler Daten: Sobald solche Daten weitergegeben werden, unterliegen sie nicht mehr ohne Weiteres der Kontrolle der Nutzer. Die Datenempfänger können die Daten dann – auch entgegen der Nutzerwünsche – auswerten und gegebenenfalls an Dritte weitergeben.

Diesem Aspekt begegnet das in dieser Arbeit erstellte Konzept folgendermaßen. Zunächst wird die Weitergabe der Daten massiv eingeschränkt. Daten werden lediglich nach ausdrücklicher Zustimmung der betroffenen Personen weitergegeben. Dies jedoch nur, nachdem die Nutzer über die möglichen Risiken aufgeklärt wurden. Des Weiteren wird die Aussagekraft und Datenqualität vor einer eventuellen Weitergabe der Daten auf ein für die Nutzer akzeptables Maß reduziert. Diese Vorkehrungen sorgen bereits dafür, dass der potenziell mögliche Schaden reduziert wird. Für den Fall, dass die Nutzer darüber hinaus gehenden Schutz wünschen, können zusätzliche, externe

Schutztechnologien in Form von Systemen für das digitale Rechtemanagement zum Einsatz kommen. Diese sind eine Ergänzung für das vorgestellte Konzept und können durch den Einsatz entsprechender Technologien eine unbefugte Weitergabe beziehungsweise eine unerwünschte Nutzung verhindern. Durch eine Kombination mit diesen Werkzeugen kann der *Data Custodian Service* alle gestellten Anforderungen vollständig erfüllen.

In einem weiteren Schritt wurde geprüft, ob das Konzept geeignet ist, für eine Reihe von typischen Anwendungsfällen die notwendigen Daten für die gewünschten Analysen zu liefern. Die Untersuchung des Konzepts, auch mithilfe des Demonstrators, ergab, dass der *Data Custodian Service* in der Lage ist, die entsprechenden Datenanfragen zu bearbeiten und die Daten zu liefern. Im Vergleich zu den nicht veränderten Rohdaten kann die Datenqualität und Aussagekraft zum Teil deutlich reduziert werden, ohne die geplanten Analysen zu beeinträchtigen. Hierdurch können die Nutzer die gewünschten Dienstleistungen in Anspruch nehmen, aber gleichzeitig – im Vergleich zur Ausgangssituation ohne Schutz der Energiedaten – die Gefährdung ihrer Privatsphäre reduzieren. Währenddessen vermittelt der *Data Custodian Service* unter anderem über die numerische und farblich codierte Darstellung der Gefährdungsbewertung den Nutzern eine Einschätzung der potenziellen Risiken im Hinblick auf die Privatsphäre, so dass die Nutzer abschätzen können, welche möglichen Folgen mit einer Datenweitergabe einhergehen können.

7.2. Einbettung in den Stand der Wissenschaft

In dieser Arbeit wurde ein Konzept für ein *nutzerorientiertes* Energiedatenmanagementsystem vorgestellt. Bei der Entwicklung standen aus Nutzersicht vor allem der Datenschutz und die Wahrung der Privatsphäre im Vordergrund. Es gibt zahlreiche verwandte Arbeiten, die sich ebenfalls dem technologischen Schutz von Energiedaten widmen (vgl. Abschnitt 3.2.2). Im Fokus sind dort jedoch häufig spezielle Anwendungsfälle, weshalb die dort vorgestellten Technologien oft auch nur für diese eingesetzt werden können. So erfordern beispielsweise Verfahren, welche Zählerdaten vieler Haushalte verrauschen und aggregieren, eine Mindestzahl von Teilnehmern, um die Sicherheit der Daten

zu gewährleisten – eine Anwendung für eine sehr kleine Zahl von Haushalten ist folglich nicht möglich [vgl. u.a. BSU10; ÁC11]. Andere Ansätze nutzen vertrauenswürdige Dritte, um die Privatsphäre der betroffenen Personen zu schützen [vgl. u.a. DK13; EK10]. Dies verlagert die Verantwortung jedoch meist lediglich auf eine andere Entität und löst nicht das Kernproblem [JJR11]. Wiederum andere Ansätze nutzen homomorphe Verschlüsselungsverfahren, um Berechnungen nicht auf Klartextdaten, sondern auf verschlüsselten Daten durchzuführen. Allerdings kann hierbei nur auf eine begrenzte Menge von Rechenoperationen zurückgegriffen werden [vgl. u.a. RD11; GJ11]. Folglich können solche spezialisierten Ansätze nicht als alleinige Lösung in allen Anwendungsfällen erfolgreich zum Einsatz kommen. Die vorliegende Arbeit hat dennoch nicht das Ziel, die zuvor angeführten Verfahren zu ersetzen. Im Gegenteil sieht das entworfene Konzept vor, solche Methoden als sinnvolle *Ergänzung* eines Energiedatenmanagementsystems zu nutzen und damit einen noch besseren Schutz der Privatsphäre zu erreichen. Beispielsweise können Technologien wie *Usage Control* (vgl. Abschnitt 3.2.2.7 und [PHB06; Pre+08]) oder *SMART-ER* (vgl. Abschnitt 3.2.2.2 und [FB14b]) im Anschluss an eine Reduktion der Aussagekraft beziehungsweise der Datenqualität durch den *Data Custodian Service* zum Einsatz kommen. Hierdurch wird einerseits in Abstimmung mit den betroffenen Personen die Aussagekraft der Daten auf ein akzeptables Maß reduziert und andererseits eine nachgelagerte Schutztechnologie verwendet. Damit treten die Verfahren nicht in Konkurrenz zueinander, sondern ergänzen sich und verstärken damit in Summe den Schutz der Energiedaten.

Die Nutzerorientierung des vorgestellten Konzepts äußert sich noch in einem weiteren Aspekt. Gegenwärtig können von einer Energiedatenerfassung betroffene Personen nur sehr eingeschränkt entscheiden, was mit ihren Daten geschieht. Allenfalls steht ihnen die Möglichkeit offen, eine Weitergabe und Nutzung der sie betreffenden Daten zu unterbinden. Ein explizites Einbeziehen der Nutzer in einen feingranularen Entscheidungsprozess ist jedoch in den meisten verwandten Arbeiten nicht vorgesehen. Die *Platform for Privacy Preferences* (P3P) ist zwar nicht für den Einsatz mit Energiedaten entworfen worden, bildet jedoch bezüglich des Einbeziehens der Nutzer eine Ausnahme (vgl. Abschnitt 3.2.2.6 und [Wor02]). Der P3P-Ansatz sieht vor, dass Besuchern von Webseiten vermittelt wird, welche personenbezogenen

Daten erfasst und wie diese verwendet werden. Über von den betroffenen Personen festgelegte Regeln wird entschieden, ob die besuchte Webseite die Daten erhalten darf oder nicht. Im Hinblick auf Energiedaten greift der *Data Custodian Service* diese Idee auf und erweitert sie nochmals deutlich: Die von einer Energiedatenanfrage betroffenen Personen können nicht mehr nur eine binäre Wahl zwischen Datenweitergabe und Abschottung treffen. Vielmehr ermöglicht es der *Data Custodian Service* Dateninteressenten und Nutzern, einen Verhandlungsprozess anzustoßen, im Rahmen dessen die Aussagekraft der Daten beziehungsweise die Datenqualität feingranular festgelegt werden kann. Durch eine entsprechende Aufbereitung der Datenanfragen erfahren die Nutzer gleichzeitig, welche Auswirkungen eine Datenweitergabe in der festgelegten Aussagekraft möglicherweise haben könnte. Hierdurch werden die betroffenen Personen aufgeklärt und explizit in den Entscheidungsprozess einbezogen: Sie können von Fall zu Fall, von Dateninteressent zu Dateninteressent, feingranular entscheiden, welche Daten mit welcher Aussagekraft weitergegeben werden dürfen. Bevor sie eine Freigabe erteilen, verstehen die Nutzer die möglichen Folgen der Datenweitergabe und können damit bewusst entscheiden, ob der erhoffte Nutzen das Risiko wert ist. Dieses Restrisiko kann wiederum, wie zuvor beschrieben, durch die Kombination des Energiedatenmanagementsystems mit Schutztechnologien, wie sie in den verwandten Arbeiten aufgezeigt wurden, weiter reduziert werden.

Im Hinblick auf die technischen Herausforderungen wird beim *Data Custodian Service* ein Kompromiss zwischen komplexen und mächtigen Data-Warehouse-Systemen wie beispielsweise *MIRABEL DW* [STP12; Boe+12] und einfachen, weniger mächtigen dateibasierten Speicherlösungen wie beispielsweise *REDD* [KJ11] gewählt (vgl. auch Abschnitt 3.1). Mit einer mächtigen Suchanfragesprache ausgestattete Data-Warehouse-Systeme erlauben zwar ausgefeilte Datenanfragen, bringen jedoch gleichzeitig eine deutlich höhere Komplexität mit sich. Dies kann den Schutz der Privatsphäre erschweren, da durch geschickte Formulierung der Datenanfragen potenziell die Schutzmechanismen des Energiedatenmanagementsystems umgangen werden könnten. Darüber hinaus gestaltet sich die Datenhaltung als deutlich aufwendiger und die Bearbeitung von Anfragen kann rechenintensiver sein. Das andere Extrem, die einfache dateibasierte Datenhaltung, ermöglicht wiederum überhaupt keine Anfragen, sondern dient lediglich als Datenspeicher. Der *Data Custodian*

Service stellt den Mittelweg zwischen diesen beiden Extremen dar. Er bietet deutlich mehr Möglichkeiten als nur eine reine Datenhaltung, ist jedoch gleichzeitig weniger mächtig und damit auch deutlich weniger komplex als eine Data-Warehouse-Lösung. Im Vergleich zu diversen kommerziellen Lösungen bietet der *Data Custodian Service* eine größere Flexibilität im Hinblick auf unterstützte Energiedatenquellen und Ausgabeformate. Es ist nachvollziehbar, dass Dienstleister – zumindest zunächst – lediglich Quellen unterstützen, die sie selbst vertreiben oder anbieten [vgl. u.a. Dis16; Plu16]. Aus Nutzersicht ist es jedoch von Vorteil, wenn von einem einzelnen Energiedatenmanagementsystem potenziell beliebige Datenquellen unterstützt werden, so dass dieses System als alleinige Datensenke dienen kann.

Zusammenfassend handelt es sich bei dem in dieser Arbeit vorgestellten *Data Custodian Service* um ein Energiedatenmanagementsystem, welches die von der Energiedatenerfassung betroffenen *Personen* – die Nutzer – in den Vordergrund stellt. Dies impliziert, dass diese intensiv in die zu treffenden Entscheidungen mit einbezogen werden und in die Lage versetzt werden, bewusste Entscheidungen über den Umgang mit ihren Energiedaten zu treffen. Darüber hinaus wird dem Datenschutz und der Wahrung der Privatsphäre besonderer Wert beigemessen. Hierfür wird nicht nur auf die Reduktion der Datenaussagekraft gesetzt, sondern es kann bei Bedarf auch auf weitere Schutzmechanismen zurückgegriffen werden. Währenddessen werden auch die technischen Herausforderungen nicht außer Acht gelassen, so dass ein auf dem entworfenen Konzept basierendes Energiedatenmanagementsystem auch für die Handhabung von Realweltdaten gerüstet ist.

7.3. Ausblick

Im Rahmen dieser Arbeit wurde ein Demonstrator implementiert und vorgestellt. Aus diesem könnte zukünftig eine, für den Produktiveinsatz geeignete, Anwendung hervorgehen. Bis zu einem solchen fertigen, sicheren Produkt ist jedoch viel Entwicklungs- und Implementierungsarbeit nötig. Der Demonstrator könnte deshalb als Ausgangsbasis für eine solche Umsetzung dienen. Die daraus hervorgehende Anwendung könnte in einem größeren Feldtest erprobt werden. Durch einen solchen Test unter Realweltbedingungen könnten

neue Anforderungen aufkommen, welche wiederum in die Weiterentwicklung des Energiedatenmanagementsystems einfließen könnten. Darüber hinaus sollte dieses Produkt – ebenfalls unter realistischen Bedingungen und von entsprechenden Experten – eingehenden Sicherheitsprüfungen unterzogen werden.

Ebenfalls im Rahmen eines Feldtests könnte geprüft werden, welche Anreize den Dateneignern geboten werden müssten, damit sie bereit wären, ihre Daten in definierten Abstufungen von unterschiedlicher Aussagekraft zu teilen. Der potenziell denkbare Schaden, der den betroffenen Personen entstehen kann, ist in der Literatur hinreichend beschrieben. Es sind zahlreiche Negativbeispiele für die missbräuchliche Verwendung von Energiedaten bekannt. Im Gegensatz dazu sind Anwendungsfälle, die einerseits eine Datenweitergabe zwingend erfordern und andererseits den Nutzern dienlich sind, jedoch deutlich weniger prominent aufgeführt. Dementsprechend wäre in diesem Kontext die Untersuchung der nötigen Anreizmechanismen interessant.

Möglicherweise kann das in dieser Arbeit vorgestellte Prinzip auch dazu beitragen, die Situation zur Weiternutzung von zunächst projektbezogen erhobenen Forschungsdaten zu verbessern. Ein entsprechendes Energiedatenmanagementsystem könnte als Sammelstelle für alle im Rahmen eines Projektes angefallenen Energiedaten dienen. Nach Projektende könnten Dateninteressenten Anfragen an die betroffenen Personen richten und die ansonsten nicht mehr nutzbaren Daten weiteren Analysen zuführen, falls die Daten freigegeben werden. Die im ursprünglichen Projekt beteiligten Wissenschaftler könnten in Zuge dessen möglicherweise auch als eingesetzte Stellvertreter handeln und die Weiternutzung der Daten koordinieren.

Abschließend bleibt zu hoffen, dass Energiedaten künftig hauptsächlich im Sinne der Dateneigner genutzt werden und eine missbräuchliche Verwendung so gut wie möglich unterbunden wird. Diese Arbeit leistet diesbezüglich einen Beitrag und stellt ein Konzept vor, welches es den betroffenen Personen ermöglicht, bewusst eine feingranulare Entscheidung über die bei ihnen eingehenden Datenanfragen zu treffen, dabei das Risiko und den Nutzen einer Datenweitergabe abzuwägen und schließlich entsprechend der persönlichen Präferenzen zu handeln.

Umsetzung des Datenmodells

Dieser Anhang zeigt, wie das in Abschnitt 5.2 beschriebene Datenmodell in *Django* mithilfe der integrierten objektrelationalen Abbildung umgesetzt wird. Die unten verwendete Notation orientiert sich dabei an einem *UML*-Klassendiagramm. Zu den einzelnen Entitäten sind jeweils deren Attribute sowie die dazugehörigen *Django Model Fields* aufgeführt. Hierbei handelt es sich um spezielle *Python*-Datentypen beziehungsweise -objekte in *Django*, die jeweils ein entsprechendes Äquivalent in der verwendeten relationalen Datenbank haben. Dies hat den Vorteil, dass einerseits in *Django* mit gewöhnlichen *Python*-Objekten gearbeitet werden kann und andererseits die definierten Restriktionen und Relationen auch in der Datenbank geprüft und durchgesetzt werden. In der Dokumentation von *Django* [Dja16b] sind ausführliche Informationen zu den einzelnen Feldtypen zu finden. Dennoch werden die im Demonstrator verwendeten Typen im Folgenden kurz erklärt.

AutoField Ein *Integer*-Feld, welches von *Django* selbstständig jeder Entität hinzugefügt wird. Hierin wird der automatisch erzeugte Primärschlüssel abgelegt.

CharField Ein Feld, um *Strings* abzuspeichern. Längere Texte sollten in einem *TextField* hinterlegt werden.

DateTimeField In diesem Feld können Zeitstempel – bestehend aus Datum, Uhrzeit und Zeitzone – hinterlegt werden.

FileField Ein Feld, um die Metadaten zu hochgeladenen Dateien abzuspeichern. Die Dateien selbst werden im normalen Dateisystem hinterlegt und

können über die Informationen aus diesem Feld referenziert beziehungsweise gelesen werden.

FloatField Hierbei handelt es sich um ein Feld für die Speicherung einer Gleitkommazahl.

ForeignKey Über dieses Feld kann eine Fremdschlüsselbeziehung definiert werden. Referenziert wird der Primärschlüssel des gewünschten Objekts, welcher in einem *AutoField* hinterlegt ist.

IntegerField Ein Feld zum Abspeichern eines *Integer*-Werts.

ManyToManyField Ein Feld, um eine *many-to-many*-Beziehung zu formulieren. *Django* erzeugt transparent die benötigten Hilfstabellen, um eine Zuordnung der jeweiligen Paare zueinander zu ermöglichen.

TextField Hiermit können wie in *CharFields* ebenfalls *Strings* gespeichert werden, jedoch sind diese Felder für längere Texte ausgelegt.

Nachfolgend sind für die Entitäten aus dem *Entity-Relationship*-Diagramm (vgl. Abbildung 5.2) deren Attribute und das dazugehörige *Django Field* aufgeführt. Die Entitäten *Nutzer* und *Dateninteressenten* fehlen, da hierfür die *Django*-eigenen Benutzerkonten in jeweils unterschiedlichen Rollen verwendet werden. Auch die Entität für die Zeitreihendaten sind nicht aufgeführt, da diese nicht in der relationalen Datenbank, sondern in einem Zeitreihenspeicher vorgehalten werden.

Entität 1: *Ausgabeformat*

Attribut	Feldtyp
beschreibung	CharField
formatterfile	FileField
name	CharField

Entität 2: *Bewertungsmethode*

Attribut	Feldtyp
beschreibung	CharField
bewertungsfile	FileField
name	CharField

Entität 3: *Datenanfrage*

Attribut	Feldtyp
aufloesung	FloatField
ausgabeformat	ForeignKey(Ausgabeformat)
bewertungsmethode	ForeignKey(Bewertungsmethode)
created	DateTimeField
dateninteressent	ForeignKey(DjangoUser)
end_timestamp	DateTimeField
entscheidung	IntegerField
kategorie	ForeignKey(Kategorie)
nutzer	ForeignKey(DjangoUser)
qualitaetsreduktions- massnahme	ManyToMany- Field(Qualitaetsreduktions- massnahme)
rueckmeldung_nutzer	TextField
start_timestamp	DateTimeField

Entität 4: *Datenbeschreibung*

Attribut	Feldtyp
created	DateTimeField
einheit	ForeignKey(Einheit)
kategorie	ManyToManyField(Kategorie)
notizen	TextField
nutzer	ForeignKey(DjangoUser)
parser	ForeignKey(Parser)
zeitreihendaten	FileField

Entität 5: *Einheit*

Attribut	Feldtyp
beschreibung	CharField
einheit	CharField

Entität 6: *Kategorie*

Attribut	Feldtyp
beschreibung	CharField
kennzahl	IntegerField
name	CharField

Entität 7: *Parser*

Attribut	Feldtyp
beschreibung	CharField
name	CharField
parserfile	FileField

Entität 8: *Qualitätsreduktionsmaßnahme*

Attribut	Feldtyp
beschreibung	CharField
massnahmenfile	FileField
name	CharField
reduktionsfaktor	FloatField

Literaturverzeichnis

- [AA13] A. M. Annaswamy und M. Amin. „IEEE vision for smart grid controls: 2030 and beyond“. In: *IEEE Vision for Smart Grid Controls: 2030 and Beyond* (Juni 2013), S. 1–168. doi: 10.1109/IEEEESTD.2013.6577608.
- [AB07] N. Armaroli und V. Balzani. „The future of energy supply: Challenges and opportunities“. In: *Angewandte Chemie International Edition* 46.1-2 (2007), S. 52–66. issn: 1521-3773. doi: 10.1002/anie.200602373.
- [Abr+05] W. Abrahamse, L. Steg, C. Vlek und T. Rothengatter. „A review of intervention studies aimed at household energy conservation“. In: *Journal of Environmental Psychology* 25.3 (2005), S. 273–291. issn: 0272-4944. doi: 10.1016/j.jenvp.2005.08.002.
- [ÁC11] G. Ács und C. Castelluccia. „I have a DREAM! (DiffeRentially privatE smArt Metering)“. In: *Information Hiding*. Hrsg. von T. Filler, T. Pevný, S. Craver und A. Ker. Bd. 6958. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, S. 118–132. isbn: 978-3-642-24177-2. doi: 10.1007/978-3-642-24178-9_9.
- [Ami12] S. Amin. „Smart grid security, privacy, and resilient architectures: Opportunities and challenges“. In: *Power and Energy Society General Meeting, 2012 IEEE*. Juli 2012, S. 1–2. doi: 10.1109/PESGM.2012.6345767.
- [Att+10] S. Z. Attari, M. L. DeKay, C. I. Davidson und W. Bruine de Bruin. „Public perceptions of energy consumption and savings“. In: *Proceedings of the National Academy of Sciences* 107.37 (2010), S. 16054–16059. doi: 10.1073/pnas.1001509107.

- [BCR12] D. Bonino, F. Corno und L. D. Russis. „Home energy consumption feedback: A user survey“. In: *Energy and Buildings* 47 (2012), S. 383–393. issn: 0378-7788. doi: 10.1016/j.enbuild.2011.12.017.
- [Ber+11] D. Bergman, D. Jin, J. Juen, N. Tanaka, C. Gunter und A. Wright. „Distributed non-intrusive load monitoring“. In: *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*. Jan. 2011, S. 1–8. doi: 10.1109/ISGT.2011.5759180.
- [BK12] H. D. Baehr und S. Kabelac. *Thermodynamik: Grundlagen und technische Anwendungen*. 15. Aufl. Springer-Lehrbuch. Berlin: Springer Vieweg, 2012.
- [BL99] G. Brandon und A. Lewis. „Reducing household energy consumption: A qualitative and quantitative field study“. In: *Journal of Environmental Psychology* 19.1 (1999), S. 75–85. issn: 0272-4944. doi: 10.1006/jevp.1998.0105.
- [BN08] J. Burgess und M. Nye. „Re-materialising energy use through transparent monitoring systems“. In: *Energy Policy* 36.12 (2008). Foresight Sustainable Energy Management and the Built Environment Project, S. 4454–4459. issn: 0301-4215. doi: 10.1016/j.enpol.2008.09.039.
- [Boe+12] M. Boehm, L. Dannecker, A. Doms, E. Dovgan, B. Filipič, U. Fischer, W. Lehner, T. B. Pedersen, Y. Pitarch, L. Šikšnys und T. Tušar. „Data management in the MIRABEL smart grid system“. In: *Proceedings of the 2012 Joint EDBT/ICDT Workshops*. EDBT-ICDT '12. Berlin, Germany: ACM, 2012, S. 95–102. isbn: 978-1-4503-1143-4. doi: 10.1145/2320765.2320797.
- [BP11] A. Barenghi und G. Pelosi. „Security and privacy in smart grid infrastructures“. In: *Database and Expert Systems Applications (DEXA), 2011 22nd International Workshop on*. Aug. 2011, S. 102–108. doi: 10.1109/DEXA.2011.74.
- [BSU10] J.-M. Bohli, C. Sorge und O. Ugus. „A privacy model for smart metering“. In: *Communications Workshops (ICC), 2010 IEEE International Conference on*. Mai 2010, S. 1–5. doi: 10.1109/ICCW.2010.5503916.

- [Bun11] Bundesministerium für Wirtschaft und Energie (BMWi). *Eckpunkte Energieeffizienz*. 2011. URL: <http://www.bmwi.de/BMWi/Redaktion/PDF/E/eckpunkte-energieeffizienz.property=pdf,bereich=bmwi,sprache=de,rwb=true.pdf>.
- [Bun12] Bundesministerium für Wirtschaft und Technologie (BMWi). *Die Energiewende in Deutschland: Mit sicherer, bezahlbarer und umweltschonender Energie ins Jahr 2050*. Sep. 2012.
- [Bun13a] Bundesamt für Sicherheit in der Informationstechnik (BSI). *Technische Richtlinie BSI TR-03109-1, Anlage IV: Feinspezifikation Drahtgebundene LMN-Schnittstelle, Teil b: SML — Smart Message Language*. März 2013. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR03109/TR-03109-1_Anlage_Feinspezifikation_Drahtgebundene_LMN-Schnittstelle_Teilb.pdf?__blob=publicationFile.
- [Bun13b] Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit (BMUB). *Ist Atomstrom wirklich CO2-frei?* Aug. 2013. URL: <http://www.umweltbundesamt.de/service/uba-fragen/ist-atomstrom-wirklich-co2-frei>.
- [Bun14] Bundesamt für Sicherheit in der Informationstechnik (BSI). *Schutzprofil Smart Meter Gateway (BSI-CC-PP-0073)*. Online. März 2014. URL: https://www.bsi.bund.de/DE/Themen/DigitaleGesellschaft/SmartMeter/SmartMeterGateway/Schutzprofil_Gateway/schutzprofil_smartmetergateway_node.html.
- [Bun15a] Bundesamt für Sicherheit in der Informationstechnik (BSI). *BSI TR-03109 Technische Vorgaben für intelligente Messsysteme und deren sicherer Betrieb*. Online. Nov. 2015. URL: <https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/tr03109/index.htm.html>.
- [Bun15b] Bundesamt für Sicherheit in der Informationstechnik (BSI). *Das Smart Meter Gateway – Sicherheit für intelligente Netze*. Online. Okt. 2015. URL: <https://www.bsi.bund.de/DE/Themen/>

- DigitaleGesellschaft / SmartMeter / SmartMeterGateway / Schutzprofil _ Gateway / schutzprofil _ smartmetergateway _ node.html.
- [Bun15c] Bundesministerium für Wirtschaft und Energie (BMWi). *Ein Strommarkt für die Energiewende – Ergebnispapier des Bundesministeriums für Wirtschaft und Energie (Weißbuch)*. BMWi, 11019 Berlin: Bundesministerium für Wirtschaft und Energie – Öffentlichkeitsarbeit, Juli 2015.
- [Bun15d] Bundesministerium für Wirtschaft und Energie (BMWi). *Gesamtausgabe der Energiedaten*. Mai 2015. URL: <http://bmwi.de/DE/Themen/Energie/Energiedaten-und-analysen/Energiedaten/gesamtausgabe>.
- [Bun16] Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit (BMUB). *Flächenverbrauch für Rohstoffabbau*. Okt. 2016. URL: <http://www.umweltbundesamt.de/daten/flaechennutzung/flaechenverbrauch-fuer-rohstoffabbau#textpart-1>.
- [BV04] M. Baranski und J. Voss. „Genetic algorithm for pattern detection in NIALM systems“. In: *Systems, Man and Cybernetics, 2004 IEEE International Conference on*. Bd. 4. Okt. 2004, 3462–3468 vol.4. doi: 10.1109/ICSMC.2004.1400878.
- [Cam+10] R. Camilli, C. M. Reddy, D. R. Yoerger, B. A. S. Van Mooy, M. V. Jakuba, J. C. Kinsey, C. P. McIntyre, S. P. Sylva und J. V. Maloney. „Tracking Hydrocarbon Plume Transport and Biodegradation at Deepwater Horizon“. In: *Science* 330.6001 (2010), S. 201–204. ISSN: 0036-8075. doi: 10.1126/science.1195223. URL: <http://science.sciencemag.org/content/330/6001/201>.
- [CL11] A. Charland und B. Leroux. „Mobile Application Development: Web vs. Native“. In: *Commun. ACM* 54.5 (Mai 2011), S. 49–53. ISSN: 0001-0782. doi: 10.1145/1941487.1941504. URL: <http://doi.acm.org/10.1145/1941487.1941504>.

- [Coh+10] G. Cohn, S. Gupta, J. Froehlich, E. Larson und S. Patel. „GasSense: Appliance-level, single-point sensing of gas activity in the home“. In: *Pervasive Computing*. Hrsg. von P. Floréen, A. Krüger und M. Spasojevic. Bd. 6030. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, S. 265–282. ISBN: 978-3-642-12653-6. doi: 10.1007/978-3-642-12654-3_16.
- [Coo+13] J. Cook, D. Nuccitelli, S. A. Green, M. Richardson, B. Winkler, R. Painting, R. Way, P. Jacobs und A. Skuce. „Quantifying the consensus on anthropogenic global warming in the scientific literature“. In: *Environmental Research Letters* 8.2 (2013), S. 024024. doi: 10.1088/1748-9326/8/2/024024.
- [CPW10] A. Cavoukian, J. Polonetsky und C. Wolf. „SmartPrivacy for the smart grid: embedding privacy into the design of electricity conservation“. English. In: *Identity in the Information Society* 3.2 (2010), S. 275–294. doi: 10.1007/s12394-010-0046-y.
- [Cra12] V. Crastan. *Elektrische Energieversorgung* 2. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. ISBN: 978-3-642-19856-4. doi: 10.1007/978-3-642-19856-4.
- [CTG08] M. Chetty, D. Tran und R. E. Grinter. „Getting to green: Understanding resource consumption in the home“. In: *Proceedings of the 10th International Conference on Ubiquitous Computing*. UbiComp '08. Seoul, Korea: ACM, 2008, S. 242–251. ISBN: 978-1-60558-136-1. doi: 10.1145/1409635.1409668.
- [Dar06] S. Darby. *The effectiveness of feedback on energy consumption*. Techn. Ber. Environmental Change Institute, University of Oxford, 2006.
- [Daw+10] S. Dawson-Haggerty, X. Jiang, G. Tolle, J. Ortiz und D. Culler. „sMAP: A simple measurement and actuation profile for physical information“. In: *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. SenSys '10. New York, NY, USA: ACM, 2010, S. 197–210. ISBN: 978-1-4503-0344-6. doi: 10.1145/1869983.1870003.
- [Dep16] U.S. Department of Energy. *Green Button*. Online. Abgerufen am 30.05.2016. 2016. URL: <http://energy.gov/data/green-button>.

- [Deu16a] Deutscher Bundestag. *Drucksache 18/7555 – Gesetzentwurf der Bundesregierung: Entwurf eines Gesetzes zur Digitalisierung der Energiewende*. Feb. 2016. URL: <http://dip21.bundestag.de/dip21/btd/18/075/1807555.pdf>.
- [Deu16b] Deutscher Bundestag. *Drucksache 18/8919 – Beschlussempfehlung und Bericht des Ausschusses für Wirtschaft und Energie (9. Ausschuss) zu dem Gesetzentwurf der Bundesregierung Drucksache 18/7555*. Juni 2016. URL: <http://dip.bundestag.de/btd/18/089/1808919.pdf>.
- [Dis16] Discovergy GmbH. *Discovergy*. Online. Abgerufen am 31.05.2016. 2016. URL: <https://www.discovery.com/>.
- [Dja16a] Django Software Foundation. *Django*. Software. 2016. URL: <https://www.djangoproject.com>.
- [Dja16b] Django Software Foundation. *Django documentation*. Online. Abgerufen am 30. Oktober 2016. 2016. URL: <https://docs.djangoproject.com/en/>.
- [DK13] T. Dimitriou und G. Karame. „Privacy-friendly tasking and trading of energy in smart grids“. In: *Proceedings of the 28th Annual ACM Symposium on Applied Computing*. SAC '13. Coimbra, Portugal: ACM, 2013, S. 652–659. ISBN: 978-1-4503-1656-9. doi: [10.1145/2480362.2480488](https://doi.org/10.1145/2480362.2480488).
- [Eas16] Easymeter GmbH. *Datenblatt Elektronischer Drehstromzähler Q3B*. Online. Abgerufen am 14.06.2016. 2016. URL: <http://www.easymeter.com/produkt/Q3B>.
- [EE13] D. Engel und G. Eibl. „Multi-resolution load curve representation with privacy-preserving aggregation“. In: *Proceedings of IEEE Innovative Smart Grid Technologies (ISGT) 2013*. Copenhagen, Denmark: IEEE, Okt. 2013, S. 1–5. doi: [10.1109/ISGETurope.2013.6695231](https://doi.org/10.1109/ISGETurope.2013.6695231).
- [EE14] G. Eibl und D. Engel. „Influence of data granularity on nonintrusive appliance load monitoring“. In: *Proceedings of the Second ACM Workshop on Information Hiding and Multimedia Security (IH&MM-*

- Sec '14). Salzburg, Austria: ACM, 2014, S. 147–151. ISBN: 978-1-4503-2647-6. doi: 10.1145/2600918.2600920.*
- [EE15] G. Eibl und D. Engel. „Influence of data granularity on smart meter privacy“. In: *IEEE Transactions on Smart Grid* 6.2 (März 2015), S. 930–939. doi: 10.1109/TSG.2014.2376613.
- [EEN15] G. Eibl, D. Engel und C. Neureiter. „Privacy-relevant smart metering use cases“. In: *Proceedings of IEEE International Conference on Industrial Technology (ICIT) 2015*. Seville, Spain: IEEE, 2015, S. 1387–1392.
- [EK10] C. Efthymiou und G. Kalogridis. „Smart grid privacy via anonymization of smart metering data“. In: *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. Okt. 2010, S. 238–243. doi: 10.1109/SMARTGRID.2010.5622050.
- [Eng13] D. Engel. „Wavelet-based load profile representation for smart meter privacy“. In: *Proceedings IEEE PES Innovative Smart Grid Technologies (ISGT'13)*. Washington, D.C., USA: IEEE, Feb. 2013, S. 1–6. doi: 10.1109/ISGT.2013.6497835.
- [Erk+13] Z. Erkin, J. Troncoso-Pastoriza, R. Lagendijk und F. Perez-Gonzalez. „Privacy-preserving data aggregation in smart metering systems: an overview“. In: *Signal Processing Magazine, IEEE* 30.2 (März 2013), S. 75–86. ISSN: 1053-5888. doi: 10.1109/MSP.2012.2228343.
- [Ern13] Ernst & Young Power & Utilities. *Kosten-Nutzen-Analyse für einen flächendeckenden Einsatz intelligenter Zähler*. Online. Abgerufen am 21.04.2016. 2013. URL: <http://www.bmwi.de/BMWi/Redaktion/PDF/Publicationen/Studien/kosten-nutzen-analyse-fuer-flaechendeckenden-einsatz-intelligenterzaehler.property=pdf,bereich=bmwi2012,sprache=de,rwb=true.pdf>.
- [Fan+10] Z. Fan, G. Kalogridis, C. Efthymiou, M. Sooriyabandara, M. Serizawa und J. McGeehan. „The new frontier of communications research: Smart grid and smart metering“. In: *Proceedings of the 1st International Conference on Energy-Efficient Computing and Net-*

- working.* e-Energy '10. Passau, Germany: ACM, 2010, S. 115–118. ISBN: 978-1-4503-0042-1. doi: 10.1145/1791314.1791331.
- [Fan+12] X. Fang, S. Misra, G. Xue und D. Yang. „Smart grid 2014; The new and improved power grid: A Survey“. In: *Communications Surveys Tutorials, IEEE* 14.4 (Apr. 2012), S. 944–980. ISSN: 1553-877X. doi: 10.1109/SURV.2011.101911.00087.
- [Fan+13] Z. Fan, P. Kulkarni, S. Gormus, C. Efthymiou, G. Kalogridis, M. Sooriyabandara, Z. Zhu, S. Lambotharan und W. H. Chin. „Smart grid communications: Overview of research challenges, solutions, and standardization activities“. In: *Communications Surveys Tutorials, IEEE* 15.1 (2013), S. 21–38. ISSN: 1553-877X. doi: 10.1109/SURV.2011.122211.00021.
- [FB14a] S. Finster und I. Baumgart. „Privacy-Aware Smart Metering: A Survey“. In: *IEEE Communications Surveys Tutorials* 16.3 (März 2014), S. 1732–1745. ISSN: 1553-877X. doi: 10.1109/SURV.2014.052914.00090.
- [FB14b] S. Finster und I. Baumgart. „SMART-ER: Peer-based privacy for smart metering“. In: *Computer Communications Workshops (INFO-COM WKSHPS), 2014 IEEE Conference on*. Apr. 2014, S. 652–657. doi: 10.1109/INFCOMW.2014.6849308.
- [Fol+11] M. Folk, G. Heber, Q. Koziol, E. Pourmal und D. Robinson. „An Overview of the HDF5 Technology Suite and Its Applications“. In: *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases*. AD '11. Uppsala, Sweden: ACM, 2011, S. 36–47. ISBN: 978-1-4503-0614-0. doi: 10.1145/1966895.1966900.
- [Fro+09] J. E. Froehlich, E. Larson, T. Campbell, C. Haggerty, J. Fogarty und S. N. Patel. „HydroSense: Infrastructure-mediated single-point sensing of whole-home water activity“. In: *Proceedings of the 11th International Conference on Ubiquitous Computing*. UbiComp '09. Orlando, Florida, USA: ACM, 2009, S. 235–244. ISBN: 978-1-60558-431-7. doi: 10.1145/1620545.1620581.

- [FSS10] A. Faruqui, S. Sergici und A. Sharif. „The impact of informational feedback on energy consumption—A survey of the experimental evidence“. In: *Energy* 35.4 (2010). Demand Response Resources: the {US} and International ExperienceDemand Response Resources: the {US} and International Experience, S. 1598–1608. issn: 0360-5442. doi: 10.1016/j.energy.2009.07.042.
- [Git+16] C. Gitte, H. Xu, F. Rigoll, J. van Eekelen und M. Kaisers. „Multi-Commodity Energy Management Applied to Micro CHPs and Electrical Heaters in Smart Buildings“. In: *5th D-A-CH+ Energy Informatics Conference in conjunction with 7th Symposium on Communications for Energy Systems (ComForEn)*. Band 84 in OVE-Schriftenreihe. Klagenfurt: Österreichischer Verband für Elektrotechnik, 2016. isbn: 978-3-85133-090-8.
- [GJ11] F. Garcia und B. Jacobs. „Privacy-friendly energy-metering via homomorphic encryption“. English. In: *Security and Trust Management*. Hrsg. von J. Cuellar, J. Lopez, G. Barthe und A. Pretschner. Bd. 6710. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, S. 226–238. isbn: 978-3-642-22443-0. doi: 10.1007/978-3-642-22444-7_15.
- [GJL12] U. Greveler, B. Justus und D. Loehr. „Multimedia content identification through smart meter power usage profiles“. In: *Computers, Privacy and Data Protection* 1 (2012), S. 10.
- [Goe+14] C. Goebel, H.-A. Jacobsen, V. del Razo, C. Doblander, J. Rivera, J. Ilg, C. Flath, H. Schmeck, C. Weinhardt, D. Pathmaperuma, H.-J. Appelrath, M. Sonnenschein, S. Lehrhoff, O. Kramer, T. Staake, E. Fleisch, D. Neumann, J. Strüker, K. Erek, R. Zarnekow, H. Ziekow und J. Lässig. „Energy Informatics“. In: *Business & Information Systems Engineering* 6.1 (2014), S. 25–31. doi: 10.1007/s12599-013-0304-2.
- [Gre16] Green Button Initiative. *Green Button*. Online. Abgerufen am 24.05.2016. 2016. url: <http://www.greenbuttondata.org/>.

- [GRP10] S. Gupta, M. S. Reynolds und S. N. Patel. „ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home“. In: *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*. UbiComp '10. Copenhagen, Denmark: ACM, 2010, S. 139–148. ISBN: 978-1-60558-843-8. doi: 10.1145/1864349.1864375.
- [GS12] M. Giampietro und A. H. Sorman. „Are Energy Statistics Useful for Making Energy Scenarios?“ In: *Energy* 37.1 (2012). 7th Biennial International Workshop “Advances in Energy Studies”, S. 5–17. ISSN: 0360-5442. doi: 10.1016/j.energy.2011.08.038. URL: <http://www.sciencedirect.com/science/article/pii/S0360544211005780>.
- [Gun+11] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati und G. P. Hancke. „Smart Grid Technologies: Communication Technologies and Standards“. In: *IEEE Transactions on Industrial Informatics* 7.4 (Nov. 2011), S. 529–539. ISSN: 1551-3203. doi: 10.1109/TII.2011.2166794.
- [Han+06] J. Hansen, M. Sato, R. Ruedy, K. Lo, D. W. Lea und M. Medina-Elizade. „Global temperature change“. In: *Proceedings of the National Academy of Sciences* 103.39 (2006), S. 14288–14293. doi: 10.1073/pnas.0606291103.
- [Har89] G. W. Hart. „Residential energy monitoring and computerized surveillance via utility power flows“. In: *IEEE Technology and Society Magazine* 8.2 (Juni 1989), S. 12–16. ISSN: 0278-0097. doi: 10.1109/44.31557.
- [Har92] G. W. Hart. „Nonintrusive appliance load monitoring“. In: *Proceedings of the IEEE* 80.12 (1992), S. 1870–1891.
- [Hei16] Heidelberger Services AG. *EVU-Kernprozesse*. Online. Abgerufen am 20.04.2016. Apr. 2016. URL: <https://www.hsag.info/leistungen/service/evu-kernprozesse/>.
- [Hil14] L. Hillemacher. „Lastmanagement mittels dynamischer Strompreissignale bei Haushaltskunden“. Karlsruhe, KIT, Diss., 2014. Diss. Karlsruher Institut für Technologie, 2014.

- [Hir15] C. Hirsch. „Fahrplanbasiertes Energiemanagement in Smart Grids“. Phdthesis. Karlsruhe: KIT, Fakultät für Wirtschaftswissenschaften, 2015.
- [HNB10] T. Hargreaves, M. Nye und J. Burgess. „Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors“. In: *Energy Policy* 38.10 (2010). The socio-economic transition towards a hydrogen economy - findings from European research, with regular papers, S. 6111–6119. issn: 0301-4215. doi: 10.1016/j.enpol.2010.05.068.
- [HNB13] T. Hargreaves, M. Nye und J. Burgess. „Keeping energy visible? Exploring how householders interact with feedback from smart energy monitors in the longer term“. In: *Energy Policy* 52 (2013). Special Section: Transition Pathways to a Low Carbon Economy, S. 126–134. issn: 0301-4215. doi: 10.1016/j.enpol.2012.03.027.
- [HP14] H.-M. Henning und A. Palzer. „A comprehensive model for the German electricity and heat sector in a future energy system with a dominant contribution from renewable energy technologies—Part I: Methodology“. In: *Renewable and Sustainable Energy Reviews* 30 (2014), S. 1003–1018. issn: 1364-0321. doi: 10.1016/j.rser.2013.09.012.
- [IA09] A. Ipakchi und F. Albuyeh. „Grid of the future“. In: *Power and Energy Magazine, IEEE* 7.2 (Apr. 2009), S. 52–62. issn: 1540-7977. doi: 10.1109/MPE.2008.931384.
- [Jia+09] X. Jiang, S. Dawson-Haggerty, P. Dutta und D. Culler. „Design and implementation of a high-fidelity AC metering network“. In: *Information Processing in Sensor Networks, 2009. IPSN 2009. International Conference on*. Apr. 2009, S. 253–264.
- [JJR11] M. Jawurek, M. Johns und K. Rieck. „Smart metering de-pseudonymization“. In: *Proceedings of the 27th Annual Computer Security Applications Conference*. ACM. 2011, S. 227–236.

- [JKD12] M. Jawurek, F. Kerschbaum und G. Danezis. *Privacy technologies for smart grids - A survey of options*. Techn. Ber. MSR-TR-2012-119. Microsoft Research, Cambridge, UK, Nov. 2012. url: <http://research.microsoft.com/apps/pubs/default.aspx?id=178055>.
- [Jun+14] C. Jung, M. Gasthuber, A. Giesler, M. Hardt, J. Meyer, F. Rigoll, K. Schwarz, R. Stotzka und A. Streit. „Optimization of data life cycles“. In: *Journal of Physics: Conference Series* 513.3 (2014). doi: 10.1088/1742-6596/513/3/032047.
- [Jun+15] C. Jung, M. Gasthuber, A. Giesler, M. Hardt, J. Meyer, A. Prabhune, F. Rigoll, K. Schwarz und A. Streit. „Progress in multi-disciplinary data life cycle management“. In: *Journal of Physics: Conference Series* 664.3 (2015), S. 032018. doi: 10.1088/1742-6596/664/3/032018.
- [Jun+16] C. Jung, S. Fleischer, M. Gasthuber, A. Giesler, M. Hardt, J. Meyer, F. Rigoll, R. Stotzka, M. Fischer und A. Streit. „Advancing data management and analysis in different scientific disciplines“. In: *Journal of Physics: Conference Series* (2016). Akzeptiert.
- [Kal+10] G. Kalogridis, C. Efthymiou, S. Denic, T. Lewis und R. Cepeda. „Privacy for smart meters: Towards undetectable appliance load signatures“. In: *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. Okt. 2010, S. 232–237. doi: 10.1109/SMARTGRID.2010.5622047.
- [Kar11] S. Karjalainen. „Consumer preferences for feedback on household electricity consumption“. In: *Energy and Buildings* 43.2–3 (2011), S. 458–467. issn: 0378-7788. doi: 10.1016/j.enbuild.2010.10.010.
- [KD11] G. Kalogridis und S. Denic. „Data mining and privacy of personal behaviour types in smart grid“. In: *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*. Dez. 2011, S. 636–642. doi: 10.1109/ICDMW.2011.58.
- [KD12] G. Kalogridis und S. Dave. „PeHEMS: privacy enabled HEMS and load balancing prototype“. In: *Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on*. Nov. 2012, S. 486–491. doi: 10.1109/SmartGridComm.2012.6486032.

- [KDK11] K. Kursawe, G. Danezis und M. Kohlweiss. „Privacy-friendly aggregation for the smart-grid“. In: *Privacy Enhancing Technologies*. Hrsg. von S. Fischer-Hübner und N. Hopper. Bd. 6794. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, S. 175–191. ISBN: 978-3-642-22262-7. doi: 10.1007/978-3-642-22263-4_10.
- [KFB11] G. Kalogridis, Z. Fan und S. Basutkar. „Affordable privacy for home smart meters“. In: *Parallel and Distributed Processing with Applications Workshops (ISPAW), 2011 Ninth IEEE International Symposium on*. Mai 2011, S. 77–84. doi: 10.1109/ISPAW.2011.42.
- [Khu+10] H. Khurana, M. Hadley, N. Lu und D. Frincke. „Smart-grid security issues“. In: *Security Privacy, IEEE* 8.1 (Jan. 2010), S. 81–85. ISSN: 1540-7993. doi: 10.1109/MSP.2010.49.
- [Kim+10] Y.-J. Kim, M. Thottan, V. Kolesnikov und W. Lee. „A secure decentralized data-centric information infrastructure for smart grid“. In: *Communications Magazine, IEEE* 48.11 (Nov. 2010), S. 58–65. ISSN: 0163-6804. doi: 10.1109/MCOM.2010.5621968.
- [Kim+11] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon und J. Han. „Unsupervised disaggregation of low frequency power measurements“. In: *SDM*. Bd. 11. SIAM. 2011, S. 747–758.
- [KJ11] J. Z. Kolter und M. J. Johnson. „REDD: A public data set for energy disaggregation research“. In: *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*. Bd. 25. Citeseer. 2011, S. 59–62.
- [KJ12] J. Z. Kolter und T. Jaakkola. „Approximate inference in additive factorial HMMS with application to energy disaggregation“. In: *International conference on artificial intelligence and statistics*. 2012, S. 1472–1482.
- [KNS11] Y. Kim, E. Ngai und M. Srivastava. „Cooperative state estimation for preserving privacy of user behaviors in smart grid“. In: *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*. Okt. 2011, S. 178–183. doi: 10.1109/SmartGridComm.2011.6102313.

- [KS14] M. Karwe und J. Strüker. „A survey on privacy in residential demand side management applications“. English. In: *Smart Grid Security*. Hrsg. von J. Cuellar. Bd. 8448. Lecture Notes in Computer Science. Springer International Publishing, 2014, S. 154–165. ISBN: 978-3-319-10328-0. doi: [10.1007/978-3-319-10329-7_10](https://doi.org/10.1007/978-3-319-10329-7_10).
- [Lau+03] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford und P. Armstrong. „Power signature analysis“. In: *IEEE Power and Energy Magazine* 1.2 (März 2003), S. 56–63. ISSN: 1540-7977. doi: [10.1109/MPAE.2003.1192027](https://doi.org/10.1109/MPAE.2003.1192027).
- [Lin+11] M. Lin, A. Wierman, L. Andrew und E. Thereska. „Dynamic right-sizing for power-proportional data centers“. In: *INFOCOM, 2011 Proceedings IEEE*. Apr. 2011, S. 1098–1106. doi: [10.1109/INFCOM.2011.5934885](https://doi.org/10.1109/INFCOM.2011.5934885).
- [LMW10] M. Lisovich, D. Mulligan und S. Wicker. „Inferring personal information from demand-response systems“. In: *Security Privacy, IEEE* 8.1 (Jan. 2010), S. 11–20. ISSN: 1540-7993. doi: [10.1109/MSP.2010.40](https://doi.org/10.1109/MSP.2010.40).
- [LSS14] J. Liao, L. Stankovic und V. Stankovic. „Detecting household activity patterns from smart meter data“. In: *Intelligent Environments (IE), 2014 International Conference on*. Juni 2014, S. 71–78. doi: [10.1109/IE.2014.18](https://doi.org/10.1109/IE.2014.18).
- [LW08] M. Lisovich und S. Wicker. „Privacy concerns in upcoming residential and commercial demand-response systems“. In: *Proc. of the Clemson University Power Systems Conference (Clemson, SC)*. Citeseer. 2008.
- [Maa+12] H. Maass, H. K. Çakmak, W. Suess, A. Quinte, W. Jakob, K. U. Stucky und U. Kuehnafel. „Introducing the Electrical Data Recorder as a new capturing device for power grid analysis“. In: *Applied Measurements for Power Systems (AMPS), 2012 IEEE International Workshop on*. Sep. 2012, S. 1–6. doi: [10.1109/AMPS.2012.6343992](https://doi.org/10.1109/AMPS.2012.6343992).

- [Mar+11] A. Marchiori, D. Hakkarinen, Q. Han und L. Earle. „Circuit-level load monitoring for household energy management“. In: *Pervasive Computing, IEEE* 10.1 (Jan. 2011), S. 40–48. ISSN: 1536-1268. doi: 10.1109/MPRV.2010.72.
- [McK+10] W. McKinney u. a. „Data structures for statistical computing in python“. In: *Proceedings of the 9th Python in Science Conference*. Bd. 445. 2010, S. 51–56.
- [McK12] W. McKinney. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, Inc., 2012.
- [MM09] P. McDaniel und S. McLaughlin. „Security and privacy challenges in the smart grid“. In: *IEEE Security & Privacy* 7.3 (2009), S. 75–77. ISSN: 1540-7993. doi: 10.1109/MSP.2009.76.
- [Mol+10] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet und D. Irwin. „Private memoirs of a smart meter“. In: *Proceedings of the 2Nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*. BuildSys '10. Zurich, Switzerland: ACM, 2010, S. 61–66. ISBN: 978-1-4503-0458-0. doi: 10.1145/1878431.1878446.
- [MR03] S. Meyer und A. Rakotonirainy. „A survey of research on context-aware homes“. In: *Proceedings of the Australasian information security workshop conference on ACSW frontiers 2003-Volume 21*. Australian Computer Society, Inc. 2003, S. 159–168.
- [MSW10] F. Mattern, T. Staake und M. Weiss. „ICT for green: How computers can help us to conserve energy“. In: *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking*. e-Energy '10. Passau, Germany: ACM, 2010, S. 1–10. ISBN: 978-1-4503-0042-1. doi: 10.1145/1791314.1791316.
- [Mül10] K. Müller. „Gewinnung von Verhaltensprofilen am intelligenten Stromzähler“. German. In: *Datenschutz und Datensicherheit - DuD* 34.6 (2010), S. 359–364. ISSN: 1614-0702. doi: 10.1007/s11623-010-0107-2.

- [Nat14] National Institute of Standards and Technology. *NIST framework and roadmap for smart grid interoperability standards, release 3.0*. Techn. Ber. National Institute of Standards and Technology, 2014. URL: <http://www.nist.gov/smartgrid/framework3.cfm>.
- [Nat15] National Institute of Standards and Technology (NIST). *Green Button Initiative*. Online. Jan. 2015. URL: <https://www2.nist.gov/property-fieldsection/green-button-initiative>.
- [Nat16] National Aeronautics and Space Administration (NASA). *HDF5 Data Model, File Format and Library – HDF5 1.6*. Online. Abgerufen am 13.09.2016. 2016. URL: <https://earthdata.nasa.gov/standards/hdf5>.
- [NL96] L. Norford und S. Leeb. „Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms“. In: *Energy and Buildings* 24.1 (1996), S. 51–64.
- [NR09] B. Neenan und J. Robinson. *Residential electricity use feedback: A research synthesis and economic framework*. Techn. Ber. Electric Power Research Institute (EPRI), 2009.
- [Par+12] O. Parson, S. Ghosh, M. Weal und A. Rogers. „Non-intrusive load monitoring using prior models of general appliance types“. In: *Twenty-Sixth Conference on Artificial Intelligence (AAAI-12)*. 2012.
- [PE14] C. Peuker und D. Engel. „Praxistaugliche Kommunikationssicherheit in einer Smart Metering Infrastruktur“. In: *Proc. 8. Forschungsforum der Österreichischen Fachhochschulen*. Hrsg. von J. Lüthi und H.-P. Steinebacher. 2014, S. 72–76. ISBN: 978-3-9503491-9-1.
- [PHB06] A. Pretschner, M. Hilty und D. Basin. „Distributed usage control“. In: *Commun. ACM* 49.9 (Sep. 2006), S. 39–44. ISSN: 0001-0782. doi: [10.1145/1151030.1151053](https://doi.org/10.1145/1151030.1151053).
- [PJ 10] P.J. Eby. *PEP 3333 – Python Web Server Gateway Interface v1.0.1*. Online. Sep. 2010. URL: <https://www.python.org/dev/peps/pep-3333/>.

- [Plu12] B. Plumer. *No, wind farms are not causing global warming*. The Washington Post (online). Apr. 2012. URL: https://www.washingtonpost.com/blogs/ezra-klein/post/no-wind-farms-are-not-causing-global-warming/2012/04/30/gIQAM12GsT_blog.html.
- [Plu14] Plugwise B.V. *Licence terms governing Plugwise Source software for the consumer and small business users (201105)*. Online. Bestandteil des Programms Plugwise Source. 2014.
- [Plu16] Plugwise B.V. *Plugwise*. Online. Abgerufen am 30.05.2016. 2016. URL: <https://www.plugwise.de/>.
- [Pre+08] A. Pretschner, M. Hilty, D. Basin, C. Schaefer und T. Walter. „Mechanisms for Usage Control“. In: *Proceedings of the 2008 ACM Symposium on Information, Computer and Communications Security*. ASIACCS '08. Tokyo, Japan: ACM, 2008, S. 240–244. ISBN: 978-1-59593-979-1. doi: 10.1145/1368310.1368344.
- [Pru02] A. Prudenzi. „A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel“. In: *Power Engineering Society Winter Meeting*, 2002. IEEE. Bd. 2. 2002, 941–946 vol.2. doi: 10.1109/PESW.2002.985144.
- [Pyt16] Python Software Foundation. *Python*. Software. 2016. URL: <https://www.python.org>.
- [Qua15] V. Quaschning. *Regenerative Energiesysteme: Technologie – Berechnung – Simulation*. 9., aktualisierte und erw. Aufl. Hanser eLibrary. München: Hanser, 2015. ISBN: 978-3-446-44333-4. doi: 10.3139/9783446443334.
- [Qui09] E. L. Quinn. „Privacy and the new energy infrastructure“. In: *Social Science Research Network (SSRN)* (Feb. 2009).
- [Raj+11] S. Rajagopalan, L. Sankar, S. Mohajer und H. Poor. „Smart meter privacy: A utility-privacy framework“. In: *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*. Okt. 2011, S. 190–195. doi: 10.1109/SmartGridComm.2011.6102315.

- [RD11] A. Rial und G. Danezis. „Privacy-preserving smart metering“. In: *Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society*. WPES '11. Chicago, Illinois, USA: ACM, 2011, S. 49–60. ISBN: 978-1-4503-1002-4. doi: 10.1145/2046556.2046564.
- [REC13] A. Reinhardt, F. Englert und D. Christin. „Enhancing user privacy by preprocessing distributed smart meter data“. In: *Sustainable Internet and ICT for Sustainability (SustainIT)*, 2013. IEEE. 2013, S. 1–7.
- [RGS14] F. Rigoll, C. Gitte und H. Schmeck. „Data Life Cycles in Future Residential Multi-Commodity Energy Management Systems“. In: *Great Lakes Symposium on Smart Grid and the New Energy Economy*. Hrsg. von IEEE. IEEE. Chicago: IEEE, Sep. 2014.
- [Rig+14] F. Rigoll, C. Hirsch, S. Kochanneck, H. Schmeck und I. Mauser. „A Privacy-Aware Architecture for Energy Management Systems in Smart Grids“. In: *Ubiquitous Intelligence and Computing, 2014 IEEE 11th Intl Conf on and IEEE 11th Intl Conf on and Autonomic and Trusted Computing, and IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UTC-ATC-ScalCom)*. Dez. 2014, S. 449–455. doi: 10.1109/UIC-ATC-ScalCom.2014.9.
- [RS14] F. Rigoll und H. Schmeck. „Konzeption eines Energiedatenmanagementsystems unter Beachtung von Datenschutz und Privatsphäre“. In: *VDE-Kongress 2014*. Hrsg. von VDE. VDE VERLAG GmbH, Okt. 2014.
- [Ruz+10] A. G. Ruzzelli, C. Nicolas, A. Schoofs und G. M. O'Hare. „Real-time recognition and profiling of appliances through a single electricity sensor“. In: *Sensor Mesh and Ad Hoc Communications and Networks (SECON), 2010 7th Annual IEEE Communications Society Conference on*. IEEE. 2010, S. 1–9.
- [Sah+14] A. Saha, S. Rahman, M. Pipattanasomporn und M. Kuzlu. „On security of a home energy management system“. In: *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, 2014 IEEE PES. Okt. 2014, S. 1–5. doi: 10.1109/ISGTEurope.2014.7028872.

- [San+11] L. Sankar, S. Kar, R. Tandon und H. Poor. „Competitive privacy in the smart grid: An information-theoretic approach“. In: *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*. Okt. 2011, S. 220–225. doi: 10.1109/SmartGridComm.2011.6102322.
- [Sch13] W. Scholze-Stubenrecht, Hrsg. *Der Duden in zwölf Bänden*. 26., völlig neu bearb. und erw. Aufl. Bd. 1: Duden - Die deutsche Rechtschreibung : [das umfassende Standardwerk] auf der Grundlage der aktuellen amtlichen Rechtschreibregeln. Berlin: Dudenverl., 2013. ISBN: 978-3-411-04016-2.
- [Sco15] N. Scovronick. *Reducing global health risks through mitigation of short-lived climate pollutants*. Techn. Ber. World Health Organization, 2015.
- [Shi+11] E. Shi, T.-H. H. Chan, E. G. Rieffel, R. Chow und D. Song. „Privacy-preserving aggregation of time-series data“. In: *NDSS*. Bd. 2. 3. 2011, S. 4.
- [Sid+12] F. Siddiqui, S. Zeadally, C. Alcaraz und S. Galvao. „Smart grid privacy: Issues and solutions“. In: *Computer Communications and Networks (ICCCN), 2012 21st International Conference on*. Juli 2012, S. 1–5. doi: 10.1109/ICCCN.2012.6289304.
- [SK12] J. Strüker und F. Kerschbaum. „From a barrier to a bridge: Data-privacy in deregulated smart grids“. In: *Proceedings of 32nd International Conference on Information Systems (ICIS)*. 2012.
- [Sma12a] Smart Grid Coordination Group (SG-CG). *Framework document*. Techn. Ber. CEN/CENELEC/ETSI, Nov. 2012.
- [Sma12b] Smart Grid Coordination Group (SG-CG). *Smart grid reference architecture*. Techn. Ber. CEN/CENELEC/ETSI, Nov. 2012.
- [SNL06] D. Srinivasan, W. S. Ng und A. C. Liew. „Neural-network-based signature recognition for harmonic source identification“. In: *IEEE Transactions on Power Delivery* 21.1 (Jan. 2006), S. 398–405. ISSN: 0885-8977. doi: 10.1109/TPWRD.2005.852370.

- [SP13] D. S. Sayogo und T. A. Pardo. „Understanding smart data disclosure policy success: The case of Green Button“. In: *Proceedings of the 14th Annual International Conference on Digital Government Research*. dg.o '13. Quebec, Canada: ACM, 2013, S. 72–81. ISBN: 978-1-4503-2057-3. doi: 10.1145/2479724.2479737.
- [Sta10] R. Stallman. „Is digital inclusion a good thing? How can we make sure it is?“ In: *Communications Magazine, IEEE* 48.2 (Feb. 2010), S. 112–118. ISSN: 0163-6804. doi: 10.1109/MCOM.2010.5402673.
- [Sta16] Statista – Das Statistik-Portal. *Zuwachs der Weltbevölkerung, angegeben in unterschiedlichen Zeiteinheiten (Stand 2016)*. Online. 2016. URL: <http://de.statista.com/statistik/daten/studie/1816/umfrage/zuwachs-der-weltbevoelkerung/>.
- [Ste08] L. Steg. „Promoting household energy conservation“. In: *Energy Policy* 36.12 (2008). Foresight Sustainable Energy Management and the Built Environment Project, S. 4449–4453. ISSN: 0301-4215. doi: 10.1016/j.enpol.2008.09.027.
- [STP12] L. Siksnys, C. Thomsen und T. Pedersen. „MIRABEL DW: Managing complex energy data in a smart grid“. English. In: *Data Warehousing and Knowledge Discovery*. Hrsg. von A. Cuzzocrea und U. Dayal. Bd. 7448. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, S. 443–457. ISBN: 978-3-642-32583-0. doi: 10.1007/978-3-642-32584-7_36.
- [Str14] Strom-Report.de. *Akzeptanz von Kraftwerken in der Nachbarschaft*. Online. 2014. URL: <http://strom-report.de/okostrom/#akzeptanz-kraftwerke-erneuerbare>.
- [Süd16] SüdWestStrom. *Energiedatenmanagement für Netzbetreiber*. Online. Abgerufen am 20.04.2016. Apr. 2016. URL: <http://www.suedweststrom.de/dienstleistungen/edm-fuer-netzbetreiber.html>.
- [The14] The Smart Grid Interoperability Panel – Smart Grid Cybersecurity Committee. *Guidelines for smart grid cybersecurity: Volume 2 – Privacy and the smart grid*. Techn. Ber. National Institute of Standards and Technology (NIST), Sep. 2014. doi: 10.6028/NIST.IR.7628r1.

- [TIL04] E. Tapia, S. Intille und K. Larson. „Activity Recognition in the Home Using Simple and Ubiquitous Sensors“. In: *Pervasive Computing*. Hrsg. von A. Ferscha und F. Mattern. Bd. 3001. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, S. 158–175. ISBN: 978-3-540-21835-7. doi: 10.1007/978-3-540-24646-6_10.
- [Umw07] Umweltbundesamt. *Ökonomische Bewertung von Umweltschäden: Methodenkonvention zur Schätzung von Umweltkosten*. Apr. 2007.
- [Umw12] Umweltbundesamt. *Ökonomische Bewertung von Umweltschäden: Methodenkonvention 2.0 zur Schätzung von Umweltkosten*. Aug. 2012.
- [Vas+12] I. Vassileva, M. Odlare, F. Wallin und E. Dahlquist. „The impact of consumers' feedback preferences on domestic electricity consumption“. In: *Applied Energy* 93 (2012). (1) Green Energy; (2) Special Section from papers presented at the 2nd International Energy 2030 Conf, S. 575–582. ISSN: 0306-2619. doi: 10.1016/j.apenergy.2011.12.067.
- [VDI07] VDI-Gesellschaft Energietechnik. *Energiemanagement: Begriffe*. VDI-Richtlinie 4602 Blatt 1. 2007.
- [VDI13] VDI-Gesellschaft Energie und Umwelt (GEU). *Energiemanagement: Beispiele*. VDI-Richtlinie 4602 Blatt 2. 2013.
- [VDI14] VDI-Gesellschaft Energie und Umwelt (GEU). *Energiekenngrößen: Grundlagen – Methodik*. VDI-Richtlinie 4661. 2014.
- [Vei+13] A. Veichtlbauer, D. Engel, F. Knirsch, O. Langthaler und F. Moser. „Advanced metering and data access infrastructures in smart grid environments“. In: *Proc. 7th International Conference on Sensor Technologies and Applications*. Barcelona, Spain, Aug. 2013, S. 63–68.
- [VK11] D. Varodayan und A. Khisti. „Smart meter privacy using a rechargeable battery: Minimizing the rate of information leakage“. In: *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. Mai 2011, S. 1932–1935. doi: 10.1109/ICASSP.2011.5946886.

- [VKD13] E. Vogiatzis, G. Kalogridis und S. Denic. „Real-time and low cost energy disaggregation of coarse meter data“. In: *Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*. Okt. 2013, S. 1–5. doi: 10.1109/ISGTEurope.2013.6695417.
- [VPS12] A. Veichtlbauer, T. Pfeiffenberger und U. Schrittesser. „Generic control architecture for heterogeneous building automation applications“. In: *Proceedings of the 6th International Conference on Sensor Technologies and Applications (SensorComm 2012), Rome*. 2012, S. 148–153.
- [VWD12] I. Vassileva, F. Wallin und E. Dahlquist. „Understanding energy consumption behavior for future demand response strategy development“. In: *Energy* 46.1 (2012). Energy and Exergy Modelling of Advance Energy Systems, S. 94–100. issn: 0360-5442. doi: 10.1016/j.energy.2012.02.069.
- [Wei+12a] M. Weiss, A. Helfenstein, F. Mattern und T. Staake. „Leveraging smart meter data to recognize home appliances“. In: *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*. März 2012, S. 190–197. doi: 10.1109/PerCom.2012.6199866.
- [Wei+12b] M. Weiss, C.-M. Loock, T. Staake, F. Mattern und E. Fleisch. „Evaluating mobile phones as energy consumption feedback devices“. In: *Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Hrsg. von P. Sénac, M. Ott und A. Seneviratne. Bd. 73. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer Berlin Heidelberg, 2012, S. 63–77. isbn: 978-3-642-29153-1. doi: 10.1007/978-3-642-29154-8_6.
- [Wei+13] D. Weißbach, G. Ruprecht, A. Huke, K. Czerski, S. Gottlieb und A. Hussein. „Energy intensities, {EROIs} (energy returned on invested), and energy payback times of electricity generating power plants“. In: *Energy* 52 (2013), S. 210–221. issn: 0360-5442. doi: 10.1016/j.energy.2013.01.029.

- [Wes13] V. Wesselak. *Regenerative Energietechnik*. Hrsg. von T. Schabbach, T. Link und J. Fischer. Berlin, Heidelberg, 2013. doi: 10.1007/978-3-642-24165-9.
- [Wez+12] J. van Wezel, A. Streit, C. Jung, R. Stotzka, S. Halstenberg, F. Rigoll, A. García, A. Heiss, K. Schwarz, M. Gasthuber und A. Giesler. „Data Life Cycle Labs, A New Concept to Support Data-Intensive Science“. In: *CoRR* abs/1212.5596 (2012). URL: <http://arxiv.org/abs/1212.5596>.
- [WN03] G. Wood und M. Newborough. „Dynamic energy-consumption indicators for domestic appliances: Environment, behaviour and design“. In: *Energy and Buildings* 35.8 (2003), S. 821–841. issn: 0378-7788. doi: 10.1016/S0378-7788(02)00241-4.
- [WN07] G. Wood und M. Newborough. „Energy-use information transfer for intelligent homes: Enabling energy conservation with central and local displays“. In: *Energy and Buildings* 39.4 (2007), S. 495–503. issn: 0378-7788. doi: 10.1016/j.enbuild.2006.06.009.
- [Wor02] World Wide Web Consortium (W3C). *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification*. Online. Abgerufen am 09.06.2016. Apr. 2002. URL: <https://www.w3.org/TR/P3P/>.
- [Zho+12] L. Zhou, Y. Tian, S. B. Roy, C. Thorncroft, L. F. Bosart und Y. Hu. „Impacts of wind farms on land surface temperature“. In: *Nature Climate Change* 2.7 (2012), S. 539–543.
- [ZR11] M. Zeifman und K. Roth. „Nonintrusive appliance load monitoring: Review and outlook“. In: *Consumer Electronics, IEEE Transactions on* 57.1 (2011), S. 76–84.