



Finale Studienergebnisse



Volker Markl Thomas Hoeren Helmut Krcmar
Innovationspotenzialanalyse für die neuen Technologien für das Verwalten und Analysieren von großen Datenmengen (Big Data Management)



November 2013

Innovationspotentialanalyse für die neuen Technologien für das Verwalten und Analysieren von großen Datenmengen (Big Data Management)

Autorenkollektiv

Volker Markl

Alexander Löser

Thomas Hoeren

Helmut Krcmar

Holmer Hemsen

Michael Schermann

Matthias Gottlieb

Christoph Buchmüller

Philip Uecker

Till Bitter

Danksagung

Die vorliegende Studie wurde durch einen Auftrag des BMWi zum Thema „Innovationspotentialanalyse für die neuen Technologien für das Verwalten und Analysieren von großen Datenmengen (Big Data Management)“ ermöglicht. Für das in uns gesetzte Vertrauen und die finanzielle Unterstützung möchten wir uns daher beim BMWi bedanken. Zudem möchten wir Frau Dr. Regine Gernert (Projektträger DLR) ganz herzlich Dank sagen für die fachliche und organisatorische Begleitung der Studie. Wir bedanken uns zudem bei den zahlreichen Mitarbeitern der jeweiligen Institute, die durch kritische Durchsicht der Studie und Kommentare dazu beigetragen haben die Studie zu verbessern. Für die Umsetzung des finalen Layout der Studie bedanken wir uns ganz herzlich bei Kerstin Forster. Im Rahmen der Studie wurden zwei Expertenworkshops durchgeführt, wir bedanken uns bei den im Workshop anwesenden Fachexperten, die durch Diskussionen die Studie positiv beeinflusst haben.

Unser Dank geht zudem an die zahlreichen Fachexperten, die uns durch Beiträge, Anregungen und konstruktive Kritik unterstützt haben.

Stefan Bender, Forschungsdatenzentrum der Bundesagentur für Arbeit

Ulf Brefeld, Technische Universität Darmstadt

Teresa Book, Technische Universität Berlin

Markus Böhm, Technische Universität München

Konrad Dongus, Technische Universität München

Stefan Hörmann, Technische Universität München

Fabian Hüske, Technische Universität Berlin

Johannes Kirschnick, Technische Universität Berlin

Dietrich Klakow, Universität des Saarlandes

Andreas Rauber, Technische Universität Wien

Steffen Staab, Universität Koblenz

Michael Säcker, ParStream GmbH

Matthias Scheffler, Fritz-Haber-Institut der Max-Planck-Gesellschaft

Christof Schütte, Bio Computing Gruppe der Freien Universität Berlin

Gottfried Vossen, Universität Münster

Michael Wiegand, Universität des Saarlandes

Kurzbeschreibung

Durch die Digitalisierung von Wirtschaft und Gesellschaft ist ein rasantes Anwachsen von Datenbeständen zu beobachten. In fast allen Unternehmens- sowie Wissenschaftsbereichen werden bereits heute schon Unmengen an Daten erzeugt, deren Größe, Erfassungsgeschwindigkeit oder Heterogenität die Fähigkeiten gängiger Datenbanksoftwareprodukte zur Verwaltung und zur Analyse übersteigt. Dieses Phänomen, welches unter dem Schlagwort „Big Data“ populärisiert wurde, stellt eine große Chance für Unternehmen, Wissenschaft und Gesellschaft dar. Allerdings ergibt sich aufgrund der neuen Komplexität der Daten und Analysen eine Vielzahl an Herausforderungen technischer, wirtschaftlicher und rechtlicher Natur. Diese Studie analysiert die Chancen und Herausforderungen von Big Data insbesondere im Hinblick auf eine nachhaltige Wettbewerbsfähigkeit Deutschlands.

Zusammenfassung

Big Data wurde von McKinsey als die nächste Grenze für Innovation, Wettbewerb und Produktivität identifiziert. Die nachhaltige Sicherung der Wettbewerbsfähigkeit des Wirtschaftsstandorts Deutschlands sowie die Sicherung zukünftiger Innovationen in der modernen Informationsgesellschaft erfordert ein in sich stimmiges Zusammenspiel aus vier Bereichen:

- **Technologie:** Bereitstellung von effektiven Methoden und Werkzeugen zur Analyse von großen Mengen heterogener Daten mit hoher Datenrate,
- **wirtschaftliche Verwertung:** Schaffung von Anwendungen zur Erschließung von neuen Märkten oder Stärkung existierender Märkte,
- **juristische Rahmenbedingungen** und
- **die Ausbildung von Fachkräften.**

Die Entwicklung konvergenter Informations- und Kommunikationstechnologie (IKT) im vorwettbewerblichen Bereich für das „Big Data Management“ ist ein folgerichtiger Schwerpunkt der Technologieförderung des Bundesministeriums für Wirtschaft und Technologie (BMWi). Die vorliegende Studie bewertet dieses Zusammenspiel, um durch das Verwalten und Analysieren von „Big Data“ zu „Smart Data“ zu gelangen.

Die technologischen Herausforderungen exponentiell wachsender Datenvolumina, kontinuierlicher Datenströme mit hohen Datenraten, enger Analysezeitfenster, enormer semantischer Heterogenität von Daten und die Zunahme von Audio- und Videoquellen neben Text und strukturierten Daten sind groß. Die Analyse dieser Daten und die daraus abgeleiteten Empfehlungen für Maßnahmen ermöglichen Unternehmen jedoch einen gewaltigen Informations- und Wettbewerbsvorsprung gegenüber ihren Mitbewerbern. Ferner können dadurch gänzlich neue Geschäftsmodelle entstehen.

Zentrales Ziel der vorliegenden Studie ist eine qualitative und quantitative Bewertung des Ist-Zustandes sowie der wirtschaftlichen Potenziale von neuen Technologien für das Management von Big Data. Daraus werden

standortspezifische Chancen und Herausforderungen für Deutschland abgeleitet. In der Studie werden insbesondere auch rechtliche Rahmenbedingungen anhand von Einzelfällen betrachtet. Die Studie beinhaltet zudem konkrete Handlungsempfehlungen. Diese adressieren die Ausrichtung der Wirtschafts- und Technologiepolitik der Bundesregierung einerseits, und die Unternehmensstrategien von IKT-Anbietern wie -Anwenderunternehmen andererseits.

Die Inhalte der Studie basieren auf umfangreicher Sekundärforschung, zahlreichen Expertengesprächen, den Diskussionen bei zwei Expertenworkshops, der statistischen Auswertung von Unternehmensbefragungen, Modellrechnungen und Instrumenten der Marktanalyse.

Kapitel 1 liefert als Hintergrund einen Überblick über die derzeitigen Veränderungen des wirtschaftlichen, rechtlichen und technologischen Umfelds, in dem Unternehmen heute agieren. Damit einhergehend wird herausgestellt, welche zentralen Anforderungen eine zukunftsfähige IKT erfüllen muss, um Unternehmen angemessen beim Management von Big Data zu unterstützen. Ferner werden wesentliche Konzepte und Technologien für das Management von Big Data dargestellt und potenzielle Chancen, Treiber und Hemmnis-Faktoren für deren Nutzung aufgezeigt. Darauf aufbauend werden erste mögliche Zukunftsaussichten in den einzelnen Technologiebereichen aufgezeigt.

Kapitel 2 skizziert Anwendungsszenarien für das Big Data Management in fünf ausgewählten Branchen und illustriert diese anhand von Fallbeispielen. Die Be- trachtung der Anwenderbranchen basiert auf zahlreichen Experteninterviews und umfangreichen Recherchen seitens des Autorenteams.

Kapitel 3 stellt die wesentlichen Technologien für das Big Data Management vor, wie neue Hardwarearchitekturen, Technologien zur interaktiven Analyse von Daten bzw. statistische und heuristische Technologien zur semantischen Analyse. Oft fehlen insbesondere kleineren Anbietern die Marktübersicht und auch das Know-how für die Kombination der richtigen Technologien, um Daten zu analysieren und Big Data Anwendungen zu erstellen. Insbesondere Datenmarktplätze können diesen Unternehmen helfen. Am Beispiel der Datenmarktplätze stellen wir in dem Kapitel wichtige Markakteure sowie deren Strategien und Marktpotenziale vor.

Kapitel 4 gibt einen Überblick über die deutsche Rechtslage im Hinblick auf das Big Data Management. Hierbei werden Forschungsperspektiven in den Bereichen Urheber-, Datenschutz-, Vertrags- und Haftungsrecht aufgezeigt. Eine umfassende Analyse rechtlich relevanter Fragestellungen befindet sich zusätzlich im Anhang der Studie. Kapitel 4 dient dazu, das derzeitige rechtliche Umfeld zu bewerten, um so zu einer Steigerung der Wettbewerbsfähigkeit der deutschen IKT-Wirtschaft beizutragen.

Kapitel 5 skizziert, basierend auf den vorhergehenden Kapiteln, ein Zukunftsbild für das Big Data Management. Aufbauend auf diesem Gesamtbild werden die wesentlichen Chancen und Herausforderungen für den Standort Deutschland im Hinblick auf diese Technologien und neue Geschäftsmodelle anhand von 10 Thesen herausgearbeitet.

Abschließend werden in Kapitel 6 konkrete Handlungsfelder und -empfehlungen formuliert. Sie dienen als Orientierungshilfe für die Ausrichtung der Wirtschafts- und Technologiepolitik der Bundesregierung einerseits und der Strategie der Unternehmen andererseits, um die Position deutscher Unternehmen im Management von Big Data zu stärken.

Inhaltsverzeichnis

„Big Data“ Management: Eine Chance für Innovation in Europa	8
3.1 Was ist die neue Qualität von Big Data?	9
3.2 Wirtschaftliche und technologische Rahmenbedingungen	11
3.3 Rechtliche Aspekte von Big Data	12
3.4 Was leisten zukünftige Systeme für das Big Data Management?	13
3.5 Ausbildung an den Hochschulen	15
Innovationspotenzial in ausgesuchten Branchen	16
2.1 Hohe Erwartungen, aber vorerst nur Piloten	17
2.2 Erhebung zu Innovationspotenzialen	45
2.3 Ableitung von Innovationspotenzialen	67
2.4 Branchenspezifische Innovationspotenziale	75
2.5 Geschäftsmodellorientierte Analyse der Innovationspotenziale	93
2.6 Priorisierung der Schwerpunkte	95
2.7 Zusammenfassung	106
Analyse von Big Data & Big Data Technologien	110
3.1 Der Datenanalyseprozess	111
Darstellung rechtlicher Rahmenbedingungen in Deutschland	164
4.1 Eigentum an Daten	165
4.2 Urheberrecht	170
4.3 Datenschutzrecht	176
4.4 Vertragsrecht	183
4.5 Informationshaftung und Informationsqualität	186
Der Blick in die Zukunft: 10 Kernthesen	190
5.1 Die 10 Kernthesen	190
5.2 Ein Stimmungsbild zu den 10 Thesen	194
5.3 Zukunftsszenarios	197
5.4 Zwischenfazit: Wegbereitung einer disruptiven Technologie	201

Empfehlungen für Entscheider	204
6.1 Prämissen.....	205
6.2 Handlungsempfehlungen	208
Glossar	218
Literatur	226
Abbildungsverzeichnis	248
Tabellenverzeichnis	254
Anhang: Big Data und Recht.....	256

1 „Big Data“ Management: Eine Chance für Innovation in Europa

Die zunehmende Vernetzung im Internet der Dinge und Dienste sowie die Verwendung von fortgeschrittenen Sensorik und Simulationsmodellen in Industrie 4.0, bei Dienstleistungsunternehmen, in der Forschung und im privaten Sektor führen zu einer immer größeren Verfügbarkeit von Daten. Die Analyse dieser Daten wird wirtschaftliche, wissenschaftliche und gesellschaftliche Prozesse revolutionieren, durch eine zeitnahe und umfassende datengetriebene Entscheidungsunterstützung. Insbesondere für Unternehmen sind hierdurch erhebliche Wettbewerbsvorteile zu erwarten (114).

Dieser Trend wird derzeit durch die Schlagwörter „Big Data“ oder „Data Science“ umschrieben. Dabei bedeutet „Big Data“, dass sowohl Daten als auch Analysen auf diesen Daten in den letzten Jahren eine neue Form der Qualität und Komplexität erreicht haben. Abbildung 1 zeigt die wesentlichen Querverbindungen des Zusammenspiels zwischen Technik im Bereich der Datenanalysealgorithmen, Datenanalysesysteme, und Anwendungen sowie den wirtschaftlichen und rechtlichen Rahmenbedingungen. Im den folgenden Abschnitten dieses Kapitels beschreiben wir näher dieses Zusammenspiel.

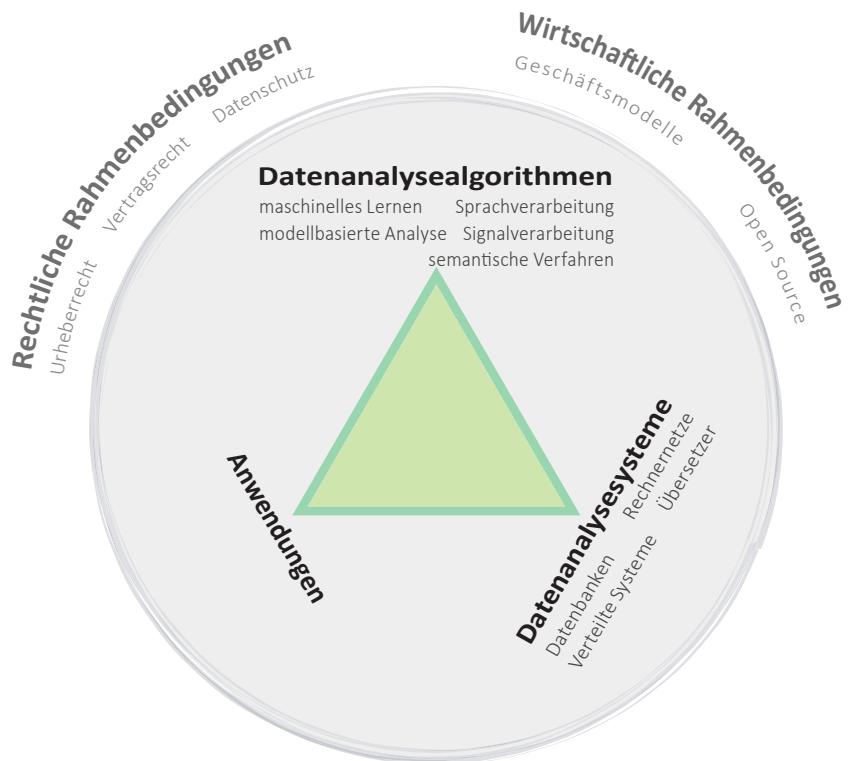


Abbildung 1: Zusammenspiel der Rahmenbedingungen

1.1 Was ist die neue Qualität von Big Data?

„Big Data“ zeichnet sich durch eine neue Komplexität im Hinblick auf die Daten aus, sowie durch eine neue Komplexität der Analysen, die auf diesen Daten durchgeführt werden.

Die neue Art der **Datenkomplexität** wird dabei durch Anforderungen an Datenvolumen (engl. volume), Datenrate (engl. velocity), Datenheterogenität (engl. variety) und Datenqualität (engl. veracity) charakterisiert, welche von handelsüblichen Datenbanksystemen nicht abgedeckt werden können. So erfordert die Analyse von Big Data die Speicherung und Verarbeitung von riesigen Datenmengen im Terabyte- oder Petabytebereich. Gleichzeitig werden die Entscheidungszeitfenster, in denen Analyseergebnisse bereitgestellt werden müssen, immer kürzer. Datenanalysesysteme müssen akkurate Analysen mit geringem Zeitverzug liefern und dies trotz potentiell hoher Datenraten, mit denen neue Daten in die Datenbasis integriert werden. Gleichzeitig werden eine Vielzahl von unterschiedlichen Datenquellen in die Datenanalyse einbezogen, welche Daten in unterschiedlichsten Formaten speichern (z.B. Zeitreihen, Tabellen, Textdokumenten, Bilder, Audio- und Videodatenströmen). Diese Daten müssen zunächst durch Informationsextraktionsverfahren in ein einheitliches, semantisch

wohldefiniertes Format überführt und dann integriert werden, ehe eine Analyse auf diesen sehr unterschiedlichen Daten durchgeführt werden kann. Aufgrund der Unschärfe von einigen Datenquellen (z.B. Sensoren mit fixer Genauigkeit) bzw. von Informationsextraktionsverfahren und Integrationsverfahren sind die Ergebnisse vieler Datenanalysen probabilistisch. Daher müssen Systeme und Analysten mit wahrscheinlichkeitsbasierten Modellen und Konfidenzen im Rahmen der Analyse von Big Data umgehen.

Neue statistische und mathematische Algorithmen, Vorhersagetechniken und Modellierungsmethoden, neue Ansätze für die Datenerfassung und –integration, Datenanalyse und –kompression, verbesserte Technologien für die Datenverarbeitung und die gemeinsame Nutzung, neue deklarative Sprachen für Spezifikation und automatische Optimierung und Parallelisierung von komplexen Datenanalyseprogrammen werden benötigt um das Datenvolumen, die Verarbeitungsgeschwindigkeit, die unterschiedlichen Datenformate und die Vertrauenswürdigkeit der Daten in den Griff zu bekommen.

Die neue Art der **Analysekomplexität** von „Big Data“ zeigt sich daran, dass zur Entscheidungsunterstützung Modelle aus den Daten generiert werden. Dies erfordert den Einsatz von fortgeschrittenen Algorithmen der Datenanalyse, insbesondere statistischer Verfahren, Verfahren des maschinellen Lernens, der linearen Algebra und Optimierung, Signalverarbeitung sowie des Data Mining, des Text Minings, des Graph Minings, Video Minings und der visuellen Analyse. So muss das Datenanalysesystem komplexe Algorithmen der linearen Algebra, Statistik oder Optimierung zeitnah verarbeiten. Diese Algorithmen zeichnen sich durch die Verbindung von benutzerdefinierten Funktionen und iterativen, zustandsbehafteten Algorithmen mit den üblichen Operationen der relationalen Algebra aus. Die Kombination von relationaler Algebra mit iterativen Algorithmen und benutzerdefinierten Funktionen wird weder von den klassischen SQL-Datenbanksystemen noch den Big Data Lösungen (Hadoop, Pig, Hive, Storm, Lambda Architektur, etc.) realisiert. Hierdurch ergibt sich hohes Innovations- und Marktpotential für die Entwicklung und Kommerzialisierung moderner Datenanalysesysteme, die relationale Datenverarbeitung mit Algorithmen insbesondere des maschinellen Lernens und der Statistik vereint. Zur breiten Nutzung derartiger Analysesysteme zur Modellbildung in Wirtschaft und Wissenschaft sind niedrige Einstiegsschwellen für die Nutzer erforderlich. Dies bedeutet, dass die Spezifikation der Datenanalysealgorithmen deklarativ erfolgen muss und keine Kenntnisse in der Systemprogrammierung und Parallelverarbeitung erfordern darf, wie es bei derzeitigen Anwendungen des wissenschaftlichen Rechens, aber auch bei aktuellen Big Data Systemen, z.B. im populären Map/Reduce-Programmiermodell, der Fall ist.

Diese Anforderungen werden zu einem Paradigmenwechsel bei Datenanalyse-sprachen, Datenanalysesystemen und auch Datenanalysealgorithmen führen

und völlig neuartige Anwendungen ermöglichen. Neuartige Datenanalysesysteme sowie Fortschritte bei der Informationsverarbeitung, der Integration, der Signalverarbeitung, im Bereich des maschinellen Lernens, des Data-Minings, der Datenkompression und Visualisierung werden neue Wege ermöglichen um nützliche, vertrauenswürdige und verifizierte Informationen zeitnah aus riesigen Datenmengen unterschiedlicher Natur zu extrahieren, abzuleiten und anzuwenden.

1.2 Wirtschaftliche und technologische Rahmenbedingungen

Das Thema „Big Data“ hat große Beachtung in der Gesellschaft, der Industrie und Forschung erfahren. Zum Beispiel beschreibt McKinsey „Big Data“ als die nächste große Barriere für Innovation, Wettbewerb und Produktivität. IBM hat neben anderen Firmen ein firmenweites Forschungs- und Innovationsprogramm zu massiver skalierbarer Datenanalyse gestartet. In den USA hat Präsident Obama im Jahre 2012 eine 200 Millionen US-Dollar Big Data Initiative gestartet (286). Im Rahmen dieser Initiative haben u.a. die National Science Foundation (NSF) und das Gesundheitsinstitut National Institut of Health (NIH) ein gemeinsames Forschungsprogramm gestartet mit dem Ziel Forschung im Bereich der Kern-technologien und –methoden von „Big Data Forschung & Technik“ fortzuentwickeln. Führende Forscher von Hochschulen und Industrie in den USA haben ein Thesenpapier verfasst, das die Herausforderungen im Bereich „Big Data“, inklusive deklarativer Datenprogrammiersprachen, skalierbarer Verarbeitungssysteme für Datenanalyse, natürlich-sprachigen Textdaten und Audio- oder Videodaten hervorhebt (285). Die britische Regierung hat zusammen mit einer privaten Stiftung eine 90 Millionen GBP Initiative gestartet, um an der Oxford Universität durch Aufbau eines Big Data Institutes die Forschung im Gesundheitswesen voranzutreiben und bei der Erforschung neuer Arzneimitteln zu unterstützen (287). Die japanische Regierung hat angekündigt Big Data Analysen zu verwenden, um zeitnahe Konjunkturprognosen zu erstellen (288).

Die neuen Herausforderungen von „Big Data“ stellen eine **große Chance für deutsche und europäische Firmen** dar, sowohl was die Technologie betrifft als auch bei Anwendungen für dieses Gebiet. Existierende Produkte im kommerziellen, überwiegend von US-Firmen dominierten Datenbankmarkt basieren auf Technologien, die aufgrund zu geringer Skalierbarkeit, mangelnder Fehlertoleranz, oder eingeschränkter Programmiermodelle den Anforderungen von Big Data nicht gewachsen sind. Somit werden derzeit die Karten im Bereich der **skalierbaren Datenanalysesysteme** international neu gemischt. Dabei ist Deutschland gut positioniert. Neben den USA besitzt Deutschland die zweitstärkste Forschungs-Gemeinschaft im Bereich des skalierbaren Datenmanagements. Diese führt bereits sehr viele Aktivitäten in der Grundlagenforschung zu Big Data durch (z.B. das Stratosphere-System der TU Berlin, das Hyper der TU München, die Forschungen zu Hadoop++ und HAIL an der Universität

Saarbrücken), die auch in Unternehmensgründungen im Bereich der Datenanalysetechnologie resultieren können. Neben einer auch in Deutschland aktiven Open-Source-Bewegung fordern derzeit viele Unternehmen, insbesondere Startups, die etablierten großen Anbieter wie IBM, Oracle und Microsoft heraus. Unter den Herausforderern, die sich Chancen und Marktanteile in dem gerade entstehenden Big Data Markt versprechen, finden sich dabei auch eine Vielzahl an deutschen Technologieunternehmen, neben Großunternehmen wie z.B. SAP mit HANA und der Software AG mit Terracotta auch viele High-Tech-Startups wie z.B. ParStream, Exasol und Q2Web. Um diese Unternehmen in ihrer Produkt- und Markteroberungsstrategie zu unterstützen, ist es wichtig, ein Technologie-transfer- und Innovationsklima zu schaffen, das es deutschen Unternehmen, besonders den mittelständischen Unternehmen, universitären Ausgründungen und Startups im Bereich der skalierbaren Datenverarbeitung sowie der Datenanalyse ermöglicht, mit den Startups insbesondere im Silicon Valley, in England und in China auf Augenhöhe im Bereich der Technologieentwicklung und Mitarbeitergewinnung zu stehen. Auf diese Weise kann die Politik einen wichtigen Beitrag für die Wettbewerbsfähigkeit deutscher Unternehmen bei der zukünftigen Entwicklung und Kommerzialisierung von Big Data Schlüsseltechnologien leisten und Deutschland durch wissenschaftliche Leistungen und Innovationen den Weg in den Milliardenmarkt Big Data ebnen. Dabei ist darauf hinzuweisen, dass die Schaffung und Erhaltung von deutschen und europäischen Datenanalyse-technologien und –systemen zur Erreichung der Unabhängigkeit von Anbietern aus den USA und Fernost auch aus datenschutzrechtlichen und sicherheitspolitischen Aspekten eine wichtige nationale Aufgabe ist, so dass es hier Handlungsbedarf seitens der Politik gibt.

Gleichzeitig ist die **Informationswirtschaft basierend auf „Big Data“** im Kommen. Europäische Firmen, wie z.B. DataMarket, Internet Memory Foundation, Okkam, VICO Research & Consulting und das deutsche Forschungsprojekt MIA erstellen **Informationsmarktplätze**, die anderweitig verbreitete Information bündeln und für Analysen bereitstellen. Diese Datenmarktplätze bieten die Chance, das gegenwärtig von US-Firmen gehaltene Monopol der Datenbereitstellung im Großen zu brechen. Insbesondere werden es Informationsmarktplätze gerade kleinen und mittelständischen Unternehmen (KMU) ermöglichen, diese Daten zu analysieren und wirtschaftlich zu nutzen und somit einen Wettbewerbsnachteil auszugleichen.

1.3 Rechtliche Aspekte von Big Data

Die technische Entwicklung der letzten Jahre hat zu einem raschen Anwachsen der verfügbaren Datenmengen geführt. Der Einsatz verschiedener Technologien ermöglicht nicht nur die quasi unbegrenzte Langzeitspeicherung dieser Daten,

sondern auch deren detailgenaue Analyse. Die hierbei herangezogenen großen Datenmengen bestehen beispielsweise aus Webcrawls, Transaktionsdaten, Sensordaten oder Daten aus Simulationen.

Diese Form der Datenspeicherung und –nutzung betrifft dabei quasi alle Bereiche des täglichen Lebens und erlangt zunehmend wirtschaftliche Bedeutung. Der Datenumgang wirft vielfältige rechtliche Fragen auf, die vor allem die Rechtsdisziplinen des Urheberrechts, des Vertragsrechts und des Datenschutzrechts betreffen.

Die systematische Speicherung großer Datenmengen im Rahmen von Big Data wirft unter anderem **urheberrechtliche Fragen** nach den Rechten des Datenbankherstellers gem. §§ 87a ff. UrhG auf. Datenbanken an sich können sowohl einen urheberrechtlichen Schutz, als auch einen *sui-generis*-Schutz genießen. Die Auswertung von Datenbanken kann folglich auch das Urheberrecht verletzen.

Die neuen technischen Möglichkeiten der Datenanalyse müssen ferner auch unter Gesichtspunkten der Kautelarpraxis untersucht werden, um so eine **Vertragsgestaltung** zu ermöglichen, die rechtliche Probleme bereits im Vorfeld vermeidet.

Die Wahrung der Grundsätze des deutschen **Datenschutzrechts** und damit der Persönlichkeitsrechte stellt ebenfalls eine zentrale Herausforderung dar. Hierbei beachtlich ist zunächst die Frage nach der grundsätzlichen Anwendbarkeit der deutschen Rechtsordnung. Daten sind heute, besonders in der Cloud, hochgradig standortflexibel und eine Speicherung und Verarbeitung kann beinahe beliebig und an verschiedenen Orten zugleich stattfinden. Soweit deutsches Datenschutzrecht anwendbar ist, gilt ein Verbot der Datenverarbeitung mit Erlaubnisvorbehalt – hier gilt es, die Erlaubnistatbestände auf ihre Vereinbarkeit mit bzw. Anwendbarkeit auf Big Data zu untersuchen. Besondere Bedeutung erlangen hier Anonymisierung und Pseudonymisierung, die die Möglichkeiten des Datenumgangs erweitern können.

1.4 Was leisten zukünftige Systeme für das Big Data Management?

Wie bereits oben skizziert, stellt „Big Data“ neue Anforderungen an die Datenhaltungs- und -analysesysteme, die der derzeitige Stand der Technik nicht leisten kann. So müssen derartige Systeme riesige Datenmengen, im Tera- oder Petabytebereich, oder in manchen Anwendungen sogar größer, verwalten können. Ferner müssen Systeme akkurate nichtlineare Analysen mit geringem Zeitverzug liefern und dies trotz potentiell hoher Datenraten, mit denen neue Daten in

die Datenbasis integriert werden. Gleichzeitig sind diese Daten oftmals heterogen und liegen in unterschiedlichen Formaten vor (z.B. relational, hierarchisch oder als Graph strukturiert bzw. in Form von Text-, Bild-, Audio-, oder Videodaten). Aufgrund der Unschärfe von Experimenten, Simulationen und Algorithmen spielt die Konfidenz in Analyseergebnisse eine wichtige Rolle. Ferner ist die Komplexität der Analysen durch Einsatz von Verfahren des maschinellen Lernens bzw. der Signal-, Sprach-, Bild- und Videoverarbeitung drastisch gestiegen. So muss ein modernes Datenanalysesystem komplexe Algorithmen der linearen Algebra, Statistik oder Optimierung zeitnah verarbeiten, welche häufig aus benutzerdefinierten Funktionen und iterativen, zustandsbehafteten Algorithmen bestehen. Dies geht weit über die üblichen Operationen der relationalen Algebra (und damit von traditionellen SQL-Datenbanksystemen) und dem Stand der Technik aktueller Big Data Lösungen (Hadoop, Pig, Hive, Storm, Lambda-Architektur, etc.) hinaus. Die Nutzung solcher Analysesysteme zur Modellbildung in Wirtschaft und Wissenschaft erfordert niedrige Einstiegsschwellen für die Nutzer, so dass die Spezifikation der Datenanalysealgorithmen deklarativ erfolgen muss. Dazu müssen die Disziplinen skalierbare Datenverarbeitung sowie Datenanalyse verschmolzen werden, um Technologien zu schaffen, die Skalierbarkeit und Intelligenz kombinieren. Hierfür müssen neue Konzepte der Datenanalyse und des maschinellen Lernens entwickelt werden, die eine Skalierbarkeit und Parallelisierbarkeit algorithmisch mitdenken. Diese sollten von ausgewählten Anwendungsschwerpunkten inspiriert sein, welche die für die Algorithmen- und Systemforschung relevanten unterschiedlichen Aspekte des Big Data Spektrums in Bezug auf Datenvolumen, Datenrate (statische Daten vs. dynamische Datenströme) und Heterogenität (numerische Daten aus Experimenten und Simulationen, Textdatenströme, Videodatenströme, Bilddaten) sowie Komplexität der Analysen (Aggregation, relationale Operationen, iteratives maschinelles Lernen, zustandsbehaftete Modellbildung) abdecken. Solcherart fortgeschrittene mathematische und statistische Datenanalyseverfahren (z.B. Optimierung, maschinelles Lernen, Signalverarbeitung, Sprachverarbeitung, Visualisierung) können durch iterative Algorithmen auf potentiell geordneten Kollektionen, benutzerdefinierte Funktionen sowie durch den Zugriff auf einen verteilten Zustand charakterisiert werden. Daher sollte ein großes Augenmerk auf der automatischen Optimierung sowie der Parallelisierung und Adaption von, in einer deklarativen Sprache spezifizierten, Datenanalyseverfahren auf neuen Rechnerarchitekturen liegen. Dazu müssen Kompetenzen in den Bereichen der skalierbaren Datenverarbeitungssysteme mit denen der Datenanalyse und des maschinellen Lernens kombiniert werden. Zur Schaffung einer Ausführungsplattform für derartige komplexe Datenanalysen müssen ferner Kompetenzen aus den Bereichen der Datenbanksysteme, der verteilten Systeme und der Rechnernetze kombiniert werden. Nur dadurch kann den Herausforderungen bzgl. Skalierbarkeit und Latenz der Analyseverfahren sowie der Heterogenität der Daten begegnet werden.

1.5 Ausbildung an den Hochschulen

Um die Wirtschaft, Wissenschaft und Gesellschaft in Deutschland auf diesen globalen Trend optimal vorzubereiten, sind hochgradig koordinierte Aktivitäten in der Forschung, der Lehre und dem Technologietransfer im Bereich der Datenanalysemethoden und skalierbaren Datenbanksysteme erforderlich. Big Data ist nicht mehr länger nur eine Herausforderung für eine spezifische Branche, es tangiert vielmehr alle Wirtschaftszweige, alle Organisationen und alle Nutzer von digitalen Technologien. Aufgrund der großen wirtschaftlichen Bedeutung stellt sich die grundlegende Frage, welche Kenntnisse erforderlich sind, um durch die Analyse von „Big Data“ Mehrwert für den jeweiligen Wirtschafts- bzw. Technologiebereich zu generieren.

McKinsey (114) prognostiziert einen riesigen Bedarf an Datenanalyseexperten, sogenannten **Data Scientists**. Dieses neuartige Berufsbild kombiniert Methodenkenntnisse in Datenanalyseverfahren (Statistik und maschinelles Lernen, Optimierung, linearer Algebra, Signalverarbeitung, Sprachverarbeitung, Data-Mining, Text Mining, Video Mining, Bildverarbeitung) mit technischen Kompetenzen im Bereich des skalierbaren Datenmanagements (Datenbanksysteme, Data-Warehousing, Informationsintegration, verteilte Systeme, Rechnernetze, Rechnerarchitekturen) sowie praktischen Kompetenzen der Systemimplementierung. Dabei sollte eine derartige Ausbildung durch praktische Anwendungsprojekte zur Vermittlung von Fachkompetenzen in bestimmten Anwendungsbereichen flankiert werden. In den USA wurden bereits verschiedene Programme realisiert, z.B. an der North Carolina State University, NCSU.

2 Innovationspotenzial in ausgesuchten Branchen

In den folgenden Abschnitten wird das Innovationspotenzial von Big Data sowohl in ausgewählten Branchen als auch branchenübergreifend untersucht. Dafür werden zunächst existierende Studien zusammengefasst und die Kernaussagen extrahiert. Anschließend werden diese Kernaussagen in einer eigenen empirischen Untersuchung untermauert sowie erste Anforderungen für eine Ausschöpfung des Innovationspotenzials abgeleitet. Anschließend werden Anwendungsbeispiele für Big Data in einzelnen Branchen dargestellt.

Die Kernaussagen dieser Abschnitte sind:

- Die Innovationspotenziale von Big Data werden von Unternehmen, auch kleinen und mittelständischen Unternehmen, erkannt und wurden zum Teil bereits realisiert.
- Damit die Innovationspotenziale in Deutschland realisiert werden können, gilt es konkrete Kombinationen von Datenarchiven, Datenanalysewerkzeugen und Anwendungsszenarien zu entwickeln. Hierfür besteht erhebliches Kooperationspotenzial bei Anbietern und Anwendern von Big Data-Technologien.

Die Entwicklung von Big Data setzt zum einen unternehmerische Kreativität und zum anderen systematische Geschäftsmodellentwicklung voraus. Big Data ist besonders bei Dienstleistungen von Relevanz. Die einzelnen Branchen werden stärker, gemeinsam und interdisziplinärer zusammenarbeiten müssen, da Big Data einen fließenden Übergang zwischen den Branchen schaffen kann. Insgesamt können die Potenziale von Big Data dann realisiert werden, wenn breitenwirksame Referenzbeispiele ermöglicht werden, welche die Wettbewerbsfähigkeit in deutschen Schlüsselbranchen verbessern. Dabei spielen insbesondere kleine und mittelständische Unternehmen eine wichtige Rolle. Diese können in ähnlichem Maße wie große Unternehmen von Big Data profitieren und weisen ein ähnliches Herausforderungsprofil auf.

2.1 Hohe Erwartungen, aber vorerst nur Piloten

Ein Überblick aktueller Studien zu den Innovationspotenzialen von Big Data.

Es gibt bereits viele Studien, Whitepapers, Umfragen und Marktanalysen, die das Innovationspotenzial von Big Data in unterschiedlichsten Branchen und Anwendungsfeldern untersuchen.

Es können Kernaussagen über alle Studien hinweg getroffen werden:

1. Big Data war bisher vor allem ein Thema für IT-Experten, das sonst noch relativ unbekannt ist.
2. Big Data wird die Unternehmensführung, Geschäftsprozesse und die betriebliche Informationslogistik verändern.
3. Konkreter Handlungsbedarf wird aktuell vor allem in der Informationslogistik gesehen.
4. Unternehmen beginnen erst damit, Strategien für den Einsatz von Big Data über die Analyse strukturierter Daten hinaus zu entwickeln.
5. Für den Einsatz von Big Data fehlen noch Fachkräfte sowie organisatorische Strukturen und Prozesse in den Unternehmen. Diese werden in Zukunft noch stärker nachgefragt werden.
6. Mit zunehmender Bereitschaft der Unternehmen wird der Big Data Markt erheblich wachsen.

Im Folgenden wird zunächst die Methodik der Analyse beschrieben und anschließend werden die Kernaussagen detailliert vorgestellt.

2.1.1 Methodik zu Auswahl und Aggregation der Studien

Es gibt bereits eine Vielzahl an Studien zu unterschiedlichen Aspekten von Big Data sowie mit unterschiedlichen regionalen Fokus. Tabelle 1 zeigt die Studien, die für diesen Überblick ausgewählt wurden. Die Studien wurden dabei anhand folgender Kriterien ausgewählt:

- Die Studie muss Aussagen zu Deutschland beinhalten.
- Die Studie muss eine quantitative Datengrundlage haben und Aussagen zur Erhebung beinhalten. Es gibt eine Vielzahl an Studien, die hierzu keine Aussage treffen bzw. auf qualitativem Datenmaterial beruhen. Obwohl diese Studien auch interessante Aspekte von Big Data behandeln, war es für die vorliegende Untersuchung das Ziel quantitative Aussagen treffen zu können.
- Die Studie muss vergleichbare Aussagen beinhalten, die zueinander verknüpft werden können. Aufgrund der Heterogenität der Studien hinsichtlich des Erhebungsinstruments mussten bei diesem Kriterium häufig

Abstriche gemacht werden. Wo möglich wurden die Zahlen ineinander überführt, sonst werden exemplarische Aussagen aus einzelnen Studien aufgeführt.

Wir konnten auf Basis dieser Kriterien die in Tabelle 1 aufgeführten neun Studien zusammenführen. Im Rahmen der Analyse wurden die Themen identifiziert, die möglichst in allen Studien behandelt werden. Jede Studie hat darüber hinaus weitere Spezialaspekte untersucht. Für diese Aspekte wird auf die jeweiligen Studien verwiesen. Ähnliche Themen wurden von drei Personen identifiziert und auf inhaltlicher Ebene aggregiert. Da die Studien bestimmte Sachverhalte (z.B. Potenziale) mit unterschiedlichen Instrumenten abfragen (Rankings, Wachstumszahlen, Likert-skalen), wurde, wo möglich ein Ranking der entsprechenden Aussagen vorgenommen. Wo es nicht möglich war, wurde das Thema anhand einer Studie exemplarisch erläutert. Die Stichprobengröße der untersuchten Studien in Tabelle 1 ist mit $n > 4492$ umfangreich.

Herausgeber	Abkürzung	Studie	Jahr	n	Geographie	Befragte
 BARC	BARC	Big Data Survey Europe	2013	274	Deutschland Frankreich Großbritannien Österreich Schweiz	Entscheidungsträger allgemein und IT
 BITKOM	BITKOM	Umfrage zur Bekanntheit von Big Data	2013	1008	Deutschland	Bevölkerung allgemein
 CG	CG	The Deciding Factor: Big Data and Decision Making	2012	607	Weltweit	Entscheidungsträger allgemein
 CR	CR	Big Data: What's Holding you Back?	2013	n.a.	n.a.	Entscheidungsträger IT

	Experton Group	Datenexplosion in der Unternehmens-IT	2012	100	Deutschland	Entscheidungsträger allgemein und IT
	IAIS	Big Data: Vorsprung durch Wissen	2013	100	Deutschland	Entscheidungsträger allgemein
	Interxion	Big Data: Beyond the Hype	2013	750	Belgien Dänemark Deutschland Frankreich Großbritannien Irland Niederlande Österreich Schweden Schweiz	Entscheidungsträger IT
	TCS	The Emerging Big Returns on Big Data	2013	643	Australien Brasilien Deutschland Großbritannien Indien Japan Mexiko Niederlande USA	Entscheidungsträger allgemein
	TNS	Quo Vadis Big Data	2012	1010	Deutschland Österreich Schweiz Frankreich Russland Südafrika Brasilien USA	Entscheidungsträger IT

Tabelle 1: Überblick der untersuchten Studien (n > 4492)

Im Folgenden werden die Kernaussagen der untersuchten Studien erläutert.

2.1.2 Big Data ist aktuell vor allem ein Thema für IT-Experten, das sonst noch relativ unbekannt ist

Der deutschen Bevölkerung ist der Begriff Big Data noch nicht geläufig. In Unternehmen wird Big Data im Wesentlichen als neues Paradigma für den Umgang mit großen Datenmengen gesehen. Kritische Stimmen sehen Big Data als Marketing-Hype.

Abbildung 2 zeigt, dass der Begriff Big Data für die überwiegende Mehrheit der deutschen Bevölkerung nicht bekannt ist. Lediglich 14 Prozent geben an, die Bedeutung des Begriffs zu kennen.

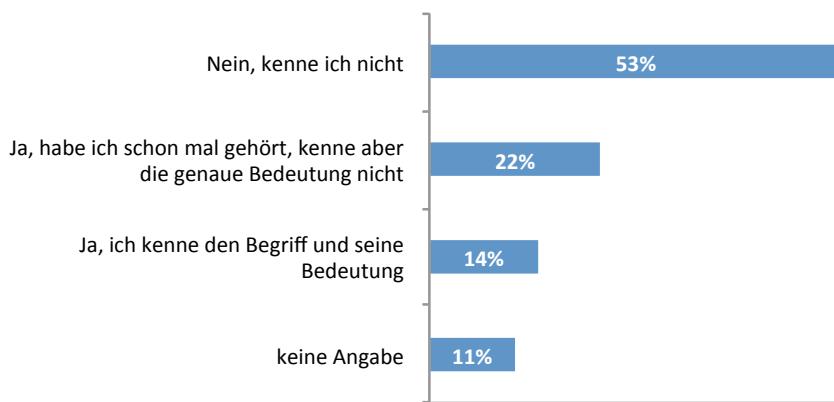


Abbildung 2: Bekanntheit des Begriffs Big Data in der deutschen Bevölkerung in Prozent
(n=1008, BITKOM 2013)

Bei Entscheidern in Unternehmen ergibt sich ein differenziertes Bild. Wie Abbildung 3 zeigt, ist der Begriff Big Data lediglich für 9 Prozent der befragten Entscheider unbekannt.

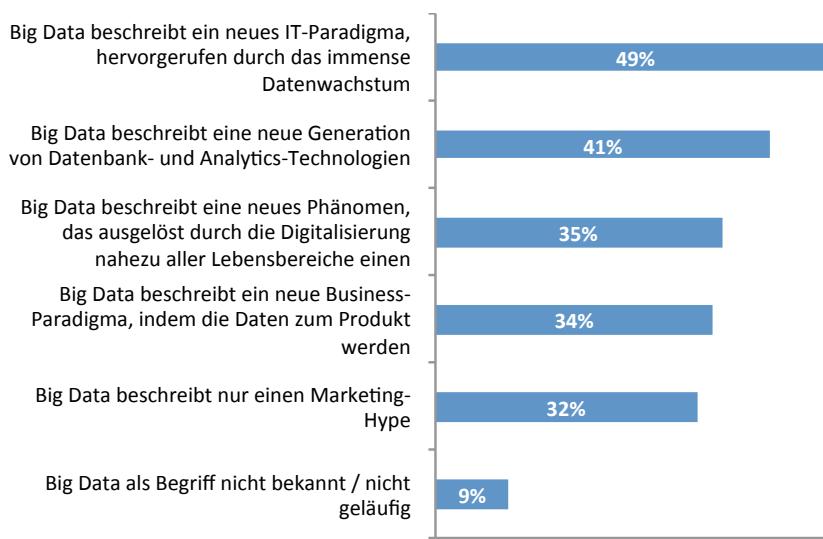


Abbildung 3: Zustimmung zu verschiedenen Beschreibungen des Big Data Begriffs in Prozent (n=79, Experton Group 2012)

Für 49 Prozent der Befragten stellt Big Data ein neues IT-Paradigma dar. 41 Prozent verbinden mit dem Begriff technologische Entwicklungen im Bereich

Datenbanken und Datenanalyse. Kritisch äußern sich 32 Prozent der Befragten, für sie ist Big Data auch ein Marketing-Hype.

Darüber hinaus zeigen andere Studien (Abbildung 4), dass Big Data vor allem innerhalb der IT-Abteilungen von Unternehmen diskutiert wird. Dort wird es von 25 Prozent der Befragten als große Chance für das eigene Unternehmen gesehen.

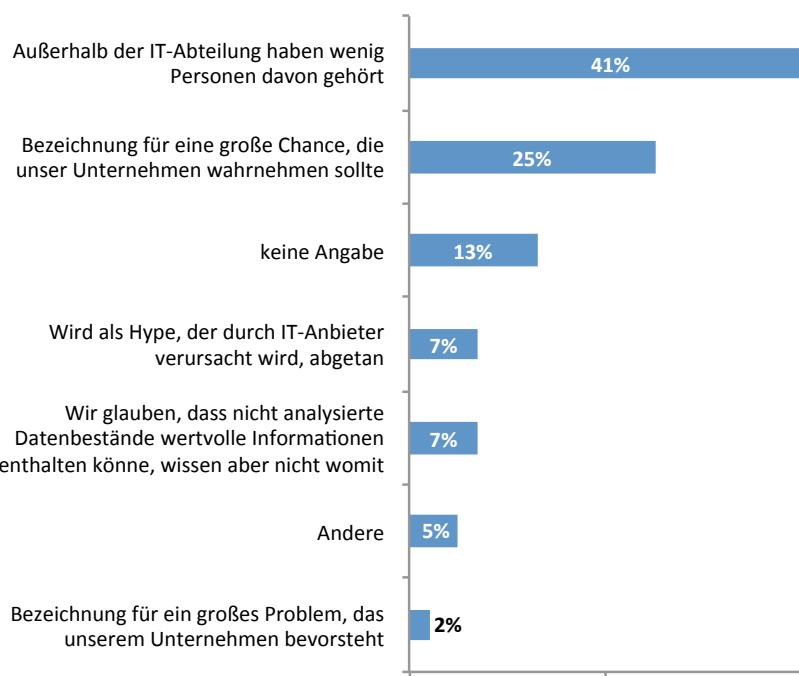


Abbildung 4: Einstellung befragter Unternehmen zu Big Data in Prozent (n=n.a., computing research 2013)

2.1.3 Big Data verändert die Unternehmensführung, Geschäftsprozesse und die betriebliche Informationslogistik

Die Potenziale von Big Data sind aufeinander aufgebaut. Für eine bessere Unternehmensführung sind effektive Geschäftsprozesse notwendig, welche wiederum eine optimale Informationslogistik voraussetzen. Für alle drei Ebenen sind die Potenziale von Big Data erkannt.

Von den neun in diesem Bericht untersuchten Studien gehen nur die Studien von Capgemini und TNS nicht auf Beweggründe von Unternehmen und Anwendern sich mit Big Data zu beschäftigen ein. Die Beweggründe lassen sich zu

dreizehn Kategorien aggregieren. Die fünf bedeutendsten Kategorien an Beweggründen für Big Data sind demnach:

1. Strategische Wettbewerbsvorteile allgemein
2. Verbesserte Entscheidungsfindung
3. Verbessertes Verständnis für Markt- / Kundenbedürfnisse
4. Erhöhung der Kundenzufriedenheit und -bindung
5. Kostenvorteile

Zu beachten ist, dass sich sowohl die Beweggründe als auch deren Kategorien teilweise auf unterschiedlichen Abstraktionsniveaus befinden. So ist bspw. die Kategorie „strategische Wettbewerbsvorteile allgemein“ wesentlich abstrakter als die Kategorie optimale Preisgestaltung. An diesem Beispiel wird auch deutlich, dass kausale Abhängigkeiten zwischen Beweggründen und deren Kategorien bestehen: So kann eine optimale Preisgestaltung zu strategischen Wettbewerbsvorteilen führen. Somit sollte die in Tabelle 2 gezeigte Priorisierung weniger als Rangliste, sondern eher als Überblick über die in den untersuchten Studien von Unternehmen genannten Beweggründe verstanden werden.

	GESAMTRANG ⁰									
	BARC	BITKOM	Capgemini	computing research	Experton Group	Fraunhofer IAIS	interxion	Tata Consultancy Services	TNS	
Strategische Wettbewerbsvorteile allgemein	1	-	5	n.a.	-	n.a.	1	-	-	n.a.
Verbesserte Entscheidungsfindung	2	3	-		4		5	1	-	
Verbessertes Verständnis für Markt- / Kundenbedürfnisse	3	8	2		4		-	-	7	
Erhöhung der Kundenzufriedenheit und -bindung	4	10	-		-		-	2	5	
Kostenvorteile	5	7	2		9		3	-	6	

Steigerung der Innovationskraft und -geschwindigkeit	6	-	-	-	-	8	4	10	
Erhöhung von Produkt- und Servicequalität	7	9	-	-	-	7	-	8	
Zielgenaues Marketing	8	6	1		12	-	-	9	
Operative Effizienzsteigerungen	9	2	-		2	4	6	11	
Verbesserung des Kundenservice	10	5	-		6	6	-	13	
Umsatzsteigerungen	11	-	-		11	2	3	11	
Verbesserung des Risikomanagements und der Compliance	12	-	-		12	9	5	-	
Optimale Preisgestaltung	13	-	4		-	-	-	17	
O: Basierend auf dem Durchschnittsrang über die verfügbaren Studienergebnisse									

Tabelle 2: Aggregation verschiedener Studien zu den Beweggründen für Big Data (Eigene Darstellung)

Im Folgenden werden die Beweggründe für den Einsatz von Big Data auf drei Ebenen detailliert dargestellt.

Auf betriebswirtschaftlicher Ebene verbinden deutsche Entscheider Big Data insbesondere mit dem Aufbau von strategischen Wettbewerbsvorteilen. Darüber hinaus werden die Steigerung der Umsätze, die Einsparung von Kosten und eine Erhöhung der Produktivität mit Big Data in Verbindung gebracht. Diese Potenziale sind vor allem mit dem optimalen Einsatz von Daten in Unternehmen gepaart.



Abbildung 5: Ziele von Big Data Anwendungen in Anzahl Nennungen (n=100, Fraunhofer IAIS 2013)

Betrachtet man die Potenziale von Big Data auf einer **geschäftsprozessorientierten** Ebene, werden bessere strategische Entscheidungen als wichtigster Vorteil von Big Data genannt (59 Prozent Zustimmung). Fast gleich wichtig sind aber ebenso eine bessere Steuerung operativer Prozesse sowie schnellere und detailliertere Analysen. Zusätzlich wird aus einer europäischen Perspektive das in Deutschland wichtige Thema der Kosteneinsparungen durch Big Data geringer eingeschätzt.

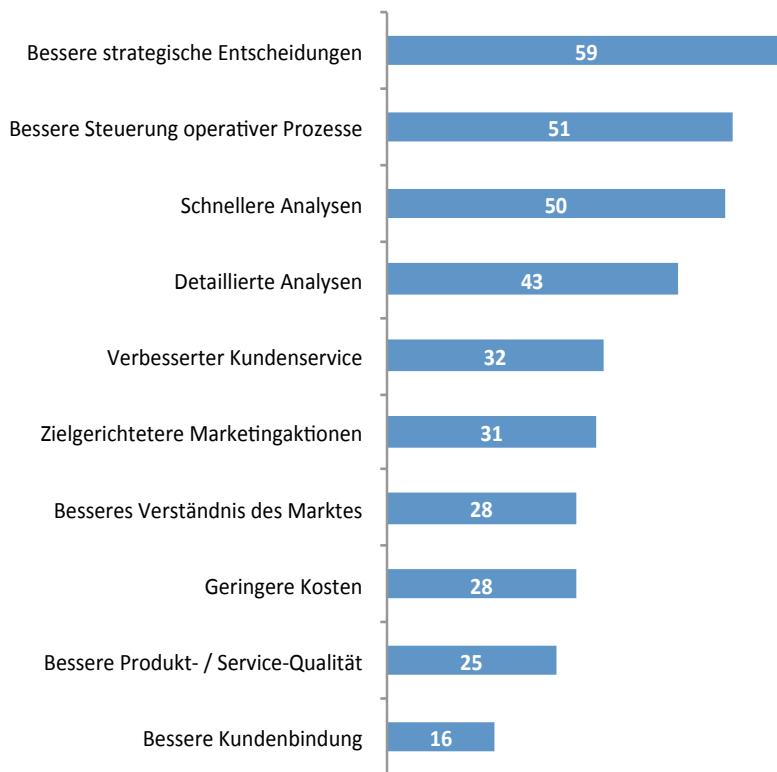


Abbildung 6: Wichtigste Vorteile einer Nutzung von Big Data-Technologien (n=167, BARC 2013)

Betrachtet man die Potenziale auf einer **informationslogistischen Ebene**, so werden insbesondere bessere Information über Informations- und Konsumverhalten der Kunden, optimierte Logistikprozesse, bessere Einschätzung von Marktpotentialen und Business Cases sowie die bessere Aussteuerung von Vertriebs- und Marketingkampagnen genannt.

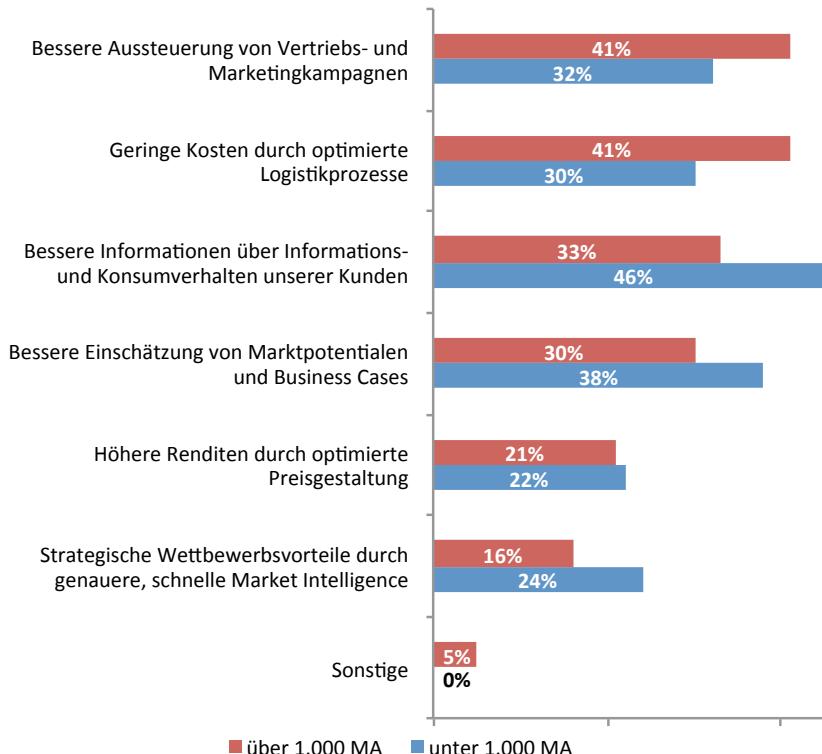


Abbildung 7: Positive Auswirkungen von Big Data (n=100, Experton Group 2012)

2.1.4 Konkreter Handlungsbedarf wird aktuell vor allem in der Informationslogistik gesehen

Der verantwortungsvolle Umgang mit dem Wirtschaftsgut Daten ist wichtig. In Unternehmen wird Handlungsbedarf bei Big Data dann gesehen, wenn es um die Steigerung der Effizienz und Effektivität der Informationslogistik geht.

Die deutsche Bevölkerung erkennt die Bedeutung von Daten als Wirtschaftsgut (Abbildung 8). Dabei sollte nach Meinung der Befragten der Datenschutz nicht außer Acht gelassen werden. Dem gegenüber sehen 25 Prozent der Befragten keinen Mehrwert in der Verarbeitung großer Datenmengen.

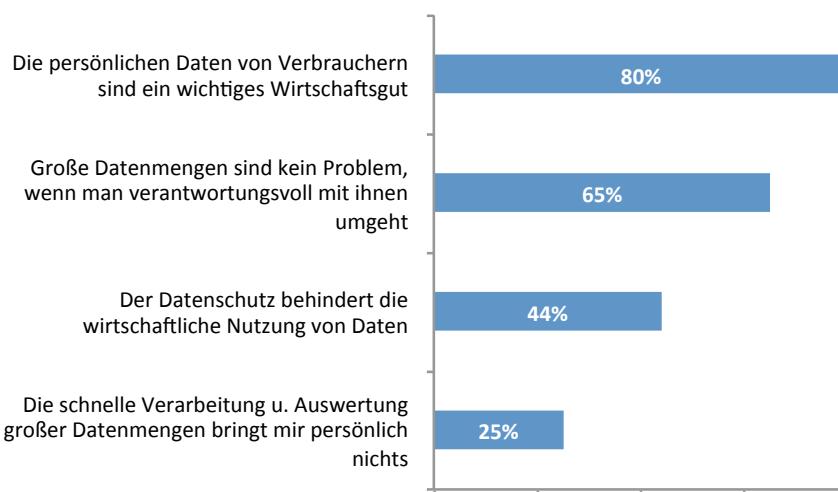


Abbildung 8: Zustimmung zu verschiedenen Aussagen über die Bedeutung von Big Daten in Prozent (n=1008, BITKOM 2013)

Diese Einschätzung teilen europäische IT-Entscheider und sehen eine hohe kurz- bis mittelfristige Bedeutung von Big Data in ihren Unternehmen (Abbildung 9). Besonders in Deutschland ist die Priorität von Big Data hoch einzuschätzen: 85 Prozent der Befragten schreiben Big Data eine hohe Priorität innerhalb der nächsten drei Jahre zu (Abbildung 10)

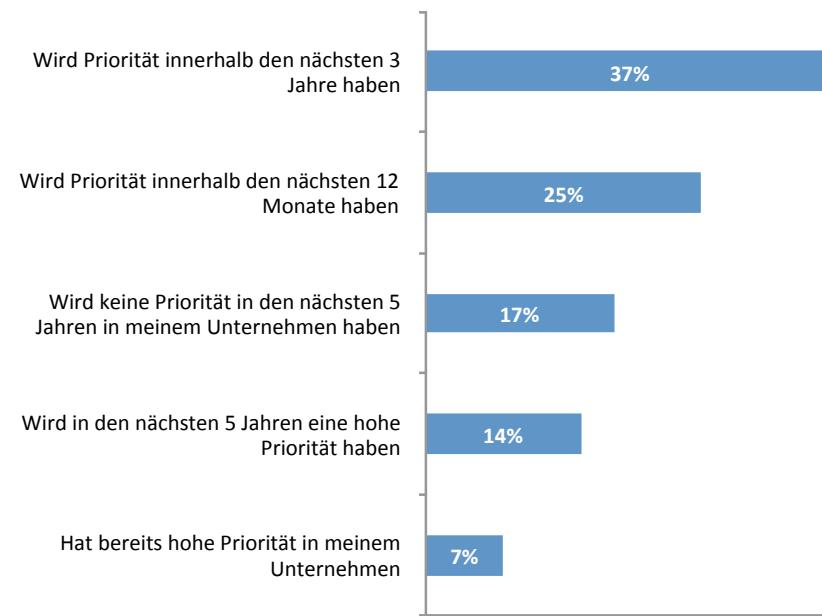


Abbildung 9: Zeithorizont und Priorität von Big Data in Unternehmen in Prozent (n=750, interxion 2013)

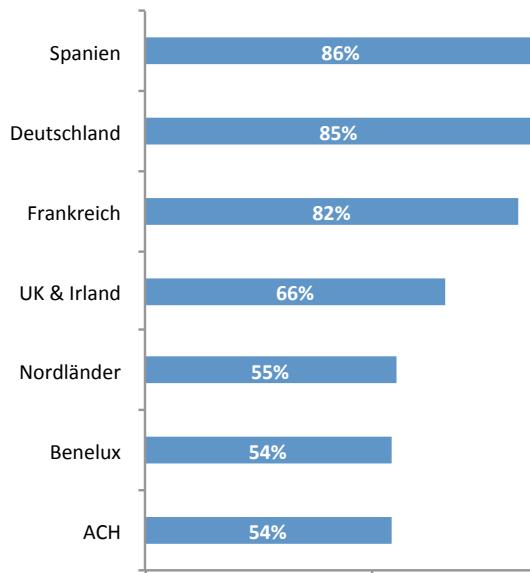


Abbildung 10: Priorität von Big Data innerhalb der nächsten drei Jahre in Prozent (n=750, interxion 2013)

Drei von neun Studien schlüsseln die Bedeutung von Big Data nach Branchen auf. Dabei legen alle drei Studien unterschiedliche Bewertungskriterien an. Während die Studie des BITKOM das Transformationspotenzial von Big Data in den einzelnen Branchen verwendet, wird die Bedeutung von Big Data in den Studien von Capgemini bzw. TNS anhand des Handlungsbedarfs in Bezug auf „Big Data Kultur“ bewertet. Zudem schlüsseln die Studien die Bedeutung von Big Data nach einer unterschiedlichen Anzahl an Branchen auf. Eine Zusammenfassung der einzelnen Studienergebnisse nach Branchen zeigt folgendes Bild:

- Die Bedeutung von Big Data in den Branchen Mobilität und Logistik wird als sehr hoch eingestuft. Dabei ist jedoch zu beachten, dass nur eine der drei Studien diese Branche betrachtet.
- Auf dem zweiten Rang befinden sich die Branchen Gesundheitswesen und Telekommunikation. Die Bedeutung von Big Data wird hier jeweils in zwei von drei Studien als sehr hoch bzw. hoch eingeschätzt.
- Auf dem dritten Rang befinden sich die Branchen Dienstleistungen und Öffentlicher Sektor. Diese Branchen werden in jeweils zwei von drei Studien betrachtet. Während die Bedeutung von Big Data in der Branche Dienstleistungen in beiden Studien als hoch bzw. mittel eingestuft wird, wird die Bedeutung von Big Data im Öffentlichen Sektor einmal als sehr hoch und einmal als niedrig eingestuft.
- Den vierten Rang belegen die Branchen Einzelhandel, Energie und Industrie. Die Bedeutung von Big Data in diesen Branchen wird in allen drei

Studien untersucht. Dabei wird die Bedeutung von Big Data jeweils in zwei der drei Studien als mittel bzw. niedrig eingeschätzt.

- Den fünften und letzten Rang belegt die Branche Finanzdienstleistungen. Auch hier wird die Bedeutung von Big Data in allen drei Studien untersucht. In zwei von drei Studien wird die Bedeutung von Big Data als sehr niedrig eingeschätzt.

Insgesamt lässt sich der Überblick in Tabelle 3 in zwei Gruppen einordnen. Die Gruppe welche ein hohes Transformationspotential und einen hohen Nachholbedarf hat. Der Wettbewerbsvorteil ist hier noch auszubauen. Die zweite Gruppe besteht aus Unternehmen aus den Branchen Einzelhandel, Energie, Industrie und Finanzdienstleistungen. Die Unternehmen dieser Branchen sind sehr wettbewerbsfähig geben aber an Big Data Technologie weiter ausbauen (Finanzdienstleistungen und Versicherungen) oder etablieren (Energie) zu wollen.

	GESAMTRANG ⁰	BITKOM ¹	Capgemini ²	TNS ³
Mobilität und Logistik	1	1	-	-
Gesundheitswesen	2	1	2	-
Telekommunikation	2	1	-	2
Dienstleistungen	3	2	-	3
Öffentlicher Sektor	3	1	-	4
Einzelhandel	4	1	3	4
Energie	4	3	4	1
Industrie	4	3	1	4
Finanzdienstleistungen und Versicherungen	5	2	5	5

O: Basierend auf dem Durchschnittsrang über die verfügbaren Studienergebnisse
 1: Rangfolgekriterium: Transformationspotenzial durch Herausgeber geschätzt
 2: Rangfolgekriterium: Nachholbedarf in Bezug auf „Big Data Kultur“, Sample: 607 allgemeine Entscheidungsträger
 3: Rangfolgekriterium: Handlungsbedarf, Sample: 1010 IT-Entscheidungsträger

Tabelle 3: Aggregation verschiedener Studien zur Bedeutung von Big Data nach Branchen (Eigene Darstellung)

Exemplarisch soll hier auf die Ergebnisse der Studie der TNS Infratest eingegangen werden. Die Ergebnisse zeigen, dass der Handlungsbedarf bei Big Data über Branchen hinweg schwankt aber durchgehend im höheren Bereich anzusiedeln ist (Abbildung 11). Insbesondere in den Branchen Energie, Telekommunikation und Dienstleistungen wird ein hoher Handlungsbedarf ausgemacht.

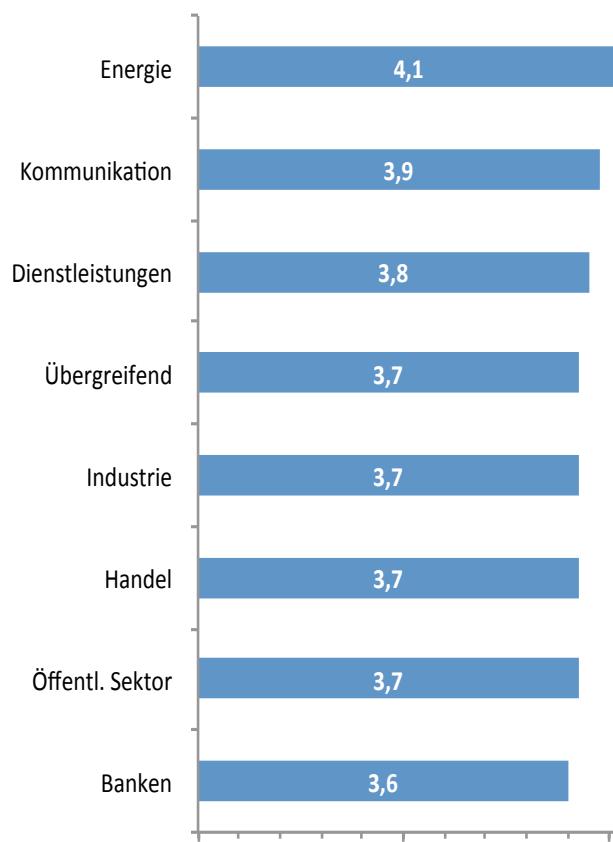


Abbildung 11: Handlungsbedarf bei Big Data nach Branche auf einer Skala von 1 bis 5
(n=1010, TNS Infratest 2012)

Ein ähnliches Bild ergibt sich, wenn man Entscheider nach den zu erwartenden Renditen von Big Data Projekten in einzelnen Branchen fragt. Auch hier finden sich informationsintensive Branchen auf den Spitzensplätzen. Lediglich der Wert für die Kommunikationsbranche fällt hier geringer aus. Branchen mit hohen erwarteten Renditen sind Versorgung, Energie und Rohstoffe sowie die High-Tech-Branche. Hier liegen die erwarteten Renditen bei 73,0 Prozent, 60,6 Prozent und 52,4 Prozent.

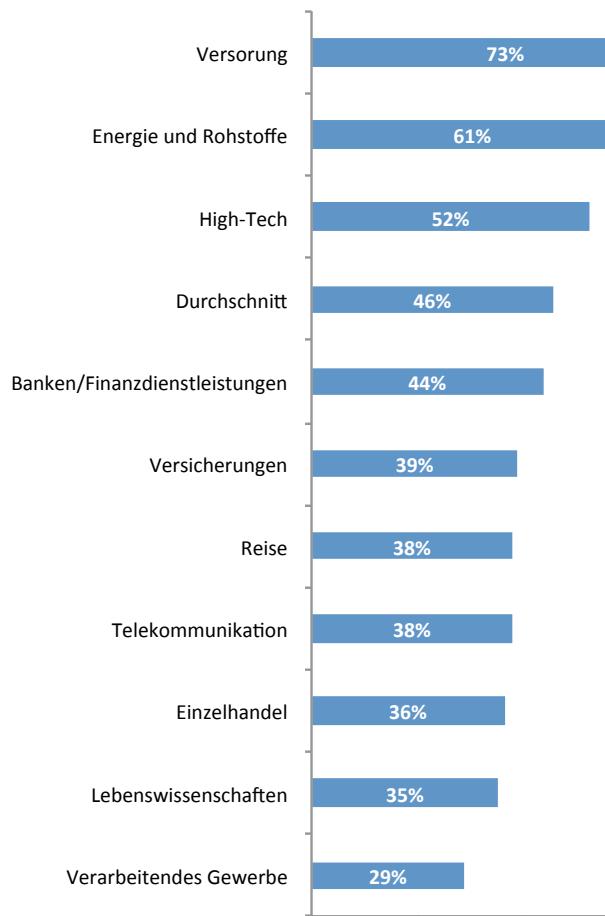


Abbildung 12: Durchschnittliche erwartete Rendite von Big Data Investition nach Branchen im Jahre 2012 in Prozent (n=643, TCS 2013)

Analysiert man die Bedeutung von Big Data für einzelne Unternehmensaktivitäten, wird deutlich, dass die größten Potenziale in sekundären Unternehmensfunktionen gesehen werden. Das höchste Potenzial wird bei der Vertriebsanalyse (55 Prozent Zustimmung), der Wettbewerbsanalyse, der Budgetierung sowie der Mitarbeiterauslastung gesehen (alle 49 Prozent Zustimmung). Dies sind alles Unternehmensaktivitäten bei denen die schnelle Analyse großer Datenmengen zu Effizienzgewinnen in der Informationslogistik führen kann.

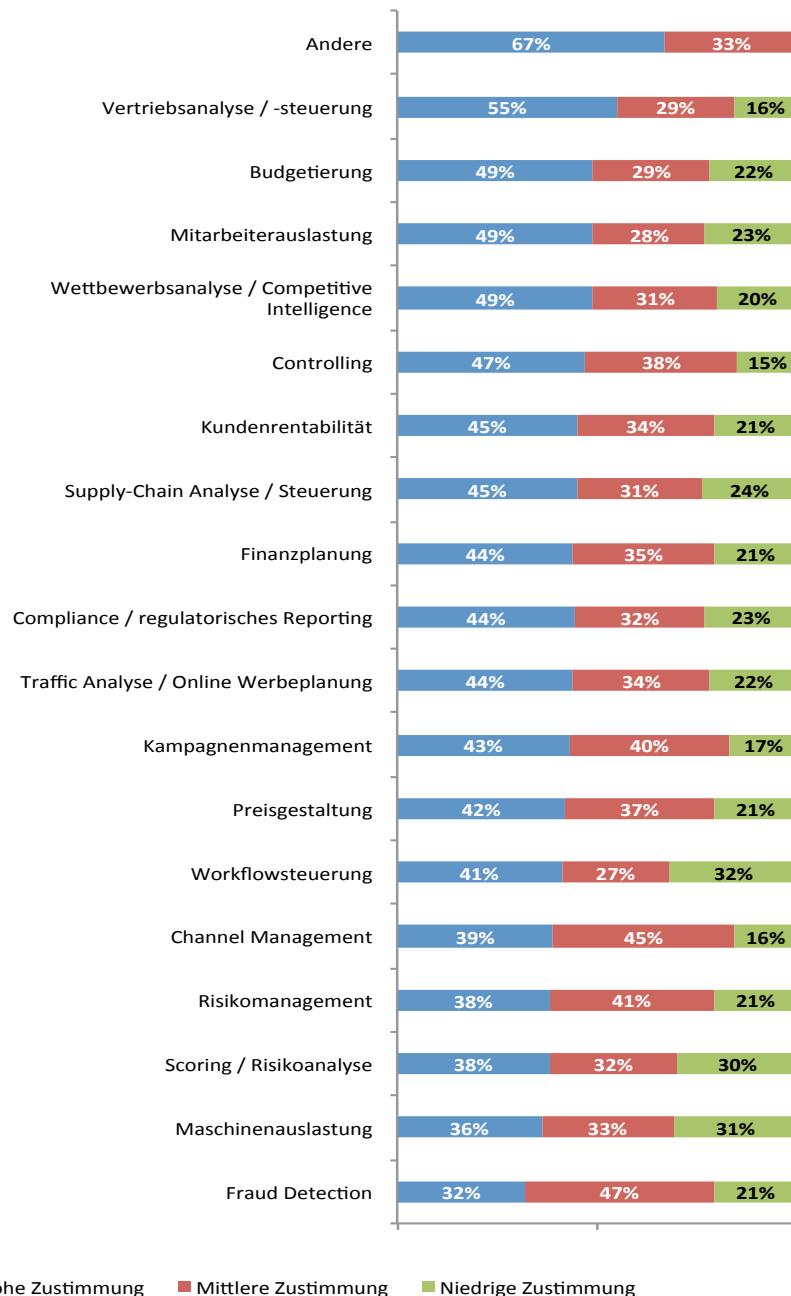


Abbildung 13: Unternehmensaktivitäten und deren Potenzial (n=85, Experton Group 2012)

Durch weitere Studien wird die Unternehmensaktivität bestätigt. Neben Vertriebs- und Marketingaufgaben wird hier einem Einsatz von Big Data in der Logistik sowie dem Qualitätsmanagement ein hohes Potenzial zugewiesen (Tabelle 4).

Rang	Aktivität	Funktion	Potenzial (auf ei- ner Ska- la von 1-5)
1	Identifikation der werthaltigsten Kunden	Vertrieb	4,05
2	Überwachung der Produktqualität	F&E	4,02
3	Überwachung von Produktlieferungen	Logistik	4,00
4	Identifikation von Kundenbedürfnissen	F&E	3,95
5	Identifikation von Risikokunden	Kunden- dienst	3,94
6	Bestimmung der Effektivität von Marketingkampagnen	Marketing	3,90
6	Bestimmung von Lagerverknappungen	Logistik	3,90
6	Identifikation von Ausschlägen in Logistikkosten	Logistik	3,90
7	Analyse von Kundenverhalten auf der Unternehmenswebsite	Kunden- dienst	3,89
8	Verbesserung des Mitarbeiterverbleibs	Personal	3,85
9	Identifikation von Cross-Selling-Möglichkeiten	Vertrieb	3,80
9	Bestimmung der Effektivität von Marketingkanälen	Marketing	3,80
10	Bestimmung der optimalen Vertriebsansätze	Vertrieb	3,78
10	Testen neuer Produktdesigns	F&E	3,78

Tabelle 4: Top 10 Unternehmensaktivitäten und deren Potenzial (n=643, TCS 2013)

Werden Entscheider nach erwarteten Renditen für Big Data Projekte gefragt, wird die aktuelle Bedeutung der Informationslogistik deutlich. So sind diese in den Funktionsbereichen Logistik und Finanzen mit Abstand am höchsten. Hier wird mit Renditen von 78,1 Prozent bzw. 69,0 Prozent gerechnet. Überraschend ist die mit 41,4 Prozent relativ niedrige erwartete Rendite im Funktionsbereich Marketing.

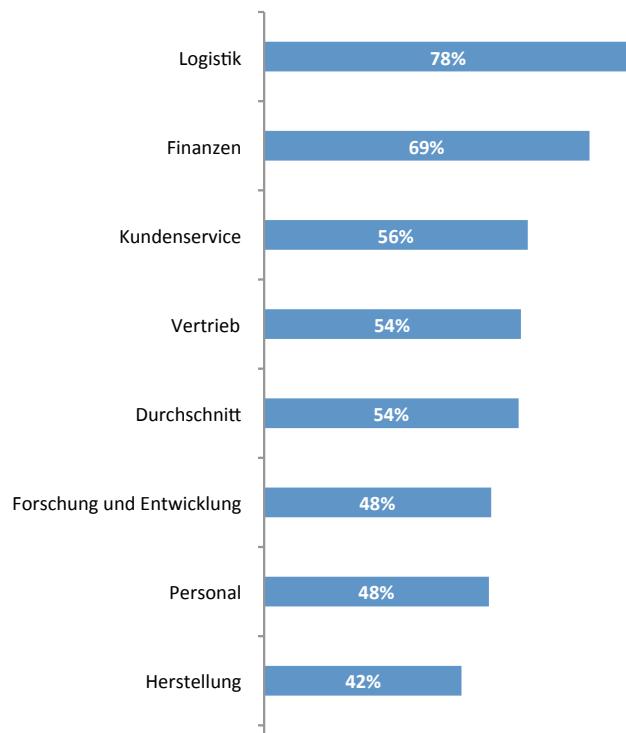


Abbildung 14: Durchschnittliche erwartete Rendite von Big Data Investition nach Funktionsbereichen im Jahre 2012 in Prozent (n=643, TCS 2013)

2.1.5 Unternehmen beginnen erst damit, Strategien für den Einsatz von Big Data über die Analyse strukturierter Daten hinaus zu entwickeln

Die Mehrheit der befragten Unternehmen explorieren aktuelle Möglichkeiten für den Einsatz von Big Data. Wenn Projekte gestartet wurden, überwiegen Projekte mit strukturierten Daten.

Werden Big Data Projekte durchgeführt, dominieren derzeit vor allen Dingen die Analyse strukturierter Daten, wie Transaktionsdaten (70 Prozent) und Logdaten (50 Prozent). Immerhin 40 Prozent der befragten Unternehmen ziehen schon unstrukturierte Daten für Analysen heran. Die Analyse von Daten aus sozialen Netzwerken werden mit 14 Prozent noch sehr wenig genutzt. Immerhin planen jedoch 50 Prozent die zukünftige Nutzung von Social Media-Daten. In den Branchen Energie und Industrie kommt dies bereits vermehrt vor, denn dort ist ein hoher Anteil an Transaktionsdaten, Logs und Sensorik im Einsatz. Es zeigt sich in diesen Branchen ebenfalls ein Aufholbedarf.

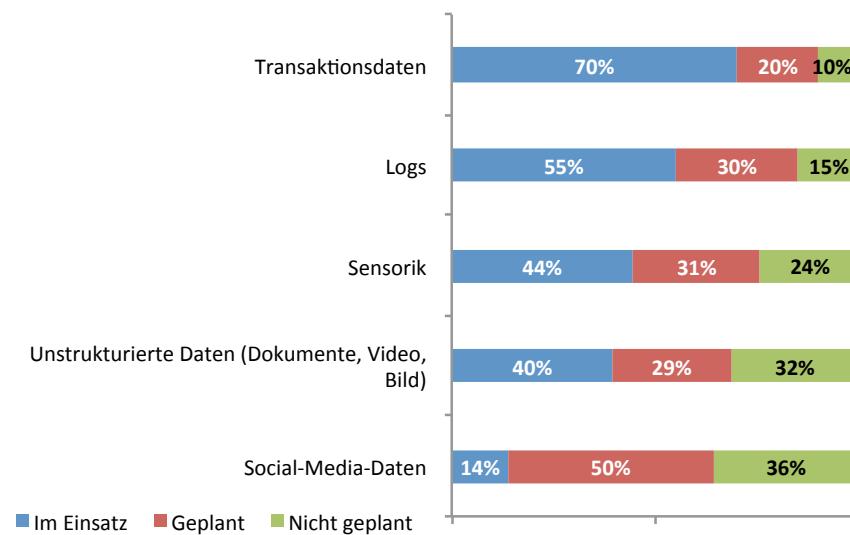


Abbildung 15: Einsatz verschiedener Datenarten für Big Data Analysen (n=274, BARC 2013)

Über alle Studien hinweg zeigt sich, dass sich ein Großteil der befragten Unternehmen bereits mit Big Data beschäftigt hat, die Big Data Initiativen jedoch noch am Anfang stehen. So haben beispielsweise 63 Prozent der befragten Entscheider angegeben, dass es in ihrem Unternehmen noch keine Strategie für den Einsatz von Big Data gibt (Abbildung 16).

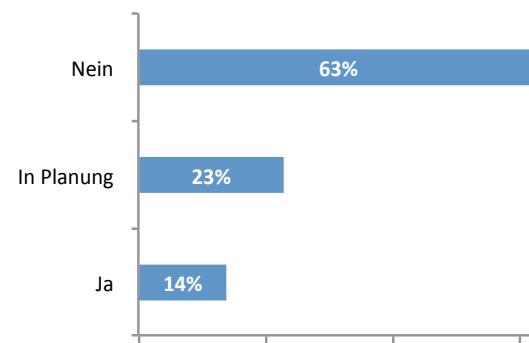


Abbildung 16: Vorhandensein einer Strategie für Big Data (n=274, BARC 2013)

Auch zeigen andere Studien, dass lediglich 25 Prozent der Unternehmen den Einsatz von Big Data komplett ausschlossen. Der Großteil der befragten Unternehmen plante eine Beschäftigung oder befand sich bereits in ersten Planungs- und Umsetzungsphasen (Abbildung 17).

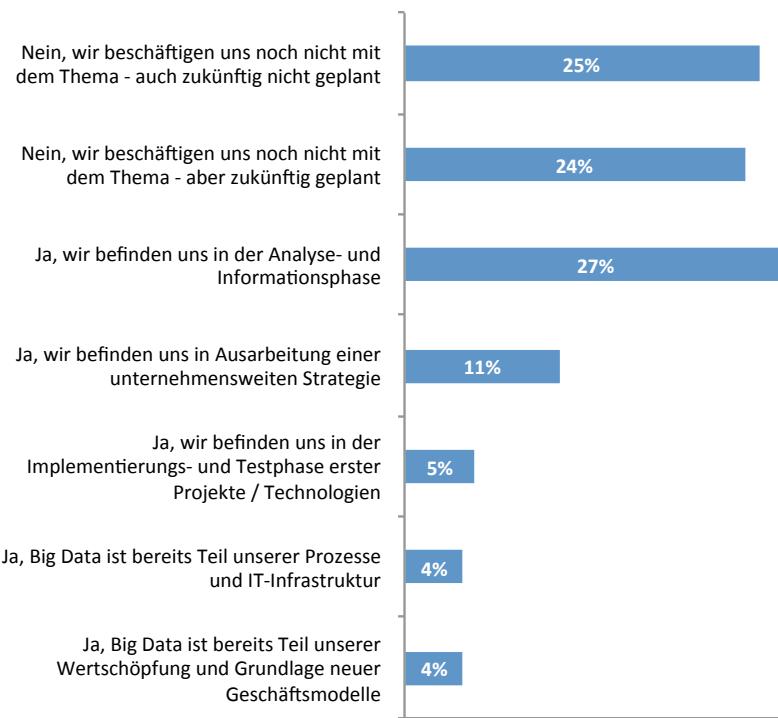


Abbildung 17: Beschäftigung mit Big Data (n=100, Experton Group 2012)

Darüber hinaus wird deutlich, dass die Beschäftigung mit Big Data von 2012 auf 2013 deutlich angestiegen ist (Abbildung 16). Während im Mai 2012 noch 61 Prozent der Unternehmen kein Interesse an Big Data zeigten, waren es im März 2013 nur noch 24 Prozent. Zudem sind viel mehr Unternehmen in der Diskussionsphase (2012: 24 Prozent, 2013: 36 Prozent) sowie in der Planungsphase (2012: 8 Prozent, 2013: 19 Prozent). Auch sind 2013 bereits 10 Prozent im Roll-out von Big Data Anwendungen.

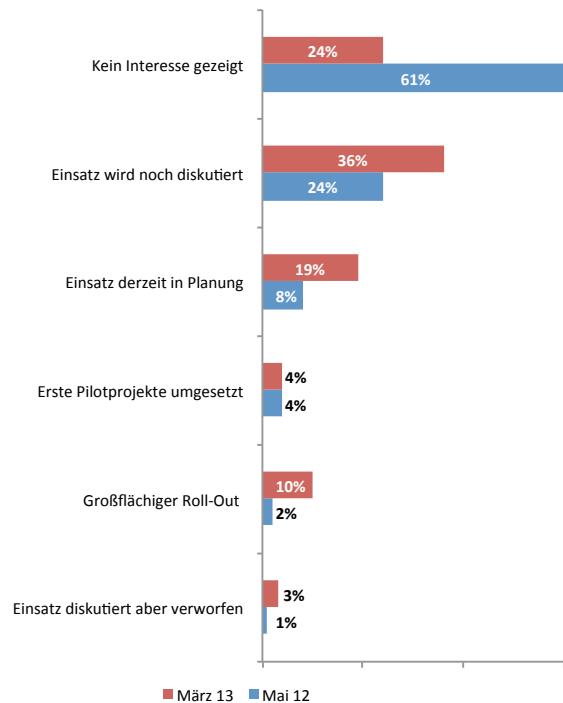


Abbildung 18: Veränderung des Ausmaßes der Beschäftigung mit Big Data 2012 bis 2013
(n=n.a., computing research 2013)

Der Einsatz von Big Data-Werkzeugen variiert geringfügig zwischen Unternehmen unterschiedlicher Größe (siehe Abbildung 19). Unternehmen mit unter 250 Mitarbeitern sowie Unternehmen mit über 2.500 Mitarbeitern setzen zu über 45 Prozent Big Data-Werkzeuge ein (unter 250 Mitarbeiter: 45 Prozent; über 2.500 Mitarbeiter: 49 Prozent). Unternehmen mit 250 bis 2.499 Mitarbeitern liegen mit einem Einsatz von 32 Prozent etwas darunter, planen aber mit einem zukünftigen Einsatz von 41 Prozent.

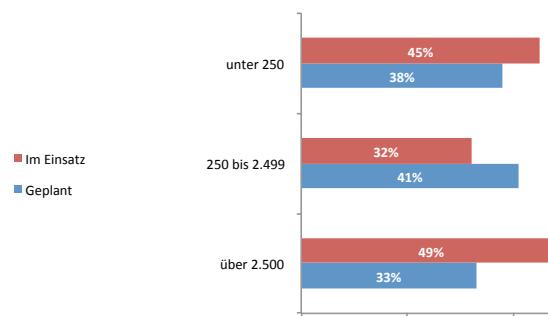


Abbildung 19: Nutzung von Big Data-Werkzeugen nach Unternehmensgröße (n=132, BARC 2013)

2.1.6 Für den Einsatz von Big Data fehlen Fachkräfte sowie organisatorische Strukturen und Prozesse in den Unternehmen

Datensilos, Fachkräftemangel und fehlende Prozesse hemmen den Einsatz von Big Data. Die Außenwirkung von Big Data, insbesondere der Datenschutz, spielt daher noch eine nachgelagerte Rolle.

Acht der neun ausgewählten Studien gehen auf Herausforderungen bei der Umsetzung von Big Data ein. Es lassen sich acht Kategorien zusammenfassen. Aus der Aggregation der Ergebnisse der Studien in Hinblick auf die Bedeutung dieser Herausforderungskategorien (vgl. Tabelle 5) ergeben sich folgende Erkenntnisse:

- Über alle Studien hinweg werden die hohen Kosten von Big Data bzw. die mangelnde Investitionsbereitschaft des Managements als eine der beiden größten Herausforderungen gesehen.
- Eng damit verbunden ist die niedrige Priorität des Themas Big Data im Unternehmen selbst.
- Entgegen der weit verbreiteten Annahme sind Datenschutz und Datensicherheit bei Weitem nicht die größten Herausforderungen. Allerdings bewerten die Studien hier sehr unterschiedlich: Während beispielsweise die Studie des Fraunhofer IAIS Datenschutz als bedeutendste Herausforderung sieht, wird Datenschutz in der Studie der TCS als weit weniger bedeutend angesehen. Dies kann so interpretiert werden, dass auf unterschiedlichen Unternehmensebenen unterschiedliche Schwerpunkte gesetzt werden. Aus unternehmerischer Sicht ist die Außenwirkung von Big Data-Projekten wichtig, aus prozessorientierter bzw. informationslogistischer Sicht überwiegen technische und organisatorische Herausforderungen.
- Für die Aggregation wurde der Durchschnittsrang über die verfügbaren Studienergebnisse ermittelt und mittels eines Gesamtrangs gereiht.

	GESAMTRANG ⁰									
	BARC	BITKOM	Capgemini	Computing Research	Experton Group	Fraunhofer IAIS	interxion	Tata Consultancy Services	TNS	
Hohe Kosten und/ oder mangelnde Investitionsbereitschaft des Managements	1	5	n.a.	6	1	5	2	2	6	3
Niedrige Priorität des Themas Big Data im Unternehmen	1	-		5	2	7	-	1	-	-
Technische Hürden	2	4		5	6	3	6	4	2	3

Unterstützung der Fachseite	3	5	1	-	-	-	-	7	4
Fehlendes Know-how	4	2	2	3	4	4	3	10	-
Schwierige Interpretation / Weiterverwertung der Daten und Analysen	5	-	4	-	-	-	-	7	-
Datenschutz und Datensicherheit	6	6	-	5	3	1	-	15	-
Management des Organisatorischen Wandels	7	-	-	6	-	-	-	9	-
O: Basierend auf dem Durchschnittsrang über die verfügbaren Studienergebnisse									

Tabelle 5: Aggregation verschiedener Studien zu den Herausforderungen für Big Data (Eigene Darstellung)

Im Folgenden wird exemplarisch auf die Ergebnisse der Studien des BARC und von Computing Research eingegangen. Hier beeinträchtigen vor allem das fehlende technische und das fehlende fachliche Know-how die Umsetzung (Abbildung 20). Im Detail stellen vor allem die Integration von organisatorischen Datensilos, die Handhabung großer, heterogener und schnell veränderlicher Datenbestände und die Bestimmung der richtigen Daten für betriebswirtschaftliche Entscheidungen Herausforderungen dar. Ebenso fehlen überzeugende Einsatzszenarien.

Abbildung 20: Herausforderungen beim Einsatz von Big Data (n=206, BARC 2013)

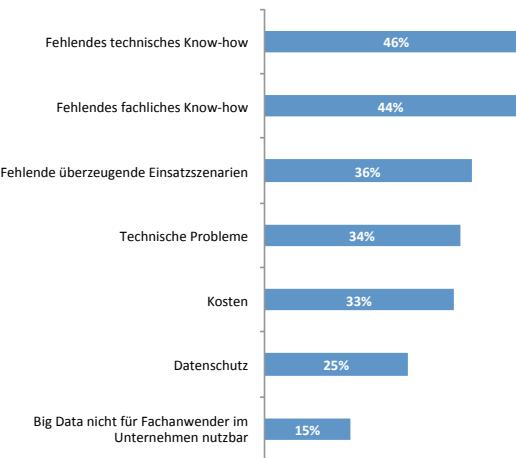


Tabelle 6 bestätigt vor allem das fehlende technische und das fehlende fachliche Know-how für die Umsetzung von Big Data-Projekten. Ebenso stellen vor allem die Integration von organisatorischen Datensilos, die Handhabung großer, heterogener und schnell veränderlicher Datenbestände und die Bestimmung der richtigen Daten für betriebswirtschaftliche Entscheidungen Herausforderungen dar. Dazu wurden zwei Studien, die von TCS und Capgemini herangezogen, welche hierzu Auskunft geben.

Herausforderung	Rang TCS	Rang Capgemini
Integration organisatorischer Datensilos	1	1
Handhabung großer, heterogener, schnell veränderlicher Datenbestände	2	5
Bestimmung der richtigen Art an Daten für betriebswirtschaftliche Entscheidungen	3	3
Bildung von Vertrauen zwischen Datenwissenschaftlern und Linienmanagern	4	-
Auffinden und Rekrutierung von Datenwissenschaftlern	5	2
Sicherung von Topmanagementunterstützung zur Finanzierung von Big Data Initiativen	6	4
Visualisierung und Präsentation der Ergebnisse von Big Data Analysen	7	-
Optimale Organisation der Big Data Initiativen innerhalb des Unternehmens	8	-
Auffinden von optimalen Unternehmensbereiche für Big Data Initiativen	9	-
Umsetzung der Erkenntnisse von Big Data Analysen in Handlungsalternativen	10	-
Umschulung der IT-Mitarbeiter zur Benutzung von Big Data Tools	11	-
Überzeugung des Linienmanagements Entscheidungen auf Big Data Erkenntnisse zu basieren	12	-
Entscheidung über richtige Big Data Tools	13	-
Sicherung der Datenbestände gegenüber externen Parteien	14	-
Überzeugung der IT-Mitarbeiter neuer Technologien und Fertigkeiten zu erlernen	15	-
Sicherung der Datenbestände gegenüber externen Parteien	16	-

Tabelle 6: Herausforderungen bei der Umsetzung von Big Data Initiativen (TCS 2013/Capgemini 2012)

Auch auf technischer Ebene werden zahlreiche Herausforderungen genannt, die bei der Umsetzung von Big Data Initiativen auftreten können. Hier sind vor allem die Vielzahl an verschiedenen Datenstrukturen, die Verknüpfung von

strukturierten und unstrukturierten Daten, sowie die Auswahl der zu speichern den Daten für Unternehmen problematisch.

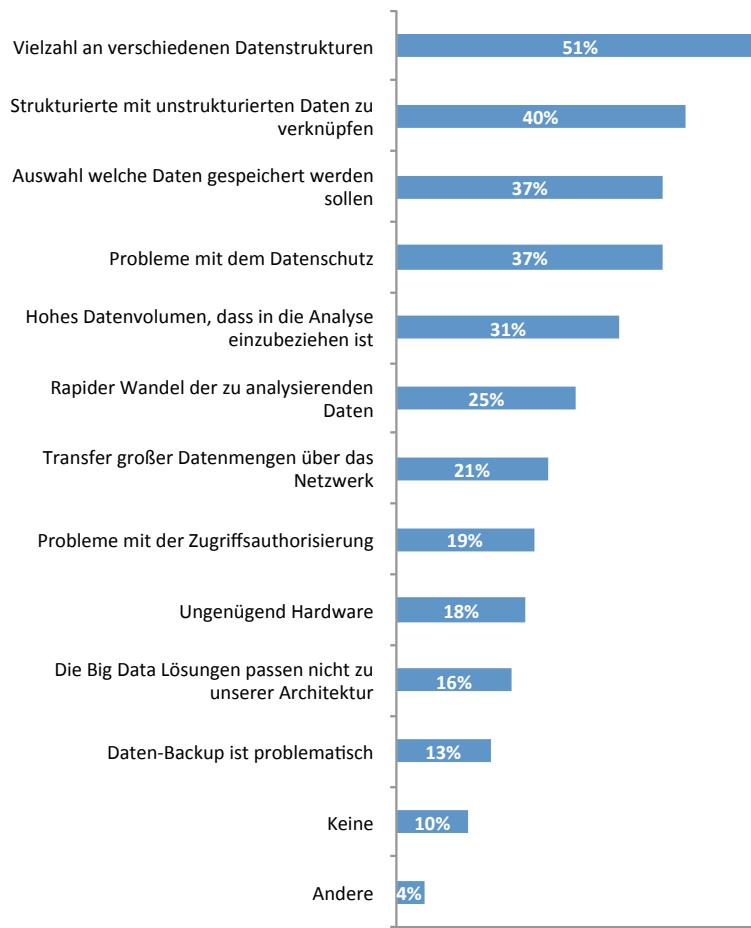


Abbildung 21: Technische Hindernisse bei der Umsetzung einer Big Data Strategie in Prozent (n=n.a., Computing Research)

Klar ersichtlich ist aber über alle Studien hinweg, dass sobald die technischen und organisatorischen Rahmenbedingungen für Big Data Projekte geschaffen wurden, der Datenschutz die zentrale Herausforderung ist.

2.1.7 Mit zunehmender Bereitschaft der Unternehmen wächst der Big Data Markt erheblich

Steigende Ausgaben und ein entsprechendes Marktwachstum werden durch alle Studien prognostiziert.

Prognosen zur Entwicklung des Big Data-Marktes zeigen, dass mit einem rasanten Wachstum gerechnet wird. Hatte der globale Big Data Markt im Jahr 2011 noch ein Volumen von 3,377 Mrd. € wird er auf über 15 Mrd. € im Jahr 2016 prognostiziert (Abbildung 22).

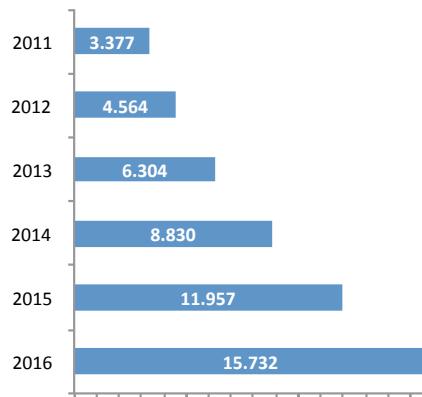


Abbildung 22: Entwicklung des globalen Big Data Marktes 2011-2016 in Mio. € (Experton Group 2012)

Eine ähnliche Entwicklung wird auch für den deutschen Big Data Markt prognostiziert (Abbildung 23). Ausgehend von einem deutschen Marktvolumen von knapp 200 Mio. € im Jahr 2011, wird auf Basis eines kontinuierlichen Marktwachstums in allen Segmenten Services, Software und Hardware das deutsche Big Data Marktvolumen im Jahr 2016 auf über 1,6 Mrd. € prognostiziert.

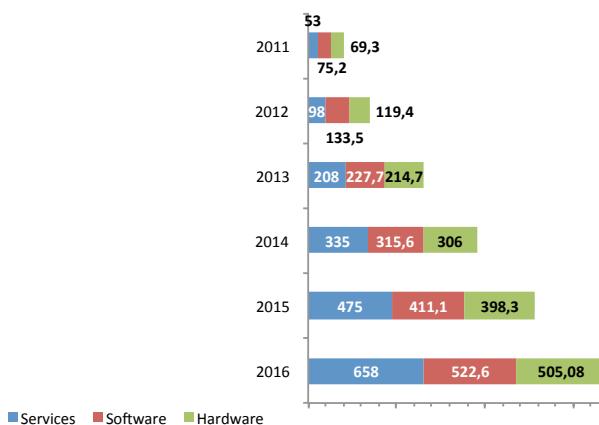


Abbildung 23: Deutscher Big Data Markt 2011-2016 nach Marktsegmenten in Mio. € (Experton Group 2012)

Ähnliche Trends lassen sich auch auf der Ausgabenseite bei den Unternehmen feststellen. Australische Unternehmen waren 2012 mit großem Abstand Spitzenreiter was die Ausgaben für Big Data Initiativen angeht. Australische

Unternehmen gaben 2012 im Durchschnitt 50 Millionen USD für Big Data Initiativen aus. Platz zwei und drei nahmen Japan und die Niederlande mit durchschnittlich 15,5 und 15 Millionen USD ein. In nahezu allen Ländern wird mit einer Erhöhung der Big Data Ausgaben gerechnet. Die Ausnahme bildet die Spitzengruppe Australien, Japan und die Niederlande. Hier wird erwartet, dass die durchschnittlichen Ausgaben bei 50 bzw. 15 Millionen USD konstant bleiben, im Falle der Niederlande sogar auf 12,5 Millionen USD sinken. Deutschland liegt hier im Mittelfeld, es werden aber erhebliche Zuwächse prognostiziert.

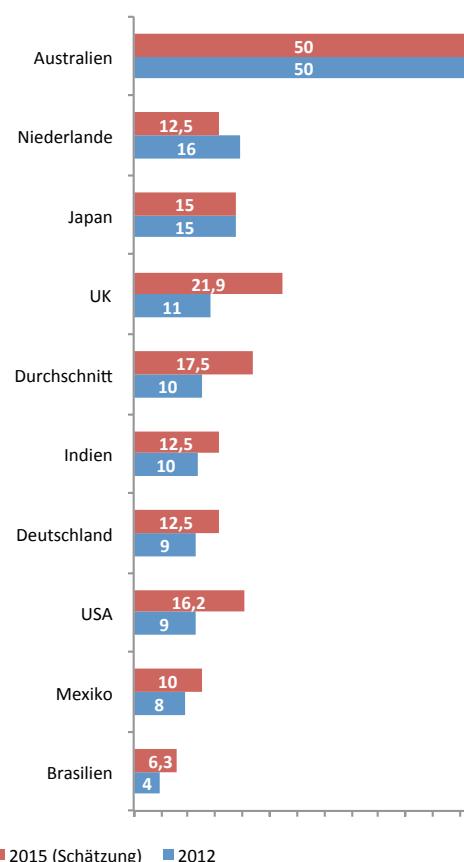


Abbildung 24: Durchschnittliche Ausgaben eines Unternehmens für Big Data nach Ländern in den Jahren 2012 und 2015 (prognostiziert) in Mio. USD (n=643, TCS 2013)

Betrachtet man die Budgetentwicklung auf europäischer Ebene für Unterschiedliche Unternehmensgrößen (Tabelle 7) ergibt sich ein gemischtes Bild: Unternehmen mit unter 250 Mitarbeitern sowie Unternehmen mit über 2.500 Mitarbeitern konnten ihr Budget für die drei Bereiche Hardware, Software und Beratung jeweils um über 7 Prozent steigern; Unternehmen mit 250 bis 2.499 Mitarbeitern wiesen hingegen eine deutlich geringere Budgetentwicklung (< 5 Prozent) für in den Bereichen Hardware, Software und Beratung aus.

Anzahl Mitarbeiter	Ausgaben für	Median	Mittelwert	Entwicklungs-Budget
Unter 250	Hardware	72.414	78.333	7,7 Prozent
	Software	89.815	122.581	9,5 Prozent
	Beratung	71.296	94.355	7,2 Prozent
250 bis 2.499	Hardware	116.667	210.526	4,6 Prozent
	Software	126.852	228.289	4,3 Prozent
	Beratung	98.276	208.784	4,8 Prozent
Über 2500	Hardware	295.588	437.500	8,6 Prozent
	Software	310.714	460.547	10,4 Prozent
	Beratung	201.136	366.393	7,7 Prozent

Tabelle 7: Durchschnittliche Ausgaben und Budgetentwicklung für Big Data nach Unternehmensgröße (n=unterschiedliche Basis, BARC 2013)

2.1.8 Zusammenfassung

Aus der integrierten Analyse der Studien lassen sich folgende Punkte für ein Zwischenfazit ableiten:

- Big Data schafft dann einen hohen Wert, wenn es zur Schaffung neuer Wettbewerbsvorteile genutzt werden kann. Big Data ist aber vorerst noch ein Thema unter IT-Experten. Dies wird daran deutlich, dass vor allem technische Herausforderungen genannt werden und erste Big Data-Projekte vor allem auf der Ebene der Informationslogistik stattfinden.
- Es fehlt noch die wirtschaftliche Rationalisierung für das Zusammenführen der verschiedenen Datenquellen in und außerhalb von Unternehmen. Einzelne Maßnahmen rechtfertigen oftmals nicht die Aufwände, die für die Aufbereitung und Analyse der Daten notwendig sind.
- Es fehlt die breite Basis an Referenzbeispielen, die zeigen, wie mit Big Data auf unternehmerischer Ebene neue Geschäftsmodelle, Produkte und Dienstleistungen geschaffen werden können. Bisher wird Big Data vor allem aus Effizienzgründen untersucht, das heißt, es sollen bestehende Prozesse schneller bzw. mit höherer Qualität durchgeführt werden.

2.2 Erhebung zu Innovationspotenzialen

Ergänzend zu obiger Analyse der existierenden Studien zu Innovationspotenzialen von Big Data wird eine eigene empirische Untersuchung durchgeführt. Das Ziel dieser Untersuchung ist es, die bestehenden Erkenntnisse zu validieren und zu ergänzen, so dass eine differenzierte Betrachtung hinsichtlich der Adoption von Big Data möglich ist.

Es können folgende Kernaussagen aus der folgenden Untersuchung abgeleitet werden:

1. Big Data wird vor dem Hintergrund effektiverer Unternehmensentscheidungen gesehen. Allerdings stehen Anbieter und Anwender vor der Herausforderung, überzeugende Kombinationen aus Datenanalyse, unternehmerischer Entscheidung und Wertbeitrag zu entwickeln. Eine Chance bieten hier unternehmensinterne Prozesse, da eine Vielzahl von Daten bisher nicht genutzt wird.
2. Deutsche Anbieter von Big Data Technologien sind in der Lage, die Anforderungen deutscher Anwender zu erfüllen. Eine zentrale Herausforderung liegt aber in der Aufklärung und Erläuterung der datenschutzrechtlichen Rahmenbedingungen sowie in der Entwicklung tragfähiger und wirtschaftlicher Lösungen.
3. Die Potenziale von Big Data werden zudem im Bereich neuer Geschäftsmodelle, Produkte und Dienstleistungen gesehen. Anwenderunternehmen sind sich im Unklaren, welche Datenanalysen für ihre jeweiligen Geschäftsprozesse relevant und wertschöpfend sind. Hier könnten Anbieter und Anwender intensiv kooperieren, um entsprechende Geschäftsmodelle, Produkte und Dienstleistungen zu entwickeln.
4. Eine wichtige Herausforderung ist der verantwortungsvolle Umgang mit personenbezogenen Daten. Hier sehen alle Beteiligten eine sehr hohe Herausforderungen. Dabei stehen neben einer Anpassung der rechtlichen Rahmenbedingungen an den Stand der Technik vor allem die Aufklärung und ein effektives Erwartungsmanagement im Fokus.
5. Eine weitere wichtige Herausforderung ist die Verfügbarkeit entsprechend qualifizierten Personals. Geeignete Ausbildungs- und Weiterbildungsangebote sind hier notwendig, um den Unternehmen das entsprechende Know-how bereitzustellen.
6. Es gibt kaum Unterschiede in der Einschätzung der Innovationspotenziale, wenn man dediziert kleine und mittelständische Unternehmen untersucht. Dies kann als Chance für den Mittelstand interpretiert werden, da es offenbar keine zusätzlichen Hürden für den Mittelstand gibt.

Im Folgenden wird zunächst die Methodik der Analyse beschrieben. Anschließend werden die Kernaussagen detailliert vorgestellt.

2.2.1 Methodik für die Durchführung der Umfrage

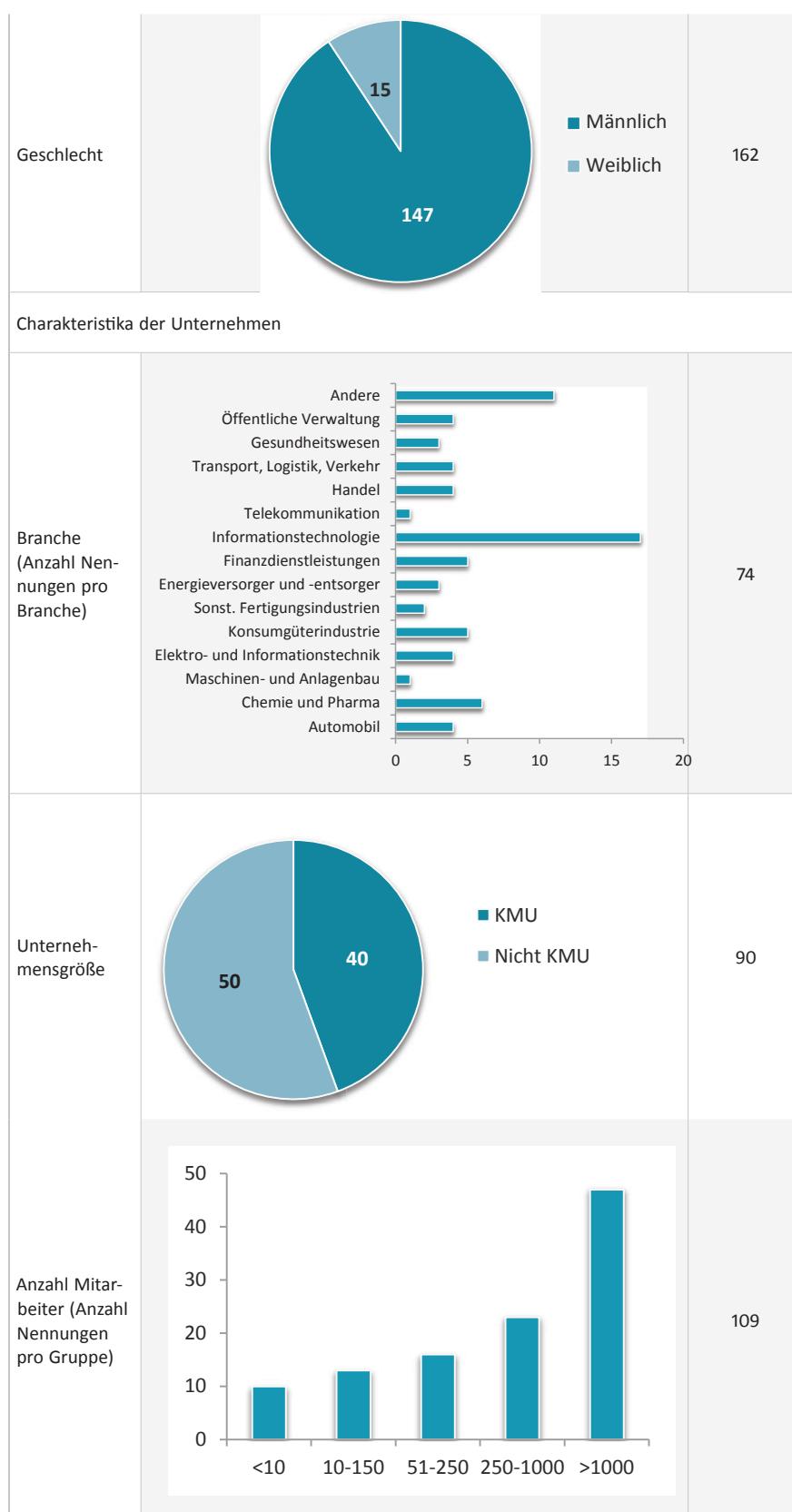
Die Umfrage basiert auf einer quantitativen Befragung von Entscheidern, Anbietern, Anwendern und Wissenschaftlern. Dazu wurde folgendes Vorgehen gewählt:

- Zunächst wurde ein Fragebogen erarbeitet und abgestimmt. Dieser Fragebogen wurde mehreren Tests unterzogen und entsprechend überarbeitet. Der Fragebogen wurde anschließend als Online-Fragebogen aufbereitet. Zur Motivation der Teilnahme wurden unter allen komplett ausgefüllten Fragebögen drei Amazon Kindle Paperwhite mit je einem 50 Euro Gutschein verlost.
- Der finale Fragebogen wurde mit Unterstützung von mehreren Verbänden (BITKOM, Gesellschaft für Informatik e.V., VOICE e.V.) an Teilnehmer verteilt. Ebenso wurde der Fragebogen in verschiedenen Online-Medien bekannt gemacht. Der Fragebogen wurde so entwickelt, dass nachvollziehbar bleibt, aus welcher Quelle die entsprechenden Antworten stammen.
- Die Auswertung der abgegebenen Antworten erfolgte mit Standardwerkzeugen der deskriptiven Statistik.

2.2.2 Hintergrund der Umfrageteilnehmer

Die folgende Tabelle beschreibt den persönlichen Hintergrund der Umfrageteilnehmer sowie Charakteristika der Unternehmen, in denen die Teilnehmer ange stellt sind.

Kriterium	Beschreibung	Anzahl Antworten																																								
Teilnehmer	Gesamte Anzahl an Fragebogenrückläufern	185																																								
Persönlicher Hintergrund der Teilnehmer	<p>Tätigkeits- bzw. Wissenshintergrund der Teilnehmer</p> <table border="1"> <thead> <tr> <th>Kategorie</th> <th>Anteil (%)</th> </tr> </thead> <tbody> <tr> <td>Anbieter</td> <td>53</td> </tr> <tr> <td>Anwender</td> <td>53</td> </tr> <tr> <td>Entscheider</td> <td>29</td> </tr> <tr> <td>Wissenschaftler</td> <td>50</td> </tr> </tbody> </table>	Kategorie	Anteil (%)	Anbieter	53	Anwender	53	Entscheider	29	Wissenschaftler	50	185																														
Kategorie	Anteil (%)																																									
Anbieter	53																																									
Anwender	53																																									
Entscheider	29																																									
Wissenschaftler	50																																									
Berufserfahrung (Anzahl Nennungen pro Jahr)	<table border="1"> <thead> <tr> <th>Jahr</th> <th>Anzahl Nennungen</th> </tr> </thead> <tbody> <tr><td>1</td><td>2</td></tr> <tr><td>3</td><td>2</td></tr> <tr><td>5</td><td>3</td></tr> <tr><td>7</td><td>4</td></tr> <tr><td>9</td><td>1</td></tr> <tr><td>11</td><td>4</td></tr> <tr><td>13</td><td>2</td></tr> <tr><td>15</td><td>6</td></tr> <tr><td>17</td><td>3</td></tr> <tr><td>19</td><td>3</td></tr> <tr><td>21</td><td>1</td></tr> <tr><td>23</td><td>1</td></tr> <tr><td>25</td><td>6</td></tr> <tr><td>27</td><td>2</td></tr> <tr><td>29</td><td>4</td></tr> <tr><td>31</td><td>0</td></tr> <tr><td>33</td><td>0</td></tr> <tr><td>35</td><td>1</td></tr> <tr><td>37</td><td>1</td></tr> </tbody> </table>	Jahr	Anzahl Nennungen	1	2	3	2	5	3	7	4	9	1	11	4	13	2	15	6	17	3	19	3	21	1	23	1	25	6	27	2	29	4	31	0	33	0	35	1	37	1	38
Jahr	Anzahl Nennungen																																									
1	2																																									
3	2																																									
5	3																																									
7	4																																									
9	1																																									
11	4																																									
13	2																																									
15	6																																									
17	3																																									
19	3																																									
21	1																																									
23	1																																									
25	6																																									
27	2																																									
29	4																																									
31	0																																									
33	0																																									
35	1																																									
37	1																																									
Ausbildungshintergrund	<table border="1"> <thead> <tr> <th>Kategorie</th> <th>Anteil (%)</th> </tr> </thead> <tbody> <tr> <td>Kaufmännisch</td> <td>22</td> </tr> <tr> <td>Ingenieurwesen</td> <td>26</td> </tr> <tr> <td>(Wirtschafts-)Informatik</td> <td>97</td> </tr> <tr> <td>Sonstige</td> <td>38</td> </tr> </tbody> </table>	Kategorie	Anteil (%)	Kaufmännisch	22	Ingenieurwesen	26	(Wirtschafts-)Informatik	97	Sonstige	38	183																														
Kategorie	Anteil (%)																																									
Kaufmännisch	22																																									
Ingenieurwesen	26																																									
(Wirtschafts-)Informatik	97																																									
Sonstige	38																																									
Alter (Anzahl Nennungen in der jeweiligen Gruppe)	<table border="1"> <thead> <tr> <th>Alter</th> <th>Anzahl Nennungen</th> </tr> </thead> <tbody> <tr><td>15-19</td><td>52</td></tr> <tr><td>20-34</td><td>51</td></tr> <tr><td>35-44</td><td>49</td></tr> <tr><td>45-54</td><td>43</td></tr> <tr><td>>54</td><td>15</td></tr> </tbody> </table>	Alter	Anzahl Nennungen	15-19	52	20-34	51	35-44	49	45-54	43	>54	15	164																												
Alter	Anzahl Nennungen																																									
15-19	52																																									
20-34	51																																									
35-44	49																																									
45-54	43																																									
>54	15																																									



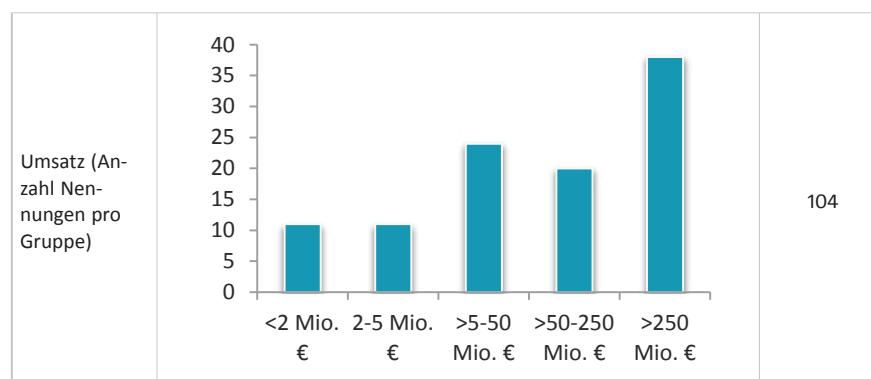


Tabelle 8: Hintergrund der Umfrageteilnehmer und Charakteristika der Unternehmen (Eigene Darstellung)

2.2.3 Big Data wird vor allem mit neuen Analysemethoden und Technologien verbunden

Für die Gewinnung eines ersten Überblickes wurde eine Tag-Cloud erstellt. Abbildung 25 zeigt die Schlagworte, welche die Befragten mit dem Begriff Big Data verbinden. Die Größe des Schlagworts repräsentiert die Häufigkeit der Nennung. Ferner ist zu beachten, dass aufgrund der automatisierten Text-Mining-Auswertung Begrifflichkeiten mit leicht unterschiedlicher Schreibweise mehrfach vorkommen können.



Abbildung 25: Tag-Cloud zum Thema Big Data (Eigene Darstellung)

Es ist zu erkennen, dass die Befragten mit Big Data vor allem technische Begriffe wie Business Analytics und Datenanalysen („Mining“) verbinden. Allerdings werden mit dem Schlagwort zum Thema Datenschutz bestehende Hemmnisse für den Einsatz von Big Data genannt.

2.2.4 Big Data wird in den kommenden fünf Jahren wichtig

Für eine Bewertung von Big Data und dessen Potential ist es entscheidend, wie Befragte aus der Praxis das Thema betrachten.

Abbildung 26 zeigt die Ergebnisse zur Einschätzung von Entscheidern und Anwendern hinsichtlich der Bedeutung von Big Data für die unternehmerische Tätigkeit in Abhängigkeit vom Zeithorizont. Die aktuelle Einschätzung im Jahr 2013 im Vergleich zu der in den kommenden fünf Jahren soll das Potential aufzeigen.

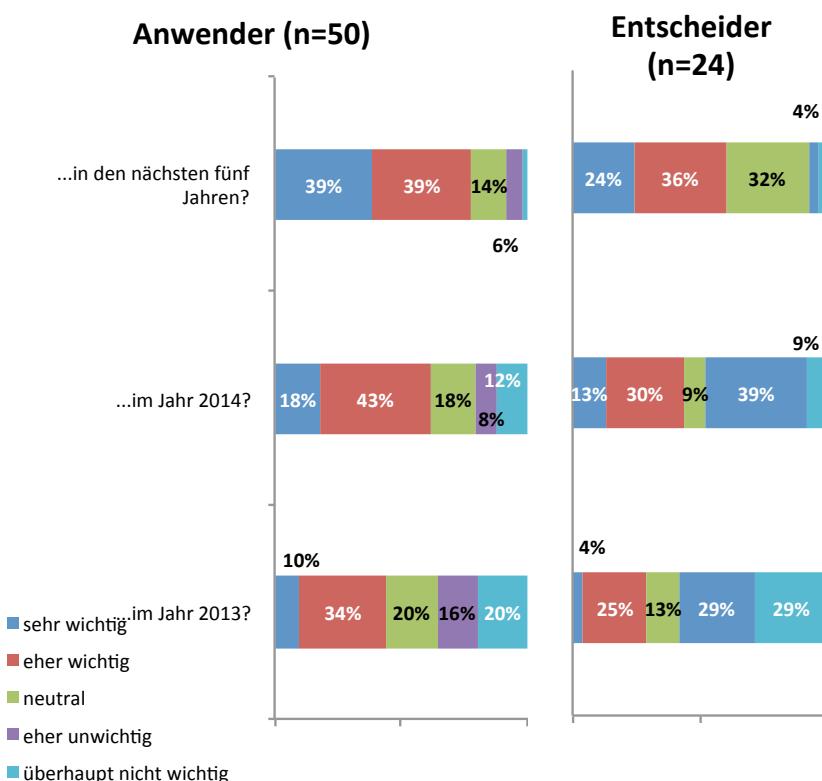


Abbildung 26: Wichtigkeit des Themas Big Data in der Praxis (Eigene Darstellung)

Alle Befragten sind sich einig, dass Big Data in den kommenden fünf Jahren ein bedeutendes Thema in der Praxis wird. Bereits im Jahr 2013 wird dem Big Data Thema eine hohe Wichtigkeit beigemessen. Es zeigt sich eine steigende Tendenz für die Relevanz der Big Data-Thematik im Jahr 2014. So halten 43 Prozent der Entscheider und 61 Prozent der Anwender das Thema für wichtig bis sehr wichtig. In den kommenden fünf Jahren, geben sogar 60 Prozent der Entscheider und 78 Prozent der Anwender an, dass es wichtig ist. Hieraus lässt sich eine deutliche Zunahme an Relevanz und damit das zu erwartende Potential erkennen.

2.2.5 Big Data hat ein hohes Wertschöpfungspotential

Aus Entscheider- und aus Anwendersicht beschreiben 50 Prozent bzw. 72 Prozent das Wertschöpfungspotential von Big Data als hoch bis sehr hoch. Von den Entscheidern sehen nur 25 Prozent und von den Anwendern nur 8 Prozent das Wertschöpfungspotential von Big Data als niedrig an. Diese Statistiken zeigen, dass die Befragten Big Data nicht nur als Trend sehen, sondern auch einen quantifizierbaren Einfluss von Big Data auf die Unternehmensergebnisse erkennen können.

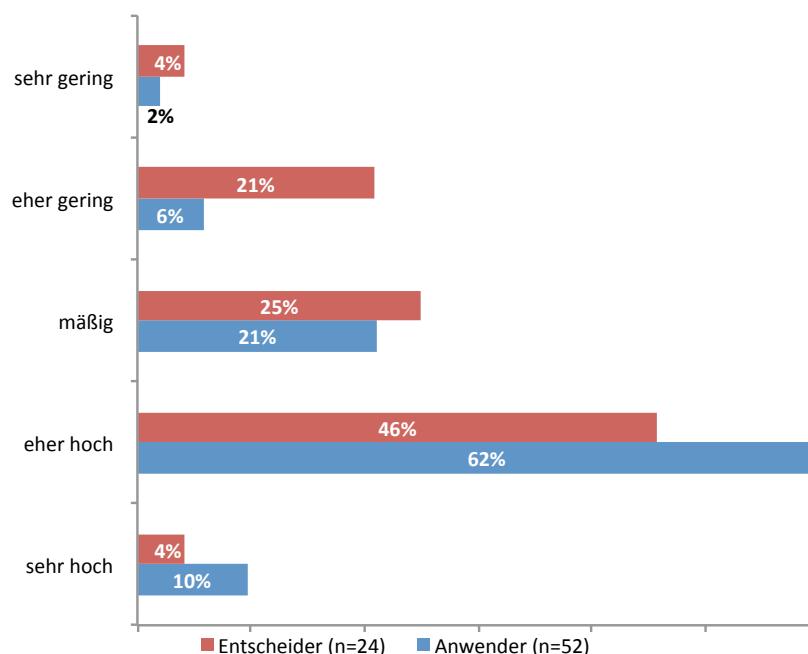


Abbildung 27: Wertschöpfungspotential von Big Data (Eigene Darstellung)

Hervorzuheben ist, dass Anwender die Potenziale von Big Data deutlich positiver bewerten als die Gruppe der Entscheider.

2.2.6 Einige Befragte konnten bereits messbaren Wert von Big Data beziffern

Von den Befragten geben 16 Prozent der Entscheider und 19 Prozent der Anwender an, dass sie bereits einen messbaren Wert mit Big Data schaffen konnten (vgl. Abbildung 28). Jedoch geben 72 Prozent der Entscheider und 54 Prozent der Anwender auch an, dass sie bis jetzt noch keinen messbaren Wert beziffern konnten. Diese Ergebnisse zeigen zum Teil noch die Diskrepanz zwischen dem Wertschöpfungspotential und dem tatsächlich bisher messbaren Beitrag.

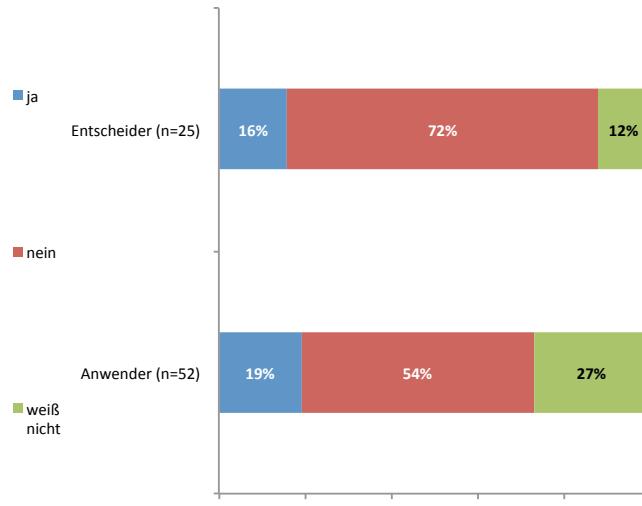


Abbildung 28: Messbarer Wertbeitrag von Big Data (Eigene Darstellung)

2.2.7 Big-Data-Lösungen werden überwiegend als Software as a Service oder als klassische Software angeboten

Von den untersuchten Anbietern geben fast ein Drittel an, dass sie Big-Data-Lösungen als Software as a Service anbieten (vgl. Abbildung 29). Hierbei wird die Lösung auf der IT-Infrastruktur eines externen Dienstleisters betrieben. Etwas mehr als ein Viertel der Anbieter geben an, die Lösung als klassische Software zu vertreiben.

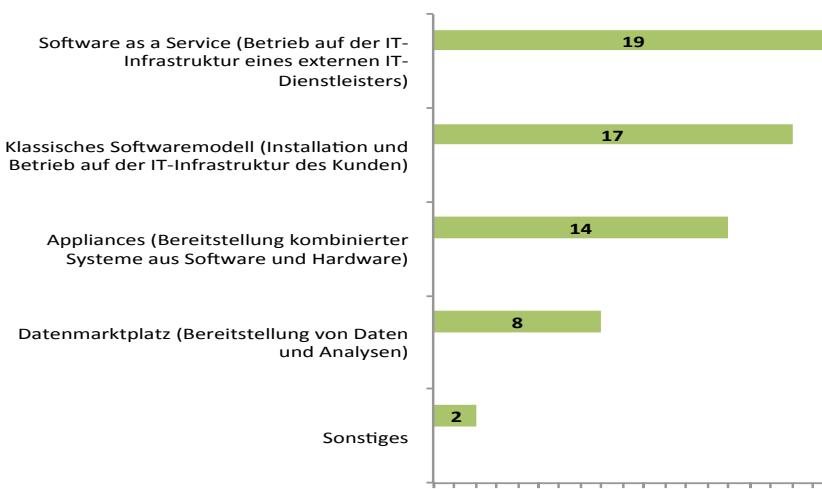


Abbildung 29: Geschäftsmodelle von Big Data-Anbietern, Mehrfachnennungen möglich, n=32 (Eigene Darstellung)

2.2.8 Big Data wird als schnelle Entscheidungen mit heterogenen Daten verstanden

Die Abbildung 30 zeigt, welche Schwerpunkte von den Befragten bei der Theematik von Big Data gelegt werden.

Besonders die große Vielfalt und Heterogenität der Datenquellen und -formate nennen 41 bis 60 Prozent der Befragten als charakteristisch für Big Data. Nur geringfügig weniger charakteristisch sehen die Befragten die zunehmende Geschwindigkeit der Datenerzeugung, -analyse und -verarbeitung. 55 bis 79 Prozent der Befragten geben bei der Frage nach der Handhabung unvollständiger, ungenauer Daten mit unvorhersehbaren Datenstrukturen und Abschätzungen von deren Korrektheit an, dass sie diese Frage mit Big Data in Verbindung bringen. Dies ist interessant, da es entgegen unseren Erwartungen von geringerer Bedeutung ist. Es kann demgemäß, wie die Abbildungen zeigen, angenommen werden, dass die technologischen Voraussetzungen bereits weitgehend vorhanden sind.

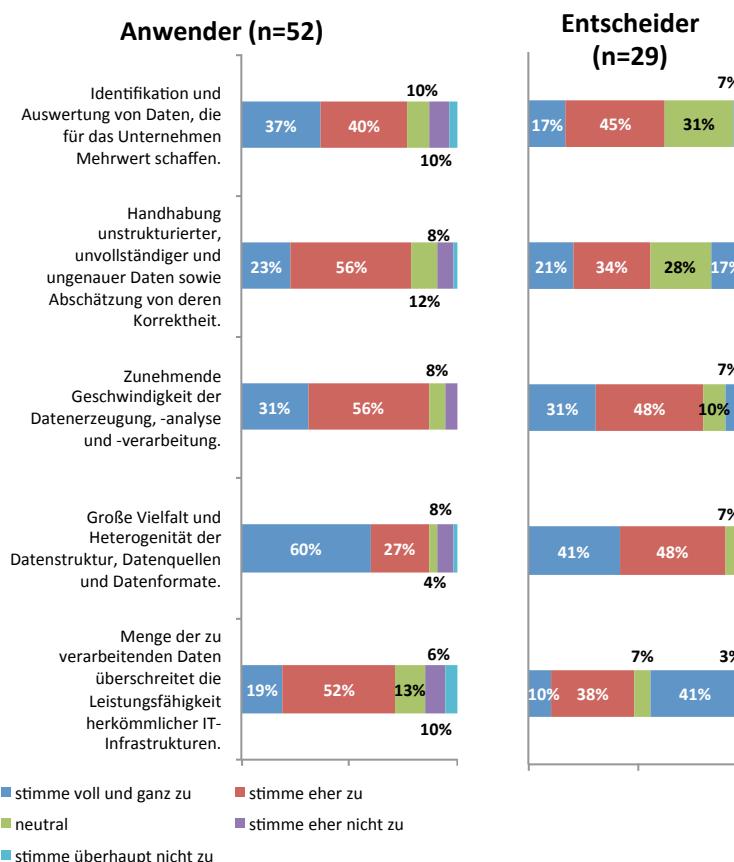


Abbildung 30: Charakteristika von Big Data 1 (Eigene Darstellung)

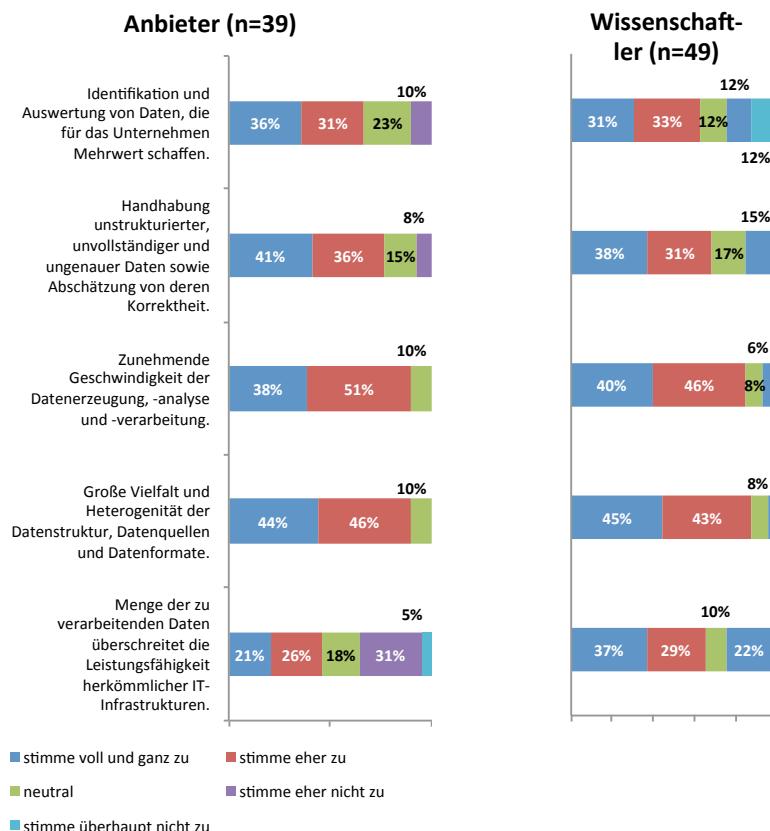


Abbildung 31: Charakteristika von Big Data 2 (Eigene Darstellung)

2.2.9 Big Data führt zu neuen Geschäftsmodellen

Die deutliche Mehrheit von 60 bis 67 Prozent der Befragten (vgl. Abbildung 32) erwartet disruptive Effekte von Big Data, d.h. innovative, neue Geschäftsmodelle, Märkte und grundlegende Veränderungen, welche unaufhaltsam bestehende Technologien, Produkte sowie Dienstleistungen teilweise oder vollständig verdrängen. Zudem geben zwischen 12 und 19 Prozent der Befragten an, dass Big Data neuartige Geschäftsprozesse ermöglicht welche zu signifikanten Kosteneinsparungen und/oder signifikantem Umsatzwachstum führen. Die verbleibenden Befragten geben an, dass Big Data bestehende Prozesse optimiert und somit zu moderaten Kosteneinsparungen und/oder moderatem Umsatzwachstum

führt. Schließlich wird in der Gestaltung neuer Geschäftsprozesse — verbunden mit signifikanten Auswirkungen auf Umsatz und Kosten — ein wesentlicher Effekt von Big Data gesehen.

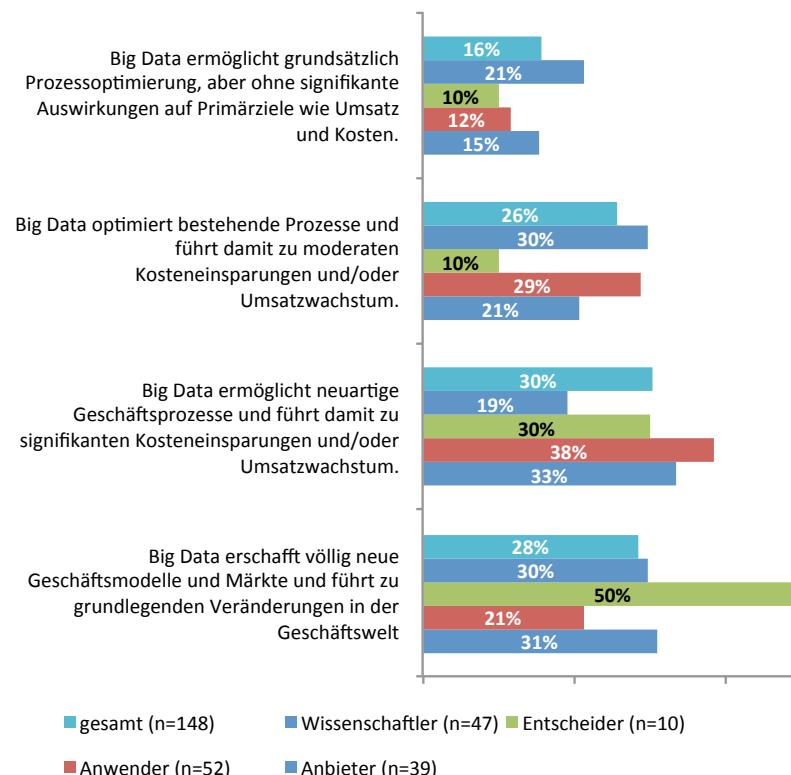


Abbildung 32: Geschäftspotentiale von Big Data (Eigene Darstellung)

2.2.10 Big Data ermöglicht bessere unternehmerische Entscheidungen

Zwischen 77 Prozent und 82 Prozent der Befragten erwarten von Big Data eine bessere Datenanalyse, die zu besseren Entscheidungsgrundlagen (vgl. Abbildung 33) führt. Ebenfalls erwarten ca. 2/3 der Befragten, dass die Reaktionsschnelligkeit der Unternehmung sowie die Datenaktualität gesteigert werden können. Ebenfalls stimmen ca. 2/3 aller Befragten zu, dass Big Data zur Entwicklung neuer Produkte und Dienstleistungen sowie zur Erschaffung neuer Geschäftsmodelle und Erlösströme beiträgt.

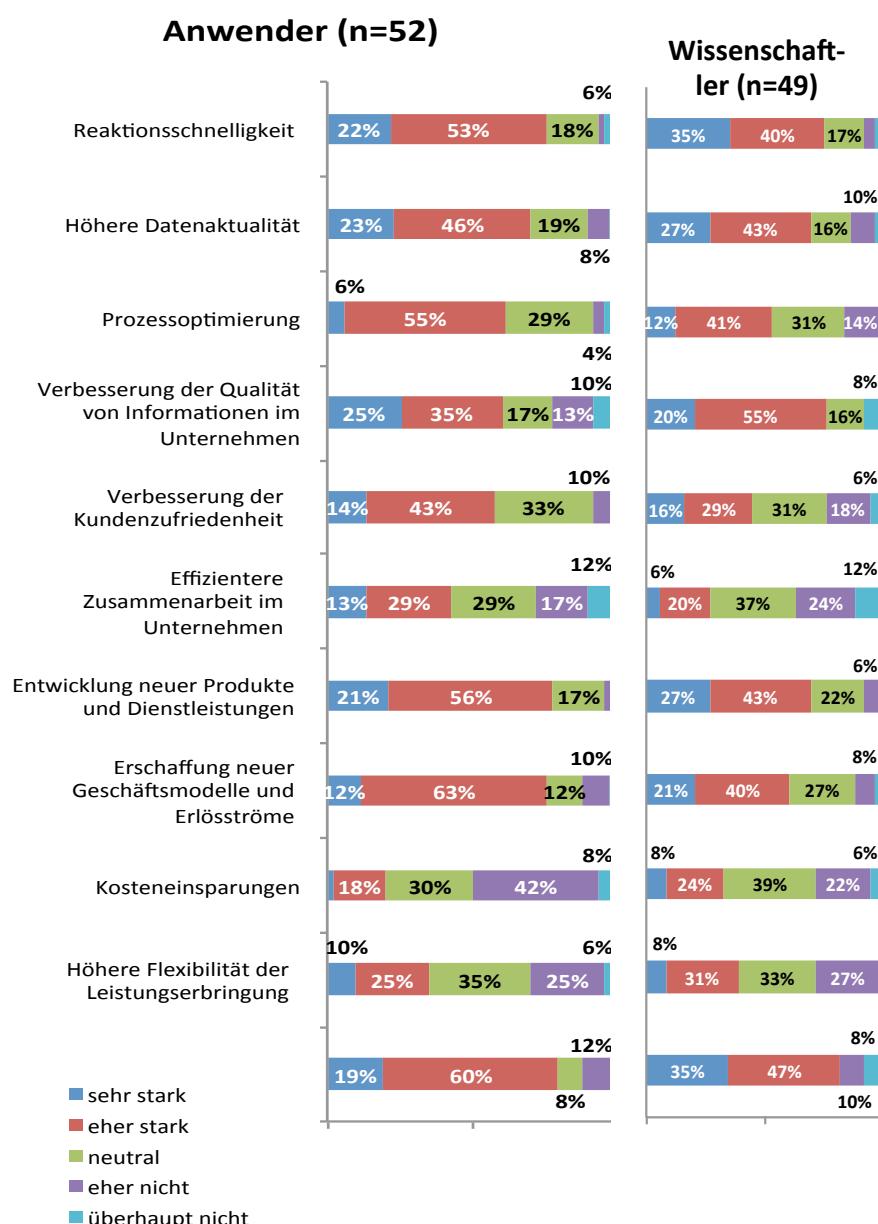


Abbildung 33: Big Data und unternehmerische Entscheidungen 1 (Eigene Darstellung)

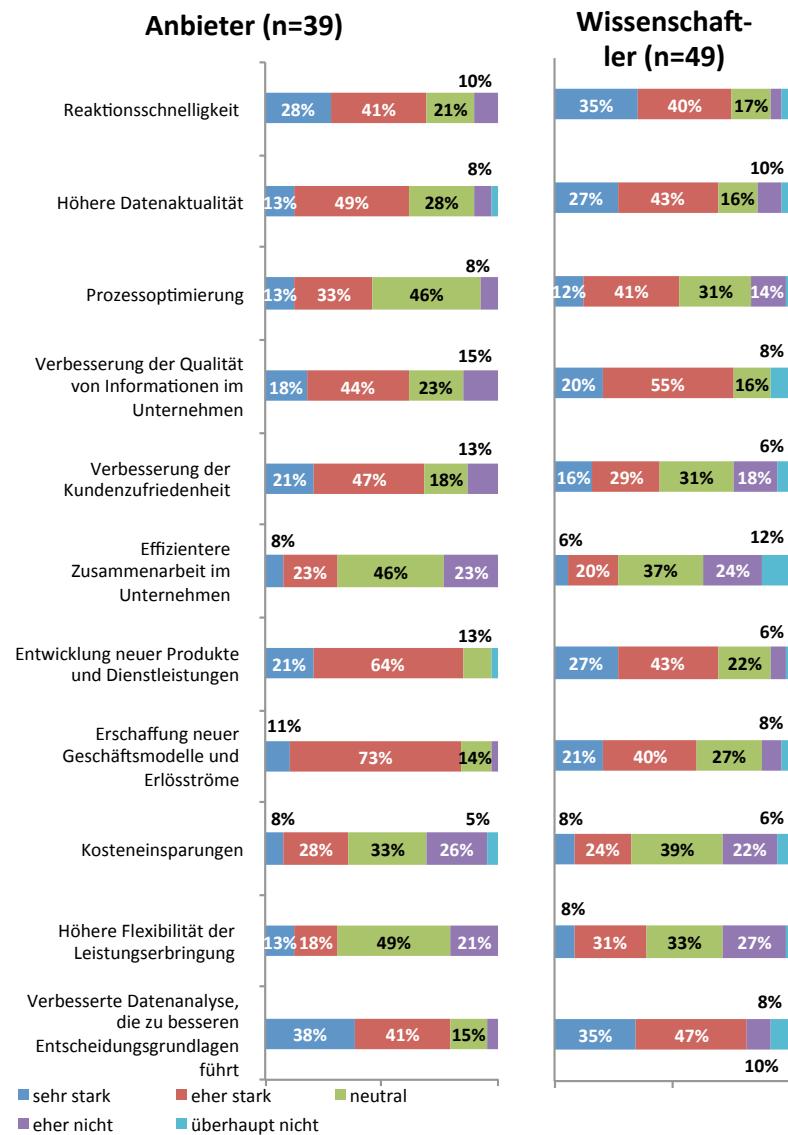


Abbildung 34: Big Data und unternehmerische Entscheidungen 2 (Eigene Darstellung)

2.2.11 Big Data hat vor allem Potenzial für eine effiziente Unternehmensführung

Experten der Entscheider- (67 Prozent) und der Anwendergruppe (86 Prozent) sehen das größte Wertschöpfungspotential für Big Data im Marketing und Vertrieb (vgl. Abbildung 35). Jedoch sehen auch ca. 2/3 der Befragten aus beiden Gruppen Potential für den Kundenservice und für die Geschäftsführung. Umgekehrt sehen die Befragten (ca. die Hälfte der befragten Experten) am wenigsten

Potential für Big Data für die Funktionen Finanzen und Buchhaltung bzw. Compliance.

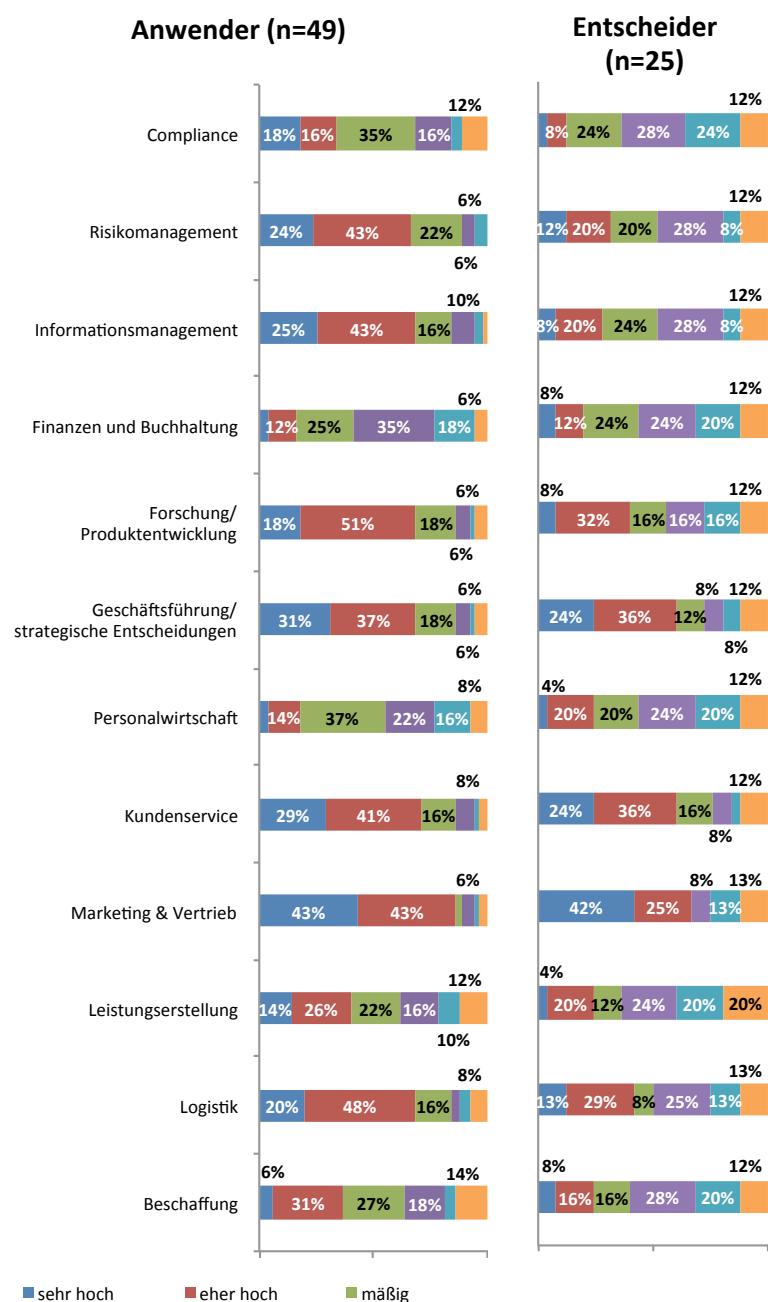


Abbildung 35: Wertschöpfungspotential für betriebliche Funktionen (Eigene Darstellung)

2.2.12 Durch Big Data wird die IT im Unternehmen zum Datenanalytiker

Abbildung 36 zeigt, welche neuen Aspekte sich für die Rolle der IT im Unternehmen aufgrund von Big Data aus Sicht der Befragten ergeben.

Die Befragten sehen vor allem, dass sich die Rolle der IT in Richtung „Data Science“ wandeln wird und dass die IT-Abteilungen/Spezialisten/Funktionen sich besonders mit komplexen und innovativen Methoden der Datenanalyse auskennen sollten. Im Weiteren wird die IT zunehmend als Gestalter von unternehmerischer Tätigkeit gesehen. Es wird davon ausgegangen, dass Big Data die Rolle der IT im Unternehmen weiter stärken wird. Zwischen 70 Prozent und 86 Prozent der Befragten stimmen zudem zu, dass in vielen Unternehmen ungenutzte Datenbestände existieren, die für das Unternehmen Mehrwert generieren können. Immerhin ca. die Hälfte aller Befragten gibt allerdings auch an, dass der Hype um Big Data zu überhöhten Erwartungen führt.

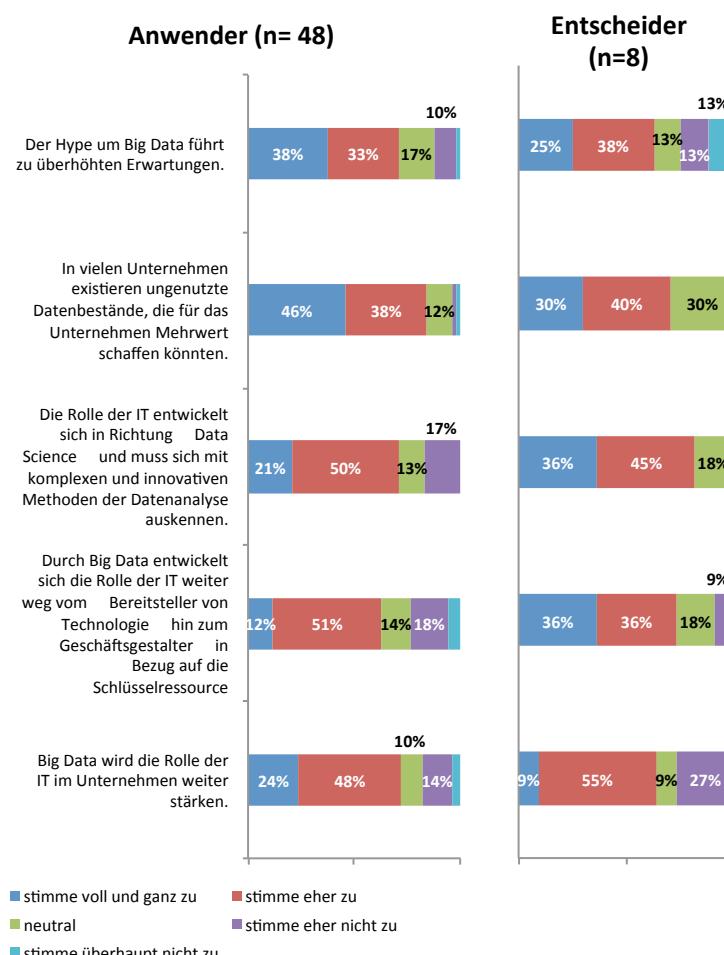


Abbildung 36: Die Rolle der IT im Big Data-Kontext 1 (Eigene Darstellung)

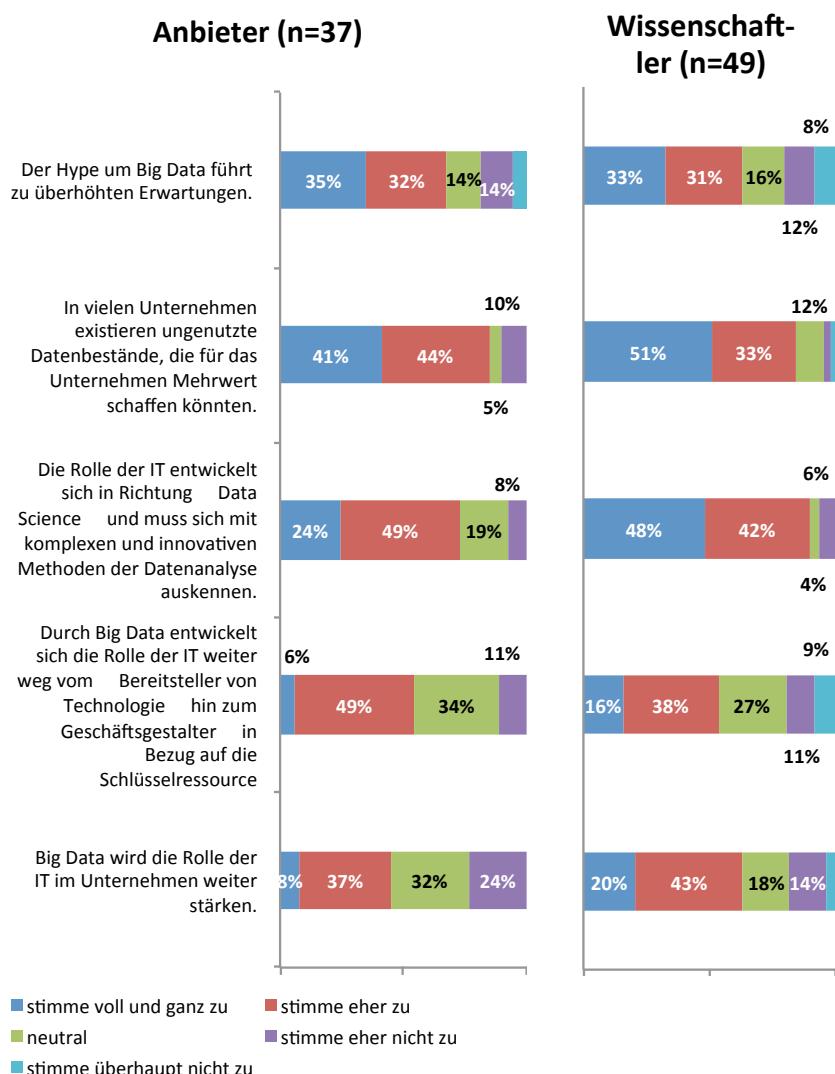


Abbildung 37: Die Rolle der IT im Big Data-Kontext 2 (Eigene Darstellung)

2.2.13 Der Wert von Daten im Unternehmen wird erkannt

Bei der Nutzung von wertvollen Unternehmensdaten besteht derzeit noch weiter Entwicklungsbedarf. 50 bzw. 64 Prozent der Entscheider und Anwender sehen zwar einen hohen Wert in bestimmten Daten, könnten diesen aber noch besser nutzen (vgl. Abbildung 38). Lediglich 20 bzw. 13 Prozent der Befragten geben an, bereits eine ganzheitliche Strategie zu haben, um die wertvollsten Daten zu verwenden. Die restlichen Befragten geben an, dass sie Daten sammeln deren Auswertung für sie aber nicht wichtig sei.

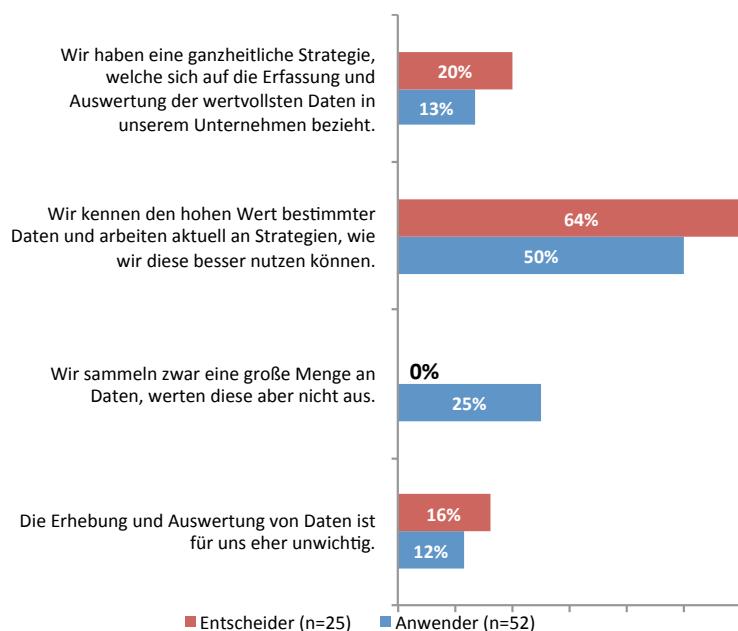


Abbildung 38: Einschätzung des Wertes von Unternehmensdaten (Eigene Darstellung)

2.2.14 Big Data-Projekte sind noch in sehr frühen Phasen

Die Abbildung 39 spiegelt den aktuellen Stand der Planung zur Einführung von Big Data wieder. 40 Prozent der Entscheider geben an, in der Informationsphase über Big Data-Technologien zu sein. Lediglich 8 Prozent der Entscheider haben sich bereits mit der Umsetzung beschäftigt. Ein Viertel der Entscheider gab an, dass sie bereits die Entwicklung von Strategien, Maßnahmen und Roadmaps planen bzw. prüfen. Dazu gehört ebenfalls eine Kosten-Nutzen-Analyse. Auf der Anwenderseite hingegen sind Unternehmen bereits weiter fortgeschritten. Hier geben bereits 15 Prozent an, eine Big Data Strategie umgesetzt zu haben. Auch befinden sich mehr Unternehmen in den Phasen Umsetzen, Prüfen und Planen.

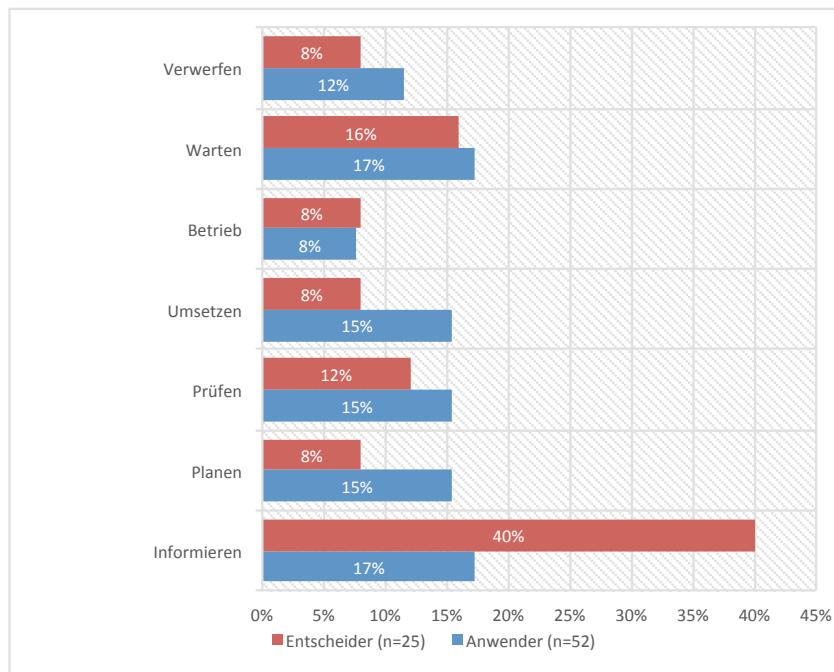


Abbildung 39: Status von Big Data-Projekten (Eigene Darstellung)

2.2.15 Big Data wird sich in weniger als 10 Jahren etabliert haben

Abbildung 40 gibt den geschätzten Zeitraum an, in dem sich Big Data flächendeckend durchsetzen wird. Der Großteil der Befragten geht davon aus, dass sich Big Data in weniger als 10 Jahren in ihrer Branche flächendeckend durchsetzen wird. Diese Einschätzung wird auch von Anbietern und Entscheidern geteilt.

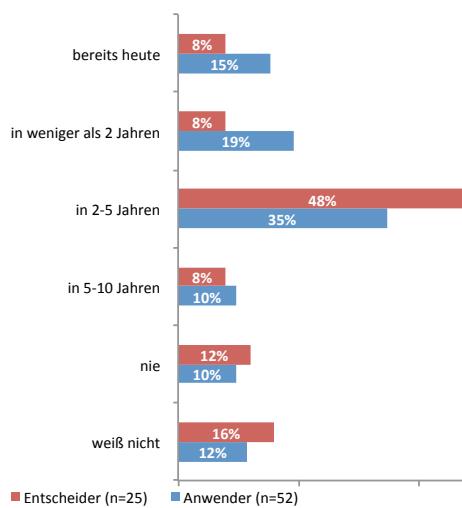


Abbildung 40: Wann wird Big Data wettbewerbsentscheidend (Eigene Darstellung)

2.2.16 Neben relationalen Daten dominieren Transaktionsdaten, Textanalysen und Webanalysen

Die Anbieter von Big Data Technologie sind breit aufgestellt und können für eine Vielzahl von Datentypen entsprechende Analysewerkzeuge anbieten. Es dominieren aber Werkzeuge für relationale Daten und stark strukturierte Daten wie Transaktionsdaten. Der Analyse von Video-, Bild-, und Audiodaten wird vergleichsweise wenig Bedeutung beigemessen.

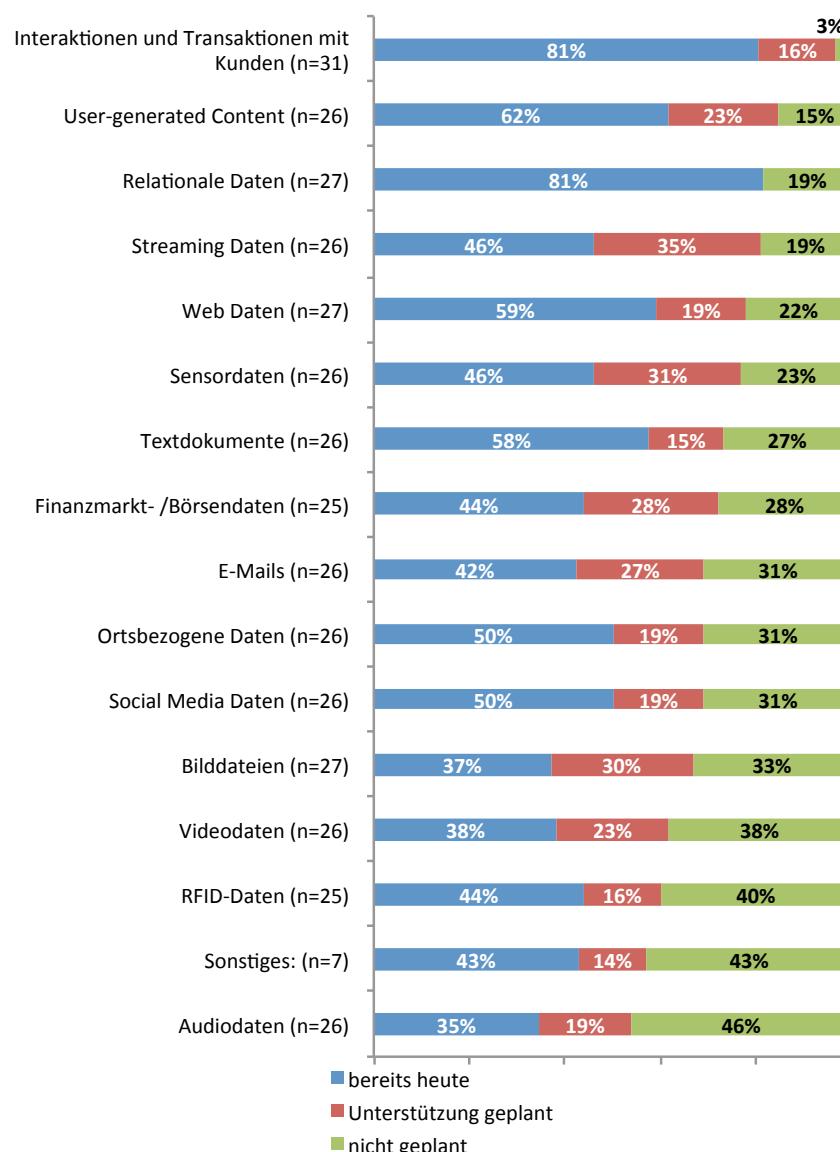


Abbildung 41: Angebot an Big Data Werkzeugen anhand der verarbeiteten Datentypen (Anbieterperspektive)

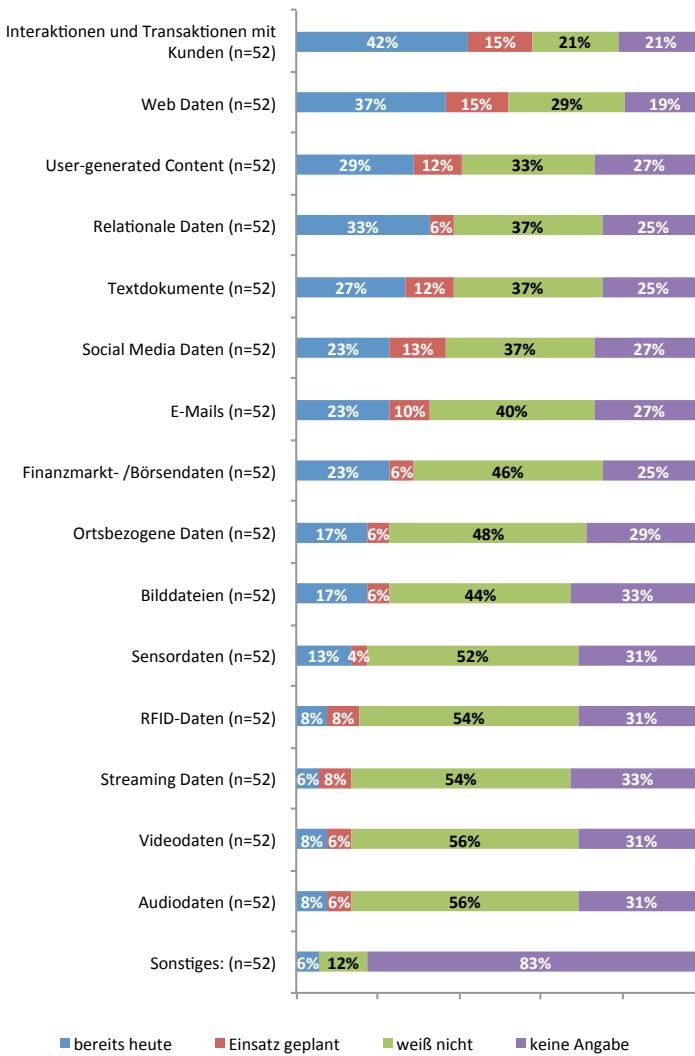


Abbildung 42: Potenzial für Big Data Werkzeuge aus Sicht der Anwender

Aus Anwendersicht steht die Analyse von Transaktionsdaten und Daten aus dem Internet im Vordergrund, wenn es um bereits durchgeführte Analysen geht. Auf Anwenderseite ist vor allem interessant, dass nur wenige Befragte den zukünftigen Einsatz planen, sondern vielmehr noch gar nicht wissen, ob bestimmte Analysen durchgeführt werden sollten.

2.2.17 Hochqualitative Daten, fehlendes Fachpersonal und mangelnde Wirtschaftlichkeit sind die größten Herausforderungen

Entscheider sehen die größten Hemmnisse für die Adoption von Big Data vor allem in fehlenden Möglichkeiten, die Datenanalysen in Entscheidungen einfließen zu lassen. Auch haben Entscheider große Bedenken bezüglich

datenschutzrechtlicher Rahmenbedingungen. Ebenso wird ein Mangel an qualifiziertem Personal für Datenanalysen konstatiert.

Auf Anwenderseite dominieren ebenso die Bedenken hinsichtlich Datenschutz und Privatsphäre. Nahezu gleichwertig herausfordernd wird die Einbindung von Datenanalysen in Entscheidungsprozesse gesehen.

Anbieter hingegen kämpfen vor allem mit Bedenken hinsichtlich des Datenschutzes. Ebenso sehen sich Anbieter vor der Herausforderung genau zeigen zu können, wie die Datenanalysen in die Entscheidungsprozesse einfließen können. Anbieter kämpfen zudem mit den komplexen Strukturen der Daten in den Unternehmen.

Zusammenfassend kann festgehalten werden, dass die größten Herausforderungen in einer wertschöpfenden Einbindung von Big Data in die Entscheidungsprozesse der Unternehmen, ein Mangel an qualifiziertem Personal und Bedenken hinsichtlich des Datenschutzes sind.

Interessanterweise gibt es keine großen Unterschiede hinsichtlich der Bedeutung der Herausforderungen, wenn man auf kleine und mittelständische Unternehmen fokussiert. Auch hier stehen die Themen des Datenschutzes sowie der fehlenden Entscheidungsorientierung von Datenanalysen im Vordergrund. Ebenso fehlen auch hier entsprechend qualifizierte Mitarbeiter. Zusätzlich kann es als besondere Chance für den Mittelstand gesehen werden, insbesondere durch strukturelle Herausforderungen wie nicht belastbare Business Cases oder fehlender Zugang zu Daten nur kaum genannt werden. Wenn kleine und mittelständische Unternehmen mit Big Data Anbietern kooperieren könnten, würde eine Ausnutzung der Potenziale von Big Data also auch im Mittelstand möglich sein.

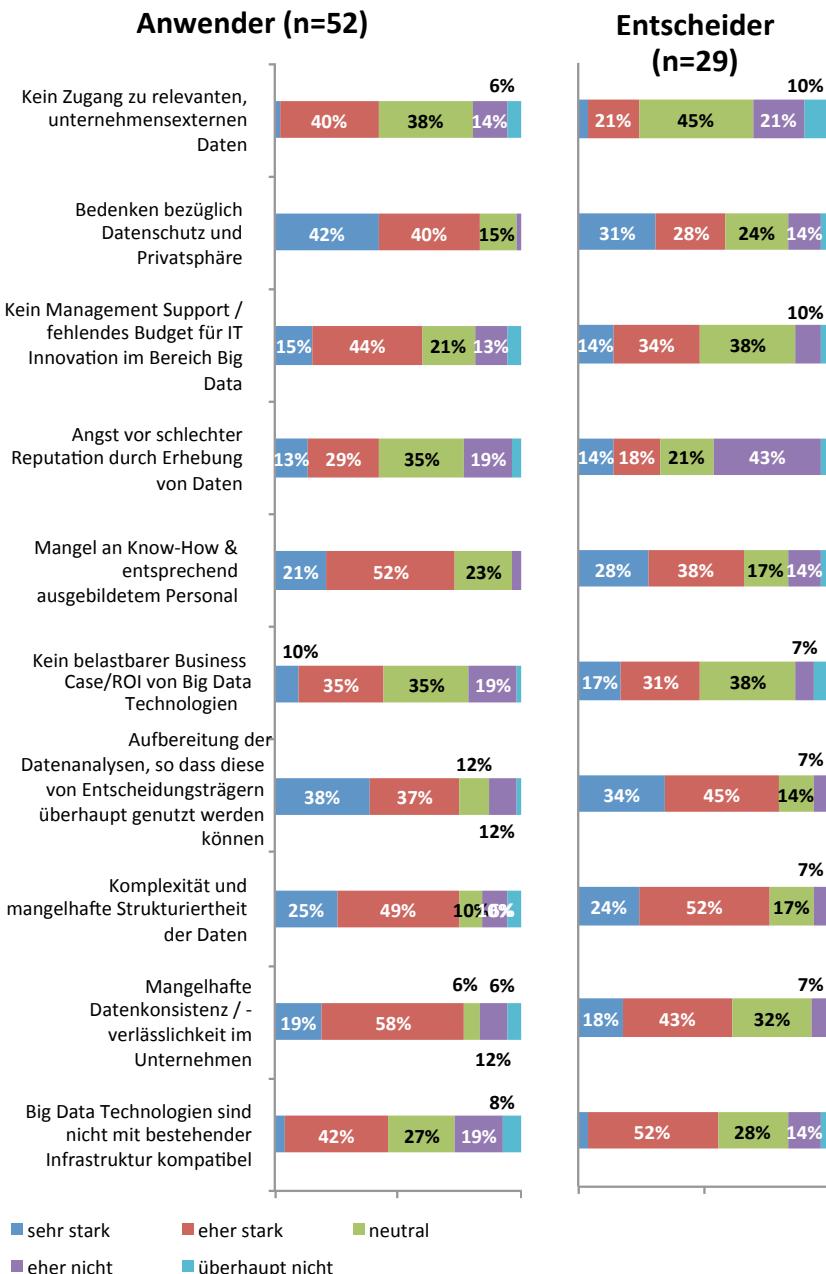


Abbildung 43: Hürden und Herausforderungen im Zusammenhang mit Big Data 1 (Eigene Darstellung)



Abbildung 44: Hürden und Herausforderungen im Zusammenhang mit Big Data 2 (Eigene Darstellung)

2.3 Ableitung von Innovationspotenzialen

Ableitung von branchenübergreifenden Innovationspotentialen von Big Data. Das Innovationspotenzial wird im nächsten Schritt im Allgemeinen und branchenübergreifend abgeleitet. Dazu werden aus den untersuchten Studien

und der eigenen Studie allgemeine branchenübergreifende Potentiale abgeleitet und die wichtigsten Begrifflichkeiten näher betrachtet. Abschließend wird eine SWOT-Analyse durchgeführt, um das Potential im Allgemeinen einschätzen zu können.

Aufgrund einer immer weiter steigenden Menge an Informationen (Big Data) wächst folglich das Potential zur Entwicklung von IKT-Anwendungen, welche in der Lage sind diese großen Datenmengen intelligent zu verarbeiten. Für eine nachhaltige technische, ökonomische und gesellschaftliche Nutzung bietet Big Data in vielen Branchen Potential für neue Dienstleistungen, neue Produkte und neue Geschäftsmodelle.

Gadatsch nennt drei wesentliche Systeme für eine Anwendung von Big Data, die Sensor- oder Logdaten, die Daten der mobilen IT und die Social Web Daten (174, S. 1616-1167). Gadatsch spricht von „Einsatzmöglichkeiten [...] in nahezu allen Branchen“ (174, S. 1618). Aus unterschiedlichen Branchen beschäftigen sich bereits verschiedene Firmen in der Praxis mit der Thematik Big Data, wie die eigene Erhebung im Unterabschnitt 2.2 ergeben hat.

Laut Schätzungen von Gartner aus dem Jahre 2011 werden 85 Prozent der Fortune-500-Unternehmen bis 2015 keinen Wettbewerbsvorteil erzielen können, da sie nicht in der Lage sein werden, Big Data effektiv zu nutzen (175). Für die effektive Nutzung sind aus der Wirtschaft, Forschung und Entwicklung demzufolge innovative Lösungsmöglichkeiten für die Big Data-Technologie gefordert. Die Studie der Tata Consulting zeigt für das Jahr 2012 auf, dass die Vereinigten Staaten mit bereits vorhandenen Big Data Initiativen von 68 Prozent im Vergleich zu Europa mit 45 Prozent führend sind (176, S. 13). Deutschland besitzt mit 34 Prozent ein erhebliches Ausbaupotential an Big Data Initiativen (176, S. 13). Der Vollständigkeitshalber sei erwähnt, dass unter Big Data Initiativen Projekte in Unternehmen zu verstehen sind, welche Big Data Technologie benutzen, dies kann zum Beispiel der Einsatz von Business Analytics sein.

In Europa ist das Vereinigte Königreich mit 63 Prozent an Big Data Initiativen in Unternehmen (176, S. 13). Dies war aufgrund des ähnlichen Sprachraumes zu den USA zu erwarten. Darüber hinaus ist der englischsprachige Raum deutlich größer und adressiert somit einen größeren Markt. Die Vielfalt der Daten in Form von übereinstimmenden Wörtern ist in Europa, besonders wegen der unterschiedlichen Sprachen größer. Jedoch können sich einzelne große deutsche Unternehmen bereits durchsetzen und neue innovative Lösungen präsentieren. Ein Beispiel hierfür ist SAP mit der HANA Plattform. Mit SAP HANA ist es durch die In-Memory Technology möglich, große Datenmengen erheblich schneller zu analysieren.

Es wird prognostiziert, dass das Datenvolumen auf bis zu 100 Zettabyte im Jahr 2020 wächst, damit wird die Verarbeitung dieser Datenmengen zunehmend schwieriger (177, S. 12). Für die Verarbeitung der Datenmengen hat Tata Consulting einen ROI von 45,5 Prozent im Durchschnitt je Branche im Jahr 2012 ermittelt (176, S. 26). Ceteris paribus ist davon auszugehen, dass die Umsätze mit Big Data in den nächsten Jahren steigen werden, insbesondere vor dem Hintergrund, dass sich die Big Data-Technologie am Anfang ihres Lebenszykluses befindet. Nach der Big Data Technologieweiterentwicklung sind die Einführung von neuen Dienstleistungen, neuen Produkten und weiteren Technologieentwicklungen denkbar.

Diese Entwicklungen in der Big Data-Technologie zeigen unterschiedliche Vorteile auf, wie z.B. Zeitersparnis, Transparenz, Kosteneinsparung, Flexibilität, Validität. Zu dem kann Big Data in einer Vielzahl von Branchen und einzelnen Themenfeldern angewendet werden, wie z.B. Cloud Computing, Industrie 4.0, E-Mobility, E-Government, Versicherungswesen, Energie und Marktforschung. Für diese Anwendungen sind insbesondere Technologien erforderlich, die die Verarbeitung von Big Data in der notwendigen Geschwindigkeit ermöglichen. Hierzu zählen beispielsweise In-Memory Datenbanken, Apache Hadoop und HBase. Darüber hinaus ist die Informationsverarbeitung und –aufbereitung für Entscheidungsträger insbesondere für den Mittelstand der deutschen Wirtschaft wichtig. In diesen Bereichen können neue Produkte und Services die KMUs unterstützen und somit das Potential der KMUs stärken sowie den eigenen Absatz erhöhen. Dies wäre eine Win-Win-Situation die sowohl die jetzigen großen Anbieter wie SAP, Deutsche Telekom, Atos stärkt aber auch den KMUs ermöglicht individuelle Lösungen anbieten zu können.

Das Potential für den Standort Deutschland steckt in der Entwicklung von Big Data verarbeitender Technologie, um so zu den Marktführern in den USA aufzuschließen. Big Data-Technologien ermöglichen ein neues Angebot von Dienstleistungen, Produkten und Technologien, nicht nur in Deutschland selbst, sondern weltweit. Dies wird besonders deutlich, wenn die vom IT-Branchenverband BITKOM prognostizierten globalen Umsätze im Bereich Big Data mit 3,3 Milliarden Euro für das Jahr 2011 betrachtet werden. Der Verband schätzt dabei eine mittlere Wachstumsrate von 36 Prozent im Zeitraum von 2012 bis 2016, was einem Umsatzzanstieg von 4,5 Milliarden auf 15,7 Milliarden Euro entspricht (177, S. 10). Dabei rechnet Gartner prognostiziert dafür 4,4 Millionen Jobs, welche im Bereich von Big Data weltweit entstehen könnten (178).

Für Big Data Anwendungen lassen sich unterschiedliche Märkte adressieren. Dazu gehören für die Auswertung der Daten Beratungsunternehmen oder Unternehmen, welche sich mit Business Intelligence und Analyse beschäftigen (z.B. von Gartner, IBM, Google Analytics). Aber auch Versicherungsunternehmen und Banken, wie Munich Re, Allianz, Deutsche Bank, Commerzbank, welche

seit Jahren bereits Business Intelligence sowie Analyseverfahren und -methoden einsetzen, um genaue Risikoanalysen durchführen zu können, sind prädestiniert. Dort ist Big Data Technologie bereits im Einsatz. Daraus ergeben sich neue Produkte und neue Dienstleistungen wie z.B. individuell zugeschnittene Verträge auf die Kundenwünsche, welche mittels Echtzeitauswertung, stabilere sowie validere Prognosen, Transparenz und Zeitersparnis unterstützt werden. Weitere Forschungsgebiete, die die „junge“ Thematik Big Data-Technologie unterstützen sind denkbar, z.B. In-Memory-Datenbanken.

Es kommt dabei nicht nur auf die Verarbeitung der Daten und die Geschwindigkeit an, sondern auch auf den Transfer der Daten von Ort A nach Ort B. Es bedeutet logistisch zu gewährleisten, dass die richtige Information zum richtigen Zeitpunkt in der richtigen Menge am richtigen Ort in der erforderlichen Qualität vorhanden ist. Für Big Data bedeutet es, die Informationen zum einen transparent und verständlich aufzubereiten sowie diese auf das Nötigste zu reduzieren, um sie in adäquater Zeit verarbeitbar und verwendbar zur Verfügung zu haben.

Daneben sind neue Visualisierungstechnologien zu entwickeln und bestehende zu verbessern. Zu beachten ist, dass 80 Prozent der Daten unstrukturiert sind (IBM, Big Data und Smarter Analytics 2012). Eine Auswertung erfordert daher i.d.R. verschiedene Analyse-Algorithmen. Dies erfolgt mit unterschiedlichen Technologien, wie RDBMS, Data Warehousing, ETL, OLAP, BPM, Data Mining, Clustering, Regression, Classification, Association Analysis, Anomaly Detection, neurale Netzwerke und so weiter.

Bevor die Auswirkung auf der Entscheidungsträgerebene genauer analysiert werden kann, fassen wir zusammen, was Big Data-Technologie leisten könnte: Die durch Big Data-Technologie ermöglichte zeitigere, transparentere, genauere, verlässlichere und umfänglichere Bereitstellung findet sich z.Z. in neuen Dienstleistungen hauptsächlich aus dem Bereich des Business Analytics, der Business Intelligence und der Marktforschung wieder. Die Informationen könnten mit Hilfe von Big Data so verarbeitet und aufbereitet werden, so dass diese für Entscheidungsträger in angemessener Menge und Qualität zur Verfügung stehen (179, S. 97). Big Data-Technologie könnte es ermöglichen den von Krcmar beschriebenen Teufelskreis einer immer weiter wachsende Informationsnachfrage in der Informationsflut zu durchbrechen (179, S. 97). Dies wäre jedoch etwas viel erwartet für diese Technologie allerdings wirkt sie der Informationsflut mit Analyseverfahren entgegen, wo durch ein erheblicher Vorteil entsteht.

Durch die entstehenden Vorteile insbesondere in der zeitigeren und verständlicheren Zurverfügungstellung von Informationen, kann dies besonders für die Entscheidungsträgerebene im Mittelstand der deutschen Wirtschaft einen Wettbewerbsvorteil bieten. Der Vorteil für KMU belegt auch eine Studie von Bange et al. vom BARC-Institut, sie sehen das Potential von Big Data-Werkzeugen bei

21 Prozent der großen, 38 Prozent bei kleinen und 41 Prozent bei mittleren Unternehmen (179, S. 7). Dies sind 79 Prozent aller Unternehmen und somit eine deutliche Mehrheit. Diese Unternehmen sind in unterschiedlichen Branchen zu finden. Abbildung 45 zeigt, wie sich die einzelnen Unternehmen nach Branchen aufteilen.

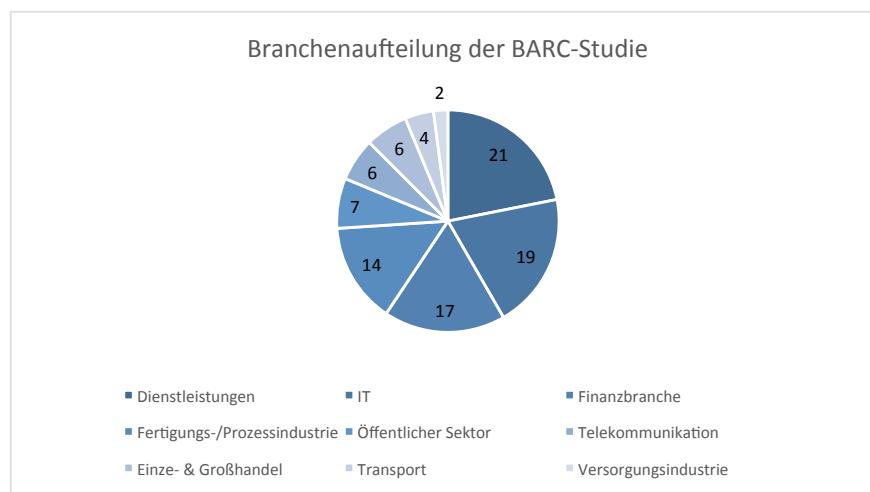


Abbildung 45: Branchenaufteilung nach der Studie des BARC-Instituts (In Anlehnung an (180, S. 9)).

Der hohe Anteil an Unternehmen aus der Dienstleistungsbranche deutet auf das Potential in diesem Bereich hin. Auch Unternehmen aus der IT- und der Finanzbranche sowie der Fertigungs-/Prozessindustrie waren in der Studie großzahlig vertreten. Die Top-5-Branchen der Studie des BARC-Instituts sind Dienstleistungen, IT, Finanzen, Fertigungs-/Prozessindustrie und öffentlicher Sektor. Das BARC-Institut zeigt in seiner Studie die Vorteile einer Nutzung von Big Data auf. Dabei wurden die Unternehmen hinsichtlich verschiedener Aussagen befragt. Der prozentuale Wert gibt die Anzahl an Unternehmen an, welche Angaben zur Aussage vorgenommen. In Tabelle 9 sind die Prozentanteile zu den Aussagen dargestellt (179, s. 25).

Aussage	Anteil in Prozent
Bessere strategische Entscheidungen	59
Bessere Steuerung operativer Prozesse	51
Schnellere Analyse	50
Detailliertere Analysen	43
Verbesserter Kundenservice	32
Zielgerichtetere Marketingaktionen	31
Besseres Verständnis des Marktes	28
Geringere Kosten	28
Bessere Produkt- / Service-Qualität	25
Bessere Kundenbindung	16

Tabelle 9: Vorteile von Big Data-Technologie für Unternehmen in Prozent (n=274)

Die drei größten Vorteile werden in besseren strategischen Entscheidungen, einer besseren Steuerung von operativen Prozessen und einer schnelleren Analyse gesehen. Was die bisherigen Aussagen zusätzlich bekräftigt.

Neben den Vorteilen sind die Hemmnisse von Big Data näher zu betrachten. Das BARC-Institut führt in seiner Studie hierzu unterschiedliche Probleme beim Einsatz von Big Data auf.

Erstens ist eines der Hauptprobleme fehlendes Wissen beim Einsatz von Big Data (179, S. 6). Würde dieses Problem behoben, könnte Big Data-Technologie effizienter, koordinierter und verlässlicher eingesetzt werden. In der Studie geben 46 Prozent der Befragten ein Fehlen von technischem und 44 Prozent ein Fehlen von fachlichem Wissen an (179, S. 6). Demgegenüber stehen mit 36 Prozent überzeugende Einsatzszenarien, mit 34 Prozent technische Probleme sowie mit 33 Prozent die Kosten für den Einsatz von Big Data. Zweitens gibt das Institut mit 25 Prozent den Datenschutz und mit 15 Prozent das Big Data nicht für Fachanwender im Unternehmen nutzbar sei an (179, S. 26). Ein Hemmnis könnte der ausgeprägte Datenschutz und die Datensicherheit sein, welche eine gesellschaftliche Akzeptanz verhindern könnten. In diesem Zusammenhang besteht Forschungsbedarf, um bekannte geeignete Schutzverfahren einfach und besonders für KMU zu ermöglichen. Insbesondere bei Start-Ups wo der Datenschutz auch ein finanzieller Faktor ist, ist dieser groß. Es gibt aktuelle Diskussionen zum Thema Datenschutz und Datensicherheit, die nicht nur Big Data betreffen, sondern auch Themen wie Cloud Computing. Die Thematik wird jedoch z.B. von Firmen, Unternehmen und Banken, wo z.B. die Bankkonten und somit die Gelder der Kunden eine erhebliche Rolle spielen, sehr ernst genommen

und ist schon allein aufgrund des Geschäftsmodells ein Muss das praktiziert wird. Das bestätigen die Ergebnisse des BARC-Institutes, so schätzen die Unternehmen mit gerade einmal 25 Prozent dieses Risiko für eher gering ein (179, S. 26-27). Das Ergebnis des BARC-Instituts berücksichtigt dabei nicht den aktuellen Diskurs im Sommer 2013. Dass digitalisierte Daten schnell übertragbar, vervielfältigbar, verbreitbar, bearbeitbar und verwertbar sind, sind Gründe für eine entsprechende Debatte über Datensicherheit und Datenschutz. Diese ist generell zu führen und wird von Einrichtungen wie der European Union Agency for Network and Information Security (ENISA) bereits ausführlich und intensiv behandelt.

Weitere Details lassen sich aus den Charakteristiken der verwendeten Daten und der Datentypen ableiten. Chen et al. unterteilen in die folgenden fünf Branchen: E-commerce & Market Intelligence, E-Government & Politics 2.0, Science & Technology, Smart Health & Wellbeing und Security & Public Safety (181, S. 1172-1173).

Für die Analyse der Strengths (Stärken), Weaknesses (Schwächen), Opportunities (Chancen) und Threats (Risiken), kurz: SWOT-Analyse wurde zunächst eine Literaturrecherche durchgeführt, um anschließend eine Bewertung vornehmen zu können. Danach werden einzelne ausgewählte Branchen untersucht. Dabei lassen sich aus der Literatur unterschiedliche Stärken und Schwächen, sowie Chancen und Risiken ableiten.

In Tabelle 10 sind die Ergebnisse der SWOT-Analyse zu Big Data im Allgemeinen dargestellt. Diese Ergebnisse basieren auf den zuvor erörterten Punkten.

		Interne Analyse	
SWOT-Analyse		Stärken	Schwächen
Externe Analyse	Chancen	Schnellere und genauere Auswertungen Bessere strategische Entscheidungen Verbesserte Steuerung operativer Prozesse Bessere Auswertung mittels BI Kostenreduktion Flexiblere Dienstleistungen, zielorientiertere Marketingaktionen Bessere Informationen zum innerbetrieblichen Risikomanagement Besseres Verständnis des Marktes Verbesserter Kundenservice	Monitoring führt zur Erhöhung der internen Kontrolle Reduktion von Datensätzen ließ Informationen unbeobachtet, Big Data schafft verlässlichere Daten Kundenindividuelle Lösungen (gläserne Kunde)
	Risiken	Distributive Daten Fraud Detection (Betrugserkennung) Unstrukturierte informelle Kommunikation Fehlendes fachliches und technischen Wissen	Datenschutz Datensicherheit Gesellschaftliche Akzeptanz Ethische Gründe Fehlende überzeugende Einsatzszenarien Technische Probleme Kosten

Tabelle 10: SWOT-Analyse zur Potentialbewertung von Big Data im Allgemeinen (Eigene Darstellung)

Eine repräsentative Umfrage von TNS zeigt überdurchschnittlich hohen Handlungsbedarf und Entwicklung bezogen auf Big Data in folgenden Branchen: Energie (61 Prozent), Kommunikation (61 Prozent), Banken (55 Prozent) und Dienstleistungssektor (54 Prozent) (182, S. 11). Der Durchschnitt in der Studie liegt bei 53 Prozent, somit weisen der öffentliche Sektor (50 Prozent) und Handel (45 Prozent) einen geringeren Handlungsbedarf auf (182, S. 11). Ferner fragte die Studie von TNS die Relevanz und den Investitionsbedarf von Big Data ab. Das größte Potential hinsichtlich der letztgenannten Dimensionen weisen unternehmensbezogene Dienstleister auf (182, S. 10). Des Weiteren ist ein erhöhter Bedarf im öffentlichen Sektor und bei Energie- und Wasserversorgern festgestellt worden (182, S. 10). Zusammenfassend weisen Versorgungs-, Dienstleistungs- und Kommunikationsunternehmen ein hohes Potential auf (182, S. 28).

Nach dem die Rahmenbedingungen näher betrachtet und mögliche Bereiche für die Potentialbewertung identifiziert wurden können nun einzelne ausgewählte Branchen genauer betrachtet werden. Den Branchen öffentlicher Sektor, Industrie 4.0, Gesundheitssektor/ Life Science, Marktforschung, (Social-) Media und Entertainment, Energie, Mobilitätsdienstleistungen sowie Risikomanagement und Versicherungswesen werden Praxisbeispiele für die Analyse zugeordnet. Der BITKOM hat 34 Praxisbeispiele untersucht, welche für die weitere Potentialbewertung näher betrachtet werden. Diese teilen sich wie in Abbildung 46 gezeigt auf die ausgewählten Branchen auf (177). In Abbildung 46 ist die Aufteilung der jeweiligen Unternehmen nach ihren Branchen in Prozent dargestellt.

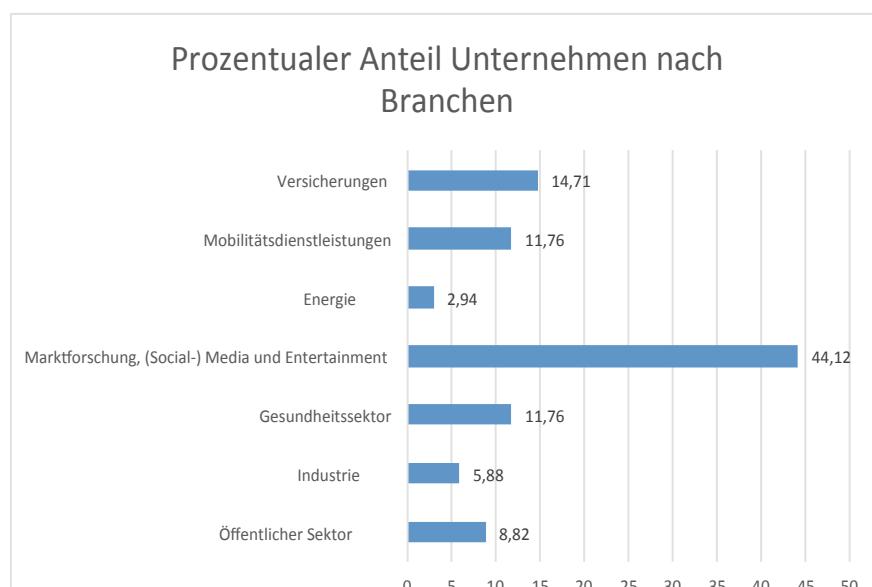


Abbildung 46: Aufteilung BITKOM Fallbeispiele auf ausgewählte Branchen (n=34)

Auffällig ist der Anteil an Unternehmen, welche Big Data-Technologie in der Marktforschung, (Social-) Media und Entertainment Branche besitzen. Dies verdeutlicht die Wichtigkeit dieser Technologie in der Marktforschung. Auf den Markt zugeschnittene Produkte und Dienstleistungen sind besonders wichtig, um gezielt auf die Wünsche des einzelnen Kunden eingehen zu können. Dies wird mittels Business Intelligence und Analyse bereits in ersten Anwendungen durchgeführt.

2.4 Branchenspezifische Innovationspotenziale

Nach dem das Potential branchenübergreifend erfasst ist, kann im Folgenden auf die als herausragend klassifizierten einzelnen Branchen eingegangen werden. Diese sind: öffentlicher Sektor, Industrie 4.0, Gesundheitssektor/Lifesciences, Marktforschung, (Social-) Media und Entertainment, Mobilitätsdienstleistungen, Energiewirtschaft sowie Risikomanagement und Versicherungswesen. Die einzelnen Branchen werden detailliert untersucht und mittels einer SWOT-Analyse betrachtet.

2.4.1 Öffentlicher Sektor

Der öffentliche Sektor bezeichnet „... in Deutschland Bund, Länder und Gemeinden (Gemeindeverbände) als Gebietskörperschaften sowie supranationale Instanzen (v.a. EU) und andere Organisationen“ (183). Als Beispiel für Big Data wird das Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ) herangezogen und näher betrachtet.

Das FDZ bereitet Individualdatensätze auf, die im Bereich der Sozialversicherung und in der Arbeitsmarkt- und Berufsforschung entstehen und stellt sie für wissenschaftliche Zwecke zur Verfügung. Das FDZ ist dabei Mittler zwischen den Datenproduzenten – insbesondere der Bundesagentur für Arbeit (BA) sowie den Forschungseinheiten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) – und den externen Datennutzerinnen und -nutzern. Zu diesem Zweck hat das FDZ transparente und standardisierte Zugangsregelungen unter Einhaltung der Datenschutzbestimmungen entwickelt. Es bereitet die Daten auf, aktualisiert, prüft und dokumentiert sie umfassend. Zudem berät es individuell über Zugang, Handling und Analysemöglichkeiten sowie über deren Reichweite und Gültigkeit. Das FDZ besteht seit neun Jahren und wurde gegründet um den Zugang zu den vorhandenen Mikrodaten der BA und des IAB für die Forschung im deutschsprachigen Raum zu verbessern. Seitdem verfolgt das FDZ zwei strategische Ziele: Den Zugang zu internationalisieren und das Angebot schrittweise auszuweiten. So wurde das Datenangebot durch die Kombination von Daten verschiedener

Datenproduzenten, wie dem Statistischen Bundesamt oder der Deutschen Rentenversicherung, erweitert. Der aktuelle Schwerpunkt der (Forschungs-)Arbeit liegt in der Weiterentwicklung dieser Ziele im Hinblick auf einen dezentralen, internationalen Zugang. Auch strebt das FDZ an, frei verfügbare Daten – wenn technisch und datenschutzrechtlich möglich – mit den Daten des FDZ zu verbinden (u.a. Big Data).

Die Datensätze des FDZ stehen der wissenschaftlichen Öffentlichkeit kostenfrei zur Verfügung. Ende 2012 stellte das FDZ insgesamt 15 Datensätze bereit, die über vier Zugangswege genutzt werden können: während eines Gastaufenthalts vor Ort, über kontrollierte Datenfernverarbeitung, als Scientific Use File (SUF) oder Campus File. Die vom FDZ angebotenen Datensätze unterscheiden sich von den „Standard“-Datensätzen in den Sozial- und Wirtschaftswissenschaften in zweierlei Hinsicht. Zum einen sind sie groß (beispielsweise beinhaltet der Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) über 1,6 Mio. Personendaten mit über 45 Mio. Datenzeilen), zum anderen komplex, da unterschiedliche Datenquellen und Untersuchungseinheiten zusammengespielt werden (beispielsweise Personen und Betriebe aus Befragungs- und/oder administrativen Daten).

Die Nutzung der FDZ-Daten haben sich seit einigen Jahren auf einem hohen Niveau von jährlich um die 130 neu genehmigten Projekten eingependelt. Die intensive Forschungsarbeit der Nutzerinnen und Nutzer mit den Daten des FDZ spiegelt sich in der Anzahl der Gastaufenthalte und der Nutzung der Datenfernverarbeitung wider. Beispielsweise wurde der Standort Nürnberg mit 505 Nutzungstagen im Jahr 2012 besucht. Die Zahl der Auswertungen mittels Datenfernverarbeitung stieg über die letzten Jahre und erreichte 2012 einen Stand von fast 1.700 Programmen, die das FDZ startete und auf Einhaltung der Datenschutzrichtlinien prüfte.

In dem vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projekt „FDZ im FDZ (PFI)“ wurden technische wie auch rechtliche Voraussetzungen geschaffen, wissenschaftliche Analysen mit FDZ-Daten von anderen Standorten aus über eine gesicherte Internetverbindung durchzuführen. Pilotstandort im Ausland ist das Institute for Social Research der University of Michigan in Ann Arbor (USA). In Kooperation mit den Forschungsdatenzentren der Statistischen Landesämter wurden im Laufe der Jahre 2011 und 2012 an den Standorten Berlin, Bremen, Düsseldorf und Dresden/Kamenz der Zugang durch Gastaufenthalte ermöglicht. Zwar ist noch immer noch der Standort Nürnberg der bevorzugte Ort an dem Gäste mit den Daten des FDZ arbeiten, allerdings kommen die fünf zusätzlichen Standorte mit 647 Nutzungstagen in der Summe auf eine höhere Zahl an Aufenthalten. Beeindruckend ist – neben der hohen Zahl an Gastaufenthalten in Düsseldorf – die große Resonanz des Standortes in Ann Arbor (327 Nutzungstage).

Eine ähnliche Zielsetzung verfolgt das EU-Drittmittelprojekt „Data without Boundaries (DwB)“. 28 Partner aus 12 europäischen Ländern arbeiten an einer europaweiten Verbesserung und Vereinheitlichung des Zugangs zu sensiblen Mikrodaten. Dazu zählt die Nutzung durch grenzüberschreitende Gastaufenthalte und über gesicherte Internetverbindungen ebenso wie die Vereinfachung von Genehmigungsprozessen. Das Projekt wird aus dem 7. Schwerpunktprogramm der EU mitfinanziert.

Ein mögliches Anwendungsszenario im öffentlichen Sektor könnten sog. Smart Cities sein. Batty et al. sehen in der Zukunft die IKT zu Big Data im Zusammenhang mit „Smart Cities“ als eines der Instrumente Analysen für Städte zu generieren. Besonders hervorgehoben wird die Mobilität, ein weitläufiges verfügbares WLAN, Global Positioning System (GPS) und Mobilfunknetzwerk als grundlegende Infrastruktur (184, S. 488). Dabei können die dabei erfassten Daten genutzt werden um die täglichen Aktivitäten und Bedürfnisse des Menschen in den Abläufen zu vereinfachen und zu beschleunigen. Ein Anwendungsfall ist z.B. der öffentliche Personennahverkehr (ÖPNV), der durch Big Data Technologie eine Verbesserung des Verkehrsmanagements ermöglicht (177, S. 69). Der Datenschutz, so die Autoren, bleibt durch eine Anonymisierung der Daten erhalten und spielt dabei eine untergeordnete Rolle (184, S. 488).

Eine Studie von TNS zeigt, dass der öffentliche Sektor einen überdurchschnittlichen Investitionsbedarf hat und Big Data als relevantes IKT-Zukunftsthema angesehen wird (182, s. 10).

Chen et al. zeigen im öffentlichen Sektor für Big Data Chancen für schnellere gegenwärtige Dienstleistungen von Seiten der Regierung sowie öffentliche Dienstleistungen auf (181). Darüber hinaus werden politische Kampagnen und elektronische Wahlen (e-polling) genannt (181).

Der §35 des Bundeswahlgesetz (BWG) und die Bundeswahlgeräteverordnung (BWO) sind die Rahmenbedingungen in Deutschland für den Einsatz von Wahlcomputern (185)(186). Jedoch sind Gesetz und Verordnung mit dem Urteil vom Bundesverfassungsgericht mit dem Aktenzeichen 2 BvC 3/07, 2 BvC 4/07 am 3. März 2009 für verfassungswidrig erklärt worden (187). Daher wird das e-polling in dieser Studie nicht als Stärke angesehen, sondern als eine Schwäche. Ferner nennen Chen et al. die Möglichkeit nach schnelleren gegenwärtigeren Dienstleistungen der Regierung (181). Ein Beispiel der Kungliga Tekniska Högskolan (KTH) zeigt dessen Potential auf, so war es nicht möglich die Vielfalt der Daten und in entsprechender Geschwindigkeit die Daten zur Verfügung zu stellen (177, S. 69). Nach der Einführung von Big Data war sowohl die Vielfalt als auch die Geschwindigkeit der Daten in entsprechend hoher Anzahl bzw. Zeit möglich (177, S.

69). Außerdem zeigen Chen et al. Chancen in der Gleichheit der Zugriffsmöglichkeiten und der öffentlichen Dienstleistungen auf (181).

Bryant et al. hingegen sehen beim Datenschutz und der Datensicherheit Schwierigkeiten und ein großes Risiko, da zum einen viele Daten und zum anderen mögliche sensible sowie sensitive Daten, die mittels Software schnell ausgewertet werden können und somit unautorisierte Zugriff und Benutzung von/auf Informationen gewährt werden könnte (188, S. 4). Als Beispiel führen sie die Überwachungskameras hinsichtlich ihrer Vernetzung an, welche heutzutage an vielen Orten angebracht sind, so z.B. in oder an Bankautomaten, Einkaufsläden, Flughafensicherheit und städtischen Straßenkreuzungen (188, S. 4). Die informelle Selbstbestimmung einer in der Bundesrepublik Deutschland lebenden Person muss mittels Pseudonymisierung und Anonymisierung, wie es das Bundesdatenschutzgesetz vorschreibt, erhalten bleiben.

In Tabelle 11 ist ein Überblick der zuvor in diesem Abschnitt erörterten Punkte als SWOT-Analyse dargestellt.

		Interne Analyse	
SWOT-Analyse		Stärken	Schwächen
Externe Analyse	Chancen	Reaktion/Abwicklung AllKostenreduktion Schnellere gegenwärtige Dienstleistungen der Regierung Öffentliche Dienstleistungen Politische Kampagnen Gleichheit der Zugriffsmöglichkeiten und öffentlichen Dienstleistungen	Kosten für Strukturierung der Daten Verteilte Informationsquellen Abstimmung von Bund, Ländern, Kommunen Datenvolumen Transparenz elektronische Wahlen (e-polling)
	Risiken	Distributive Daten Abstimmung von Bund, Ländern, Kommunen? Unstrukturierte informelle Kommunikation mit dem Bürger Stark textueller Inhalt	Datenschutz Datensicherheit

Tabelle 11: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche öffentlicher Sektor (Eigene Darstellung)

2.4.2 Industrie

Für den Industriesektor wird in dieser Studie stellvertretend Industrie 4.0 verwendet, da gerade mit dem sog. Internet der Dinge erhebliche Datenmengen entstehen.

Die Bundesregierung versteht unter Industrie 4.0 ein Zukunftsprojekt, das mit wichtigen technologie-, wirtschafts- und gesellschaftspolitischen Standortperspektiven verbunden ist. Durch das Zukunftsprojekt werden neue Perspektiven für industrielle Prozesse und industrienahen Dienstleistungen geschaffen (284).

Die „Intelligenz“ der Industrie 4.0 Systeme ergeben neben neuen Möglichkeiten eine erhebliche u.U. nicht beherrschbare Datenmenge. Für die Verarbeitung sind unterschiedliche Systeme notwendig, die in der Lage sind zeitnah, effizient, ortsflexibel, mit hoher Validität und Qualität Lösungen zu bieten. Die Systeme werden dadurch komplexer und benötigen neben der Hardware entsprechende Software. Dabei sind die Anforderungen an Systeme auch bedingt durch Cloud Computing eher dezentral ausgerichtet und benötigen eine lokale Steuerungsmöglichkeit. Die Möglichkeit der lokalen Steuerung ist besonders bei Mensch-Maschine-Interaktionen bezogen auf die Bedienbarkeit und Benutzbarkeit in der Entwicklung wichtig.

In der Entwicklung von Industrie 4.0 spielen sog. Cyber-Physische-Systeme eine entscheidende Rolle. Acatech schreibt dazu: „Gerade im Maschinen- und Anlagenbau beziehungsweise in der Automatisierungstechnik werden das Potenzial und die Herausforderungen von Cyber-Physischen-Systemen deutlich. Die sensorgestützte Vernetzung von intelligenten Maschinen und Produkten untereinander und mit übergreifenden Produktionsplanungs-, Energiemanagement- oder Lagersystemen – auch über Unternehmensgrenzen hinweg – ermöglicht Qualitäts-, Optimierungs- und Effizienzsteigerungen. Diese ergeben sich besonders aus flexiblen Anpassungen der Produktion an Kundenprozesse sowie der entsprechenden Steuerung global verteilter Logistikprozesse“ (189, S. 16).

Kagermann et al. ergänzen die Aussage von Acatech. Cyber-Physische-Systeme „... schaffen Smart Factories, der Inbegriff des Zukunftsprojekts Industrie 4.0. In der Smart Factory herrscht eine völlig neue Produktionslogik: Die Produkte sind eindeutig identifizierbar, jederzeit lokalisierbar und kennen ihre Historie, den aktuellen Zustand sowie alternative Wege zum Zielzustand. Die eingebetteten Produktionssysteme sind vertikal mit betriebswirtschaftlichen Prozessen in Fabriken und Unternehmen vernetzt und horizontal zu verteilen, in Echtzeit steuerbaren Wertschöpfungsnetzwerken verknüpft – von der Bestellung bis zur Lieferung. Gleichzeitig ermöglichen und erfordern sie ein durchgängiges Engineering über den gesamten Lebenszyklus eines Produkts einschließlich seines Produktionssystems hinweg“ (190).

Dies bedeutet jedoch, dass besonders durch die damit verbundene Vernetzung und Sensorik im Lebenszyklus eines Produktes eine nicht zu vernachlässigende Datenmenge und Komplexität entsteht. Insbesondere mangelt es an Referenzmodellen zur Umsetzung des Industrie 4.0 Ansatzes. Dies zeigt, dass an dieser Stelle weiterer Forschungsbedarf besteht. Um dies leisten zu können, benötigt

es einen Big Data-Technologieansatz, der bei der Umsetzung von Big Data in der Industrie 4.0 behilflich ist. Die neuen Geschäftsprozesse könnten z.B. durch KMU erbracht werden. Die deutsche Industrie wird im Hinblick auf die Wettbewerbsfähigkeit durch Industrie 4.0 entscheidend gestärkt. Insbesondere ermöglicht der Einsatz von Big Data-Technologie in der Industrie 4.0 eine Vielfalt von neuen Geschäftsprozessen, Dienstleistungen und Produkten.

Das Potential von Industrie 4.0 wird dabei an folgenden Punkten festgehalten:

- individuelle Kundenwünsche,
- Geschäfts- und Engineering-Prozesse dynamisch,
- durchgängige Transparenz,
- optimale Entscheidungen,
- neue Formen von Wertschöpfung und
- neuartige Geschäftsmodelle.

Insbesondere wird ein Vorteil für Start-ups und kleine Unternehmen in nachgelagerten Dienstleistungen gesehen. Neue Geschäftsperspektiven und –modelle werden laut der Studie auch eine Nutzung von Diensten und Software-Systemen für KMU ermöglichen (190, S. 26).

Dem gegenüber sind für die Einführung von Industrie 4.0 jedoch zunächst Investitionen in Form von Geld, Know-how und Forschung notwendig. Des Weiteren wird angeregt Themen wie Sicherheit, Vertrauen, Verlässlichkeit, Nutzung, Konvergenz der Betreibermodelle, Echtzeitanalyse und Prognose zu überdenken. Zudem kommt das Risiko der Bewerkstelligung von heterogenen Datenquellen und Endgeräten zu ermöglichen hinzu (190, S. 29). Des Weiteren ist i.d.R. eine starke Distributivität der Daten gegeben, dies führt bekanntermaßen zu Latenzen. Diese zu minimieren ist ein wichtiger Punkt für die global agierende deutsche Wirtschaft und ist aus diesem Grund ein weiteres Forschungsthema.

In der Tabelle 12 sind die zuvor erörterten Punkte zu Industrie 4.0 in einer SWOT-Analyse zusammengefasst dargestellt.

		Interne Analyse	
SWOT-Analyse		Stärken	Schwächen
Externe Analyse	Chancen	On-Demand Kostenreduktion Individuelle Kundenwünsche Geschäfts- und Engineering-Prozesse dynamisch Durchgängige Transparenz Optimale Entscheidungen Neue Formen von Wertschöpfung Neuartige Geschäftsmodelle	Investitionen
	Risiken	Stark distributive Daten Vertrauen Verlässlichkeit Nutzung Konvergenz der Betreibermodelle	Datenschutz Datensicherheit Latenzen

Tabelle 12: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Industrie 4.0 (Eigene Darstellung)

2.4.3 Gesundheitssektor

Der Gesundheitssektor steht hier für das Gesundheitswesen in Deutschland. In einem ersten Schritt gehen wir kurz auf die Bedeutung sowie ein Anwendungsbeispiel ein und führen in einem zweiten Schritt eine SWOT-Analyse durch.

Das Gabler Wirtschaftslexikon definiert den Begriff wie folgt: „Das Gesundheitswesen ist differenziert gegliedert. Man unterscheidet die ambulante und statio-näre Leistungserbringung durch niedergelassene Ärzte und Zahnärzte, Krankenhäuser sowie sonstige Leistungserbringer; einen eigenen Bereich stellt die Arzneimittelversorgung dar. Die Finanzierung dieser Leistungen erfolgt im We-sentlichen durch die Krankenversicherungen, die wiederum in die gesetzliche und in die private Krankenversicherung unterteilt sind. Daneben gibt es seit 1995 die Pflegeversicherung“ (191).

Anwendungsbeispiel: Informationsbasierte Medizin. Die in der Medizin erfassten Datenmengen steigen exponentiell; begleitend zeichnet sich ein Para-digma-Wechsel in der medizinischen Entscheidungsfindung ab von der Diagno-se- und Therapieentscheidung durch einen einzelnen Mediziner auf Basis einer recht beschränkten Datenmenge hin zu Hochdurchsatzverfahren beruhend auf nichttrivialen Berechnungen und Datenauswertungen. Im Bereich der Diagnos-tik werden vermehrt omics-Daten aus Hochdurchsatzverfahren für viele Patien-ten ermittelt, um dann mittels Klassifikationsalgorithmen Diagnosewerkzeuge mit hoher statistischer Aussagekraft zu entwickeln (90). Bei der Therapieplanung werden bildgebende Verfahren herangezogen, um die Ana-tomie eines individuellen Patienten zu rekonstruieren und den chirurgischen

Eingriff optimal zu planen. Dabei erfolgt die medizinische Bildverarbeitung in einer Datenverarbeitungspipeline von „region of interest detection“ und Registrierung (109) über Segmentierung und Anatomierekonstruktion (88) (162) bis zur nachfolgenden Klassifikation. Die automatische Realisierung dieser Verarbeitungsschritte ist hochkomplex und fehleranfällig und daher oft der Grund für suboptimale diagnostische oder therapeutische Ergebnisse. Algorithmen, die direkt einstufig verfahren, sind bislang noch wenig verbreitet (83). Wesentliche Anforderungen an solche Algorithmen sind nicht nur hohe Klassifikationssicherheit (hohe Sensitivität bei gleichzeitig hoher Spezifität) sondern neben Skalierbarkeit und hoher Performanz auch hohe Transparenz, d.h. die Fähigkeit des Algorithmus eigene Fehler zu erkennen und die geeignete Aufbereitung des Klassifikationsergebnisses. Hier sind noch wesentliche Forschungsschritte zu leisten.

Ein weiteres Beispiel liefert die Diversität in der Biologie, die es oft erfordert schnelle Auswertungsergebnisse bzgl. neuer Krankheiten zu erzielen. Dazu ist es u.a. erforderlich die Wirkungsweise von Medikamenten simulieren zu können. Die European Screening Port GmbH ist eine Schnittstelle zwischen Forschung und Praxis.

Zur Untersuchung der komplexen Phänomene in der Medizin bietet Big Data-Technologie neue Möglichkeiten der Analyse und der Untersuchung solcher Phänomene. Dabei kann Big Data zusätzlich einen Ansatz liefern, der es ermöglicht den Patienten als Individuum und nicht als Durchschnittsgruppe von ähnlichen Patienten zu untersuchen. Dies bedeutet in der Medizin völlig neue Erkenntnisse, bessere medizinische Behandlungsmethoden und eine daraus ggf. resultierende Kostensenkung im Gesundheitswesen. Pietsch nennt dabei als eine offene Frage, wie in Zukunft Diagnosen erfolgen. Dies könnte sowohl auf reiner Rechnerbasis und automatisiert sein, semiautomatisiert oder klassisch mit Rechnerunterstützung. Für diese Untersuchungen im Zusammenhang mit Big Data in der Medizin wurden an den Standorten der Oxford University in Großbritannien und der Stanford University in den USA in den vergangenen Jahren bereits Forschungszentren gegründet. Dabei hat das Big Data Institute Forschungszentrum in Oxford gerade im Mai diesen Jahres eine Förderung in Höhe von 30 Millionen Pfund (ca. 35 Millionen Euro¹) erhalten. Das Institut wird dabei den Fokus auf große anonymisierte medizinische Daten, wie DANN-Sequenzen, Patientendaten, nationale Registrierung und klinische Proben legen. Dies dient dem Zweck der besseren Erkennung, ärztlichen Versorgung (Heilbehandlung) und Prävention, so Times Higher Education. Das Li Ka Shing Center for Learning and Knowledge an der Stanford University wird der nächste Austragungsort für die Konferenz „BIG DATA in biomedicine“ sein (302). Ein weiteres Projekt im Ge-

¹ Stand 19.07.2013

sundheitswesen ist das Human Brain Project aus der Europäischen Union und wird mit 1,190 Milliarden Euro gefördert. Partner des Projektes sind: Ecole Polytechnique Fédérale de Lausanne (CH), Ruprecht-Karls-Universität Heidelberg (DE), Forschungszentrum Jülich GmbH (DE), Centre Hospitalier Universitaire Vaudois (CH), Karolinska Institutet (SE), Universidad Politecnica de Madrid (ES), Wellcome Trust Sanger Institute, Genome Research Limited – Genes to Cognition (UK), Technische Universität München - Fortiss GmbH (DE), Interuniversity Microelectronics Centre (IMEC) (BE), Hebrew University of Jerusalem (IL), Institut Pasteur (FR), Innsbruck Medical University (AT), Commissariat à l'énergie atomique et aux énergies alternatives (FR). In dem Projekt werden Neurowissenschaften, Medizin und wissenschaftliches Rechnen zu einer völlig neuen IKT verknüpft. Deutschland kann hier aufschließen, indem es seine hervorragend aufgestellten Disziplinen der Biologie, Medizin respektive Medizintechnik unter dem Schwerpunkt der Informatik und Wirtschaftsinformatik interdisziplinär in einem Kompetenzzentrum verknüpft und vereint.

Chen et al. führen zusätzlich ethische Gründe als Risiko für eine Umsetzung an (181). Darüber hinaus sind Patientendaten bekanntermaßen per Gesetz besonders zu schützen.

SWOT-Analyse		Interne Analyse	
Externe Analyse	Chancen	Stärken	Schwächen
		Kostensenkung im Gesundheitswesen Genaue Analysen Neue Dienstleistungen durch z.B. Apps, Augmented Reality Bessere Erkennung von Krankheiten Bessere Ärztliche Versorgung Prävention Genanalysierung und –visualisierung Langfristigere Behandlung Individuelle Patientenbehandlung Wissen über Risikofaktoren Nebenwirkungen Reaktionen	Social Media-Patientenplattformen analysieren Medikamentnebenwirkungen analysieren Patienten mehr einbinden/ mehr Verantwortung übertragen
	Risiken	Personenspezifischer Inhalt Verschiedener aber eng verwandter Inhalt	Datenschutz Datensicherheit Ethische Gründe

Tabelle 13: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Gesundheitssektor (Eigene Darstellung)

2.4.4 Marktforschung, (Social-) Media und Entertainment

Zunächst geben wir einen Überblick, was unter der gewählten Branche zu verstehen ist. In einem nächsten Schritt gehen wir auf zwei Anwendungsbeispiele

genauer ein. Abschließend werden die Vor- und Nachteile diskutiert und mittels einer SWOT-Analyse zusammengefasst und bewertet.

Bei vielen alltäglichen Fragen, die Menschen stellen, ist heute das Internet die erste Anlaufstelle. In unzähligen Foren beantworten private Personen Fragen von anderen. Dabei entsteht eine große Menge an Wissen, wobei aber auch beachtet werden muss, dass das auf diese Art geteilte Wissen nicht Expertenwissen ist. Dieses Wissen liegt in nicht strukturierter Form vor (177, S. 76). Jedoch bietet eine Analyse der Social Media Plattformen ein großes Potential, um z.B. häufig auftretende Nebenwirkungen eines Medikaments schneller zu bemerken (298) (299) (300) oder um Marktforschung zu betreiben und neue Trends frühzeitig erkennen zu können (177, S.76).

Zwei Anwendungsbeispiele sind die Videoannotation und die Textanalytik.

Das Anwendungsbeispiel der Videoannotation zeigt, dass Datenbanken mit Medieninhalten wie Video- und Audiosignalen, sowie potentiell zugehörige Metadaten, sich in einem rasanten Wachstum mit vielen Petabytes an Daten befinden. Dabei können die Inhalte sowohl von professionellen Produzenten (z.B. Video on Demand-Portale) als auch von Nutzern stammen (z.B. nutzergenerierte Inhalte, soziale Netze). Für eine effiziente Speicherung und Übertragung werden die Medieninhalte nach dem Hochladen in die Datenbank komprimiert und in einem oder mehreren vom Mediaplayer unterstützten Formaten zur Verfügung gestellt. Die Annotation der Mediendaten erfolgt anhand von Metadaten, welche durch den Nutzer zur Verfügung gestellt werden oder durch den Kontext, welcher bei Uploads in sozialen Netzen potentiell verfügbar ist (125). Die inhaltliche Analyse von Video- und Audiosignalen erfolgt üblicherweise nach einer Vorfilterung durch Metadaten. Wichtige Analysemethoden sind u.a. lineare Klassifikation, Hashing und graphbasierte Methoden (78). Dabei ist die Ausnutzung tiefer multimodaler Korrelationen eine große Chance, weil das Ableiten von zusätzlichen Metadaten zu stark verbesserten Resultaten führen kann (86). Aufgrund der riesigen Datenmengen ist die Entwicklung von Methoden, die Daten in codierter Form klassifizieren können, ein kritischer Erfolgsfaktor für die breite wissenschaftliche und wirtschaftliche Anwendung von Videoanalysen (97).

Ein weiteres Anwendungsbeispiel ist die Textanalytik. Ziele der Textanalytik sind die Extraktion und Strukturierung von potentiell relevanter Information aus der ständig wachsenden Menge digitaler Texte, die aus dem offenen Web oder in Unternehmen oder öffentlichen Einrichtungen verfügbar sind. Sie ermöglicht die Erschließung, Navigation und Suche textueller Daten und ihre Verknüpfung mit anderen Daten- und Informationsquellen. Textanalytik besteht aus Methoden und Verfahren der Computerlinguistik und Informatik für die automatische Erkennung von sprachlichen Einheiten oder Mustern in freien Texten und ihre Anwendung bei der Gewinnung von strukturierter Information und Wissen.

Dafür werden maschinelle Lernverfahren insbesondere statistische Methoden eingesetzt (124) (153). Zur Textanalytik gehören u.a. die linguistische Analyse (z.B. Tokenisierung und Wortartenerkennung, sowie die morphologische, syntaktische und semantische Analyse), Eigennamenerkennung, Relationsextraktion, Koreferenz-Auflösung und Sentiment-Analyse. Viele webbasierte Informationsextraktionssysteme können schon heute einen brauchbaren "Recall" aufgrund von Datenredundanz und großer Datenabdeckung erzielen, leiden aber dennoch unter unzureichender Präzision wegen der unkontrollierbaren Datenqualität im Web, Unzulänglichkeiten der Analyse und Fehlern beim automatischen Lernen (66, 65, 113, 107). Der Verbesserung der "Precision" kommt daher eine hohe Priorität in der Skalierung der Sprachtechnologie auf Big Data zu.

Big Data Technologie ermöglicht es soziale Medien auf neue Erkenntnisse hin zu durchsuchen und direkt mittels Big Data Werkzeugen zu analysieren (181) (192). Mit dem so gewonnenen Wissen ist ein langfristigeres Marketing möglich, da personalisierte Empfehlungen für ausgesuchte Gruppen oder Lösungen für einzelne Kunden erstellt werden können (181). Dies steigert die Kundenzufriedenheit und den Umsatz des Unternehmens (181). Ein Hindernis, das dieser Nutzung jedoch im Weg steht ist der unstrukturierte Inhalt im Internet, der in Form und Genauigkeit von Website zu Website variiert (181). Hinzu kommt, dass die inoffiziellen Nutzermeinungen eine korrekte Analyse erschweren. Wie bereits bei den vorherigen Branchen erwähnt liegen die Gefahren beim Datenschutz und bei der Datensicherheit, da ein nachlässiger Umgang mit diesen Themen die Nutzer von der Partizipation abhalten könnte (179)(181). Facebook, Twitter und weitere soziale Medien reagierten mit der Möglichkeit entsprechende Einstellungen hinsichtlich der Privatsphäre vornehmen zu können. Die neuen Dienstleistungen und Produkte müssen eine gesellschaftliche Akzeptanz finden, die mit entsprechenden Vorkehrungen versehen sind, wie z.B. Anonymisierung, Pseudonymisierung, Verschlüsselung. Folglich ist es notwendig, dass ein Risikomanagement in den Unternehmen auch das IT-Risikomanagement der Daten beinhaltet.

Ein weiteres Beispiel ist der Bereich, welcher sich um den Fernseher als Bildwiedergabemedium gebildet hat. Die Fernsehgeräte sind heute inzwischen interaktiv und internettauglich. So experimentiert ProSieben als Teil der ProSieben-Sat1 Medien-Gruppe mit Augmented-Reality-Fernsehen und verbindet auf diese Weise das Smartphone mit dem Fernsehprogramm. Daneben entstehen ganz neue Möglichkeiten und Dienstleistungen, da es mittels der sog. Apps möglich ist nicht nur genauere Kundendaten zu erfahren sondern auch Zugriff auf Kontakte zu bekommen. Auf diese Weise kann das Fernsehprogramm an Mehrheitsfähigkeiten angepasst werden. Es wird unmittelbar sichtbar, sowohl für den Fernsehsender, als auch für den Zuschauer was „in“ und „out“ ist. Informationen könnten dann speziell für den Zuschauer zugeschnitten werden. Diese Möglichkeit des zweiten Bildschirms ist ebenso in anderen Bereichen denkbar, wo

Personen interaktiv in Echtzeit unterschiedliche Informationen je nach Rolle der Person zugespielt bekommen. Am Beispiel Fernsehen, bekommt der Zuschauer z.B. Fragen gestellt, während der Programmchef die aktuellen Einschaltquoten bekommt. Es ist in diesem Zusammenhang mittels geeigneter sowie zu entwickelnden Schutzverfahren darauf zu achten, dass der Zuschauer bzw. Fernsehprogramm-Konsument nicht zum „gläsernen Kunden“ wird.

In der Tabelle 14 sind die in diesem Abschnitt beschriebenen Punkte mittels einer SWOT-Analyse dargestellt.

		Interne Analyse	
SWOT-Analyse		Stärken	Schwächen
Externe Analyse	Chancen	Schnellere BI-Ergebnisse Neue BI-Ergebnisse ermöglichen Flexible Reaktionen auf Marktumfeld Social Media auf neue Erkenntnisse hin durchsuchen und analysieren	Kundenindividuelle Lösungen (gläserne Kunde) Langfristiges Marketing und gezielte, personalisierte Empfehlungen Reichhaltigen Informationsschatz nutzen
	Risiken	Mehrsprachiger Inhalt Unstrukturierte, informelle Kundenmeinungen	Datenschutz Datensicherheit

Tabelle 14: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Marktfor- schung, (Social-) Media und Entertainment (Eigene Darstellung)

2.4.5 Mobilitätsdienstleistungen

Mobilitätsdienstleistungen werden auch in Zukunft eine entscheidende Komponente in unserer Gesellschaft. Im Hinblick auf Elektromobilität bietet Big Data neue Perspektiven und neue Möglichkeiten für den Endkunden, das Gewerbe und die Industrie selbst. Für die Nutzbarmachung sind verschiedene Herausforderungen zu lösen. Diese sind neben technischer, ökonomischer auch gesellschaftlicher Natur. IKT nehmen für die Erzielung neuer nachhaltiger Dienstleistungen, neuer Mobilitätskonzepte und neuer Geschäftsmodelle die entscheidende Rolle als Unterstützer und Beförderer ein. VW bezeichnet als Konzern die Elektromobilität als eine Schlüsseltechnologie, ebenso setzt die Bundesregierung auf politischer Ebene auf Elektromobilität. Sie wird daher im Folgenden genauer betrachtet. Dabei bildet das Regierungsprogramm zur Elektromobilität die Rahmenbedingungen für die Nutzung dieser Technologie. Die Regierung will die Technologie soweit entwickeln, dass im Jahre 2020 der Massenmarkt mit einer Million E-Fahrzeugen beginnen kann.

Zusätzlich zur E-Mobilität ist es bedeutsam, dass die Energie, welche hierfür notwendig ist ebenfalls aus erneuerbaren Energie besteht. Deutschland kann in dieser Kombination in Europa die Marktführerschaft einnehmen.

Toll Collect sorgt dafür, dass die Maut auf 1:1.000 Fehler genau berechnet wird (177, S. 91). Das System von TomTom ermöglicht die Verarbeitung der gesendeten Nutzerdaten, so ist z.B. ein Flottenmanagement in Echtzeit oder eine schnellere Ankunft durch Echtzeit-Verzögerungsinformationen realisierbar (177, S. 77) (193). Dieses sog. „HD-Traffic“ führt neben einer Verkürzung der Fahrtzeit bei Stauumfahrung auch oft zu einem geringeren Spritverbrauch. Das zeigen Praxisbeispiele (194, S. 12-17). HD-Traffic ermöglicht es Daten an das Fahrzeug zu senden, weitere Möglichkeiten wäre es Fahrzeugdaten an einen Server zu übermitteln, um so weitere Dienstleistungen anbieten zu können (195, S. 666). Durch Big Data werden die für den Endkunden angebotenen Dienstleistungen vereinfacht, so bietet z.B. die Deutsche Bahn die Dienstleistung „Touch and travel“ an, bei der die reisende Person lediglich den Start- und Endpunkt vermerken muss; die Berechnung des Preises erfolgt automatisiert. Die Tarifauswahl durch den Kunden ist somit nicht weiter notwendig. In Hongkong sind nicht nur öffentliche Verkehrsmittel sondern auch viele Geschäfte mit der Octopus Card bezahlbar.

Für Innovationen wie E-Fahrzeuge fehlen neben anderen Herausforderungen noch die passende Ladeinfrastruktur um das E-Fahrzeug alltagstauglich zu machen. Für das Erreichen des Ziels eines emissionsfreien Fahrzeugmarktes ist die Herkunft des Stroms im Hinblick auf erneuerbare Energien wichtig. Steckdosen sind hingegen fast überall zu finden. Zudem muss das E-Fahrzeug wettbewerbsfähig werden, momentan sind die Preise für die Batterien der teure Faktor. Ein Ansatz zur Erhöhung der Nutzung von Fahrzeugen insbesondere E-Fahrzeugen ist Car-Sharing. Car-Sharing wird von immer mehr Personen benutzt (270.000 Fahrberechtigte, 22,7 Prozent mehr als 2012), auf ein Car-Sharing-Fahrzeug kommen durchschnittlich 40 Kunden. Es gibt 6.700 Fahrzeuge (196).

Durch die umfangreichen Umstrukturierungen, die die Elektromobilität notwendig macht, stellen sich nicht nur neue Herausforderungen an das Produkt Fahrzeug selbst, sondern ebenso an die innerbetrieblichen Prozesse hinsichtlich ihrer Fertigungsmethoden bis zur Mitarbeiterschulung. Außerdem sind in der Beschaffung und dem Vertrieb neue Ansätze und evtl. neue Geschäftsmodelle gefragt. Für die genauen Einflüsse von Big Data ist der Unterabschnitt Industrie 4.0 zu beachten.

In der Tabelle 15 sind die Potentialbewertungen dargestellt.

SWOT-Analyse		Interne Analyse	
Externe Analyse	Chancen	Stärken	Schwächen
	Risiken		
		Genauere und vielfältigere Angebote Weniger Stau Höherer Komfort durch mehr Service	Kundenindividuelle Lösungen (gläserne Kunde) Infrastruktur (Lademöglichkeiten, Car-Sharing)
		Genauere Betrugserkennung	Datenschutz Datensicherheit

Tabelle 15: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Mobilitätsdienstleistungen (Eigene Darstellung)

2.4.6 Energie

Zur Energiebranche gehören Unternehmen, welche aus unterschiedlichen Resourcen Energie erzeugen und diese über Energienetze den Verbrauchern anbieten. „Zur Energiebranche gehören alle Produzenten und Händler der Energieträger Erdöl, Gas, Kohle, Nuklearenergie und regenerativer Energie“.

Für die Energiebranche ist die Nachhaltigkeit eine wichtige Begrifflichkeit. Für diese Untersuchung wird die Definition der UN – auch als Brundtland Report bekannt – herangezogen (303). Das bedeutet, dass insbesondere die Akzeptanz in der deutschen Gesellschaft von Big Data-Produkten eine entscheidende Rolle spielt. Ohne die soziale Komponente, des Vertrauens in die Sammlung von Daten über Personen ist die Vermarktung von neuen Produkten und Dienstleistungen in diesem Bereich schwierig. Zusätzlich ist es notwendig, dass die Wirtschaftlichkeit neben dem Umweltaspekt gegeben ist. Das bedeutet, wenn wir in Deutschland und Europa erneuerbare Energien für E-Fahrzeuge nutzen, dann ist dies eine zunächst umweltgerechte Lösung. In China im Jahre 2011, wo hingegen 78,2 Prozent des elektrischen Stroms aus Steinkohle gewonnen wird erzeugt ein Golf mit einem 85 kW E-Motor im Vergleich zum Strommix der EU mit 96g/km mehr an CO₂. Wäre der Strom aus Windenergie erzeugt worden, so würde die Belastung bei 1g/km liegen, das entspräche 99,5 Prozent weniger CO₂-Belastung pro Kilometer im Vergleich zum Strommix aus China im Jahre 2011.

- Verschiedene Formen elektrische Leistung, Wärme, Kühlung, Wasserkraft, Wind, etc.
- Verbrauch und Speicherung von Energie
- Prognoseerstellung
- Datenaufbereitung
- Simulation

Energie ist eine immer teurer werdende Ressource, um Kosten einzusparen bzw. stabil zu halten ist ein geringer Verbrauch notwendig. Besonders die stark steigende Anzahl an intelligenten Stromzählern, welche in schnellen Takten Energieverbrauchsdaten übermitteln macht den Energiekonzernen Probleme (197). Die Versorger sind dabei auf die Auswertung dieser Messdaten noch nicht vorbereitet (197). Mit Big Data-Analysen wird die Abrechnung der Energieversorger hinsichtlich des Energieverbrauchs verbessert und für den Kunden transparenter (197) (198). Maschinelles Lernen in Form von Daten-Analysen erlaubt bessere, genauere und valide Verbrauchsprognosen (197). Die Energieversorger steuern dadurch den Einkauf und dessen Produktion von Energie genauer als es bisher möglich war (197). Im Weiteren ermöglicht dabei Big Data kundenspezifischer Dienstleistungen zu gestalten. Es ließen sich so z.B. Tarife besser auf verschiedene Kundengruppen zuschneiden oder unzufriedene Kunden leichter erkennen und somit die Kundenbindung erhöhen (197). Dabei werden die Energieversorger zusätzlich zu den bisher strukturierten Daten, unstrukturierte Daten aus dem Social Media-Bereich heranziehen (199).

Die Energieversorger werden somit verbrauchsorientierter in dem zur richtigen Zeit an den richtigen Ort in einer flexiblen Struktur die benötigte Energie liefert wird. Die Neuentwicklung von Geschäftsmodellen, die diesen ökologischen Anforderungen und diesem mit bidirektionalen Produkten gerecht wird (197, S. 2). Des Weiteren entstehen neue Märkte in denen IT-Dienstleistungen eine entscheidende Rolle spielen (197, S. 2).

Ein Anwendungsbeispiel findet sich in den Materialwissenschaften. Die Methoden zur Entwicklung neuer Materialien für die Grundlagenforschung und neue technologische Anwendungen haben sich in den letzten Jahren um eine vielversprechende Richtung erweitert. Im Sommer 2011 kündigte der Präsident der Vereinigten Staaten von Amerika die „Materials Genome Initiative for Global Competitiveness“ an (129). Hiermit ist die Entwicklung neuer Materialien gemeint, die z.B. für bessere Batterien, bessere oder neuartige Katalysatoren, Solarzellen, etc. dringend benötigt werden. Die Vorarbeiten, die hinter dem Konzept der Obama-Initiative stehen, wurden im Wesentlichen in Europa geleistet (z.B. Computer Codes für Materialien) (132). Bei der Anwendung liegen die USA jetzt aber vorn, was - wie der Titel der Initiative aussagt - ihre Wettbewerbsfähigkeit im Bereich von Grundlagenforschung und Ingenieurwissenschaften wesentlich verbessern kann. Da praktisch alle Förderorganisationen von Wissenschaft, Energie, Technologie, Sicherheit und Verteidigung der USA die „Materials Genome Initiative“ seit dem Winter 2012/2013 aktiv unterstützen, wachsen die weltweit verfügbaren Datenmengen zu Materialien (z. B. (112), (75)) mit einer rasanten Geschwindigkeit. Es existieren bereits Daten für mehrere zehntausend Materialien, und diese Zahl verdoppelt sich jedes Jahr.

Bislang wurden diese Material-Datenbanken in Wesentlichen verwendet, um die darin gespeicherten Daten bei Bedarf abzufragen (z.B. „High Throughput Screening“). Es ist nun aber dringend notwendig den nächsten Schritt zu forcieren und die digitale Verfügbarkeit der Daten zu nutzen, um Korrelationen, versteckte Trends, bislang unerkannte Kausalitäten und spezielle Materialeigenschaften zu identifizieren und zu verstehen (137) (77). Die intellektuelle Herausforderung in diesem Bereich ist immens und erfordert primär eine solide Modellbildung zur Identifizierung von physikalischen Deskriptoren. Für einen breiteren Einsatz solcher Methoden, die auch nach versteckten Korrelationen und Informationen suchen, werden sich die große Heterogenität und der Umfang Daten (10GB-10TB pro Material) zunehmend hemmend auswirken. Neue Methoden der Indizierung und eine schnellere Verarbeitung (Kommunikation, Parallelisierung, etc.) werden deshalb zukünftig deutlich an Bedeutung gewinnen.

In der Tabelle 16 ist die Potentialbewertung mittels einer SWOT-Analyse, welche aus den in diesem Abschnitt erläuterten Punkten hervorgeht zusammengefasst dargestellt.

SWOT-Analyse		Interne Analyse	
Externe Analyse	Chancen	Stärken	Schwächen
	Risiken	Prognose Qualitätssteigerung Auswertungserstellung Höhere Vielfalt Kostenreduktion Kundenspezifischere Dienstleistungen	Von statischer zu dynamischer Steuerung
		Einsparpotential kann negativ sein oder zu einer erhöhten Nachhaltigkeit führen	Gesellschaftliche Akzeptanz Latenzen

Tabelle 16: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Energie (Eigene Darstellung)

2.4.7 Versicherungen

Zur Branche der Versicherungen zählt in dieser Studie das Risikomanagement und Versicherungswesen sowie Unternehmen, die das Risiko bewerten und zu minimieren beabsichtigen. Dies kann z.B. durch einen Risikoausgleich erfolgen. Eine Versicherung ist nach dem Gabler Wirtschaftslexikon „eine zur Deckung eines im Einzelnen ungewissen, insgesamt geschätzten Mittelbedarfs auf der Grundlage des Risikoausgleichs im Kollektiv und in der Zeit“.

Probleme können von rechtlichen Auswirkungen bei fehlendem Datenschutz bis hin zu gegnerischem Data-Mining reichen, bei dem wir herausfinden möchten

auf welche Weise unsere Kontrahenten auf Teile des Internets abzielen und wo nach sie suchen wie z.B. Honeypots. Eine große Frage ist wie man unbekannte Teile des Internets abdecken kann und woraus man wertvolle Informationen über Aktivitäten der Internetkriminalität bekommen könnte.

Mehrere Vorgehensweisen wurden mithilfe von Data-Mining entwickelt um der Internetkriminalität vorzubeugen. Das AntiPhish-Projekt (<http://www.antiphish-research.org/home.html>) hatte als Ziel, Phishing-Attacken wie die Aufforderung an den Empfänger einer offiziell aussehenden E-Mail vertrauliche Informationen wie Kontodaten einzugeben, zu stoppen. Die in AntiPhish entwickelten Filter teilen E-Mail Nachrichten in die Kategorien Phishing, Spam oder Berechtigt unter Verwendung verschiedener content features ein (175). Die Angreifer entwickelten stets neue Arten von Phishing Nachrichten um diese Filter zu umgehen. Mehrere gegnerische Data-Mining Ansätze wurden verwendet um mit diesen Situationen umzugehen. Richard Weber und seine Kollegen benutzen die Theorie des Signalspiels um Strategien zu entwickeln (183).

In AntiPhish haben wir vorher unbekannte Verschleierungstricks (hidden salting) wie weiße Buchstaben auf weißem Hintergrund entdeckt (176).

Bei einer Befragung von The Data Warehouse Institut (TDWI) Research gaben 33 Prozent an, dass Big Data bessere Betrugserkennung für ihr Unternehmen ermöglichen würde (200, S. 11). Für die Betrugserkennung kann Big Data einen wesentlichen Beitrag leisten. Mit Hilfe von Big Data-Technologie ist es möglich Betrugsmuster zu erkennen. Hierzu werden, wie ein Praxisbeispiel aus der BIT-KOM Studie zeigt, die täglich anfallenden Transaktionen mit Mustern aus älteren Transaktionen verglichen und Betrugsfälle können schneller, genauer und verlässlicher erkannt werden (177, S. 78).

Mit den Geodaten der Kunden von Mobilfunkanbietern wie z.B. Deutsche Telekom, Vodafone, O2, etc. ergibt sich die Möglichkeit die Kreditkartensicherheit zu erhöhen. Ist der Kreditkartenkunde örtlich stark von seinem Smartphone getrennt, z.B. befindet sich sein Smartphone in München und seine Kreditkarte in einem anderen Land, so kann dies auf einen Betrugsfall hindeuten. Bei der Kreditkartensicherheit können durch Big-Data-Analysen die Transaktionen anhand von Mustern untersucht und so Betrugsversuche erkannt werden (177, S. 78).

Des Weiteren können unter zur Hilfenahme von sog. Social-Network-Analysis kriminelle Netzwerke entdeckt und analysiert werden. Durch die Einbindung von Social Media in die Analyse können Betrugsfälle deutlich früher bemerkt werden. Die meist unstrukturierten Daten aus den sozialen Netzwerken werden bis jetzt selten von Unternehmen genutzt, obwohl die Unstrukturiertheit der Daten für moderne Programme kein Problem darstellt und viele Details für eine verlässlichere sowie beständiger Prognose gebraucht werden (200, S. 9 und S. 27). Unstrukturierte Daten, vor allem geschriebene Sprache, nehmen zu so

gaben 35 Prozent der Unternehmen eine Nutzung der Daten an (200, S. 18). Versicherungsunternehmen filtern z.B. aus Schadensanträgen strukturierte Fakten, welche sie wiederum für die Analyse des Risiko-Managements und Betrugs einsetzen (200, S. 27).

Neue Dienstleistungen, welche diese unstrukturierten Daten auswerten können sind zum einen weiterer Gegenstand der Forschung und zum anderen helfen sie die Interessen der Gesellschaft (z.B. Schutz) zu wahren.

Im Bereich des Versicherungswesens ist die Zuordnung des zu managenden Risikos entscheidend. Nach Hansen und Neumann umfasst „das Risikomanagement [...] eine große Menge von Tätigkeiten, die dazu beitragen sollen, Risiken zu erkennen, in ihrem Ausmaß abzuschätzen und deren Folgen zu mindern“ (201, S. 403). Der Branche „Finance and Insurance“ wird für die US-Wirtschaft ein hohes Potential zu geordnet (202, S. 29). Big Data-Technologie besitzt die notwendigen Voraussetzungen um eine verbesserte Betrugserkennung zu erbringen.

Für eine bessere Bewertungsgrundlage sind bekanntlich eine große Menge an Daten zu berücksichtigen z.B. für Hagelschäden eine Vielzahl von Geodaten und Wetterdaten zu berücksichtigen. So stellen sich eine Vielzahl von Fragen, wo trat der Schaden auf, wie groß sind die zu erwartenden Schäden, wie viele Versicherte sind im Bereich des Auftrittsortes, wie oft kommt es zu Hagelschäden in der geografischen Region, wie viele Fahrzeuge sind als Garagenwagen beim Versicherer geführt, etc. Durch die sozialen Netzwerke lassen sich mit Hilfe von Big Data-Technologie genauere Vorhersagen treffen, die das Risiko exakter abschätzen ließen. Dadurch entstehen ganz neue Dienstleistungen, neue Versicherungsmodelle und neue Produkte.

Dies zeigt auch die Studie von IBM. Diese hat den Einsatz von Big Data im Versicherungswesen in einem ersten Schritt in die drei Bereiche strategisch, fachlich und technisch eingeteilt (203). Im strategischen Bereich werden Geschäftsmodelle, im fachlichen Bereich werden Entscheidungs- und Prozessverbesserung und im technischen Bereich werden die Beherrschung der Datenmenge und -vielfalt betrachtet (203). Dabei wird der Status Quo des Geschäftsmodells bzgl. der Absicherung und Vermeidung betrachtet und das zukünftige Modell in die zwei Bereiche die reine Versicherung und die Rücklage/Vorsorge gegliedert, dadurch soll die Unsicherheit bzgl. des zukünftigen Zustands und des Risikos / der Gefährdung gemindert werden (203). Big Data werden dabei zwei wesentliche Potentiale zugeordnet zum einen höhere Technologie- und Informationsverarbeitungseffizienz mit der Folge „Do more with less“ zum anderen besser informierte Kunden mit der Folge „mehr Power“ (203). Kognitiven Analyse bedeuten für die Versicherungen eine Erhöhung von Richtigkeit, Schnelligkeit, Vielfalt, Sichtbarkeit und des Volumen (203). Dies bedeutet zunächst werden die Daten erfasst, Reaktionen und Ergebnisse im erweiterten Informationsraum externen

und/oder unstrukturierten Daten gemessen (203). In einem zweiten Schritt erfolgt die Auswertung und Bewertung der Informationen in Echtzeit, welches durch den Einsatz eines Regelwerks der Entscheidungsvorbereitung dient (203). In einem dritten Schritt werden Ad-hoc Aktualisierung und Adjustierung von Entscheidungsprozessen ermöglicht. Die Handlung wird dabei mit Echtzeit-Informationen unterstützt (203). Für die Realität bedeutet dies eine optimierte Einsatzsteuerung/Prozesse auf Grundlage von Big Data (203).

In der Tabelle 17 ist das Potential von Big Data für die Branche des Risikomanagements und Versicherungswesens anhand der in diesem Abschnitt beschriebenen Punkte dargestellt.

SWOT-Analyse		Interne Analyse	
Externe Analyse	Chancen	Stärken	Schwächen
	Schnellere, genauere und verlässlichere Betrugserkennung Anomalieerkennung Criminal association rule mining und clustering Analyse von kriminellen Netzwerken Social Media Analyse Unstrukturierte Daten Verbesserte öffentliche Sicherheit Kundenorientiertere Dienstleistungen Schnellere BI Ergebnisse Neue BI Ergebnisse ermöglichen Ad-hoc Aktualisierungen und Adjustierung von Entscheidungsprozessen Höhere Technologie- und Informationsverarbeitungseffizienz		Internetsicherheit Immer neuere Betrugssstrategien
	Risiken	Fake Daten Muster müssen gefunden werden, sind nicht sofort klar Echtzeit-Informationen	Datenschutz Datensicherheit

Tabelle 17: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Risikomanagement und Versicherungswesen (Eigene Darstellung)

2.5 Geschäftsmodellorientierte Analyse der Innovationspotenziale

Für die Ableitung eines Rankings der Schwerpunkte werden die Fallbeispiele der BITKOM Studie in ein Business Model Canvas nach Osterwald und Pigneur eingruppiert. Das Modell zeigt i.d.R. Geschäftsprozesse, bei dieser Untersuchung ist die Betrachtung des Prozesses auf einer Metaebene durchgeführt worden. In der Abbildung 45 werden die Fallbeispiele den Kategorien: Partner, Schlüsselaktivitäten, Schlüsselressourcen, Wertversprechen, Kundenbeziehungen, Kanäle und Kundensegmente zugewiesen. Als Partner konnten neben Unternehmen aus der

Branche Marktforschung, (Social-) Media und Entertainment mit dem Fokus auf (Social-) Media und Entertainment Beispiele aus dem öffentlichen Sektor zugeordnet werden. Schlüsselaktivitäten kommen aus der Risikomanagement und Versicherungswesen-Branche. Schlüsselressourcen sind in der Branche Energie und der Marktforschung mit Blick auf die Datensicherheit und den Datenschutz. Weitere mögliche Projekte wären aus den Branchen des öffentlichen Sektors, der Industrie 4.0 und dem Gesundheitssektor / Life Science denkbar. Jedoch sind diese Fallbeispiele bedingt und aufgrund des Forschungsbedarfs nicht als Schlüsselressourcen zugeordnet. Wie jedoch schon im Unterabschnitt 2.3 beschrieben sind Projekte wie z.B. das der EU Human Brain Projekt entscheidende Schlüsselressourcen, welche technologisch zu erschließen sind. Dies allein aufgrund der Tatsache, dass diese Projekte disruptive Technologien generieren. Der Kategorie Wertversprechen wurden Unternehmen aus dem Bereich der Marktforschung, (Social-) Media und Entertainment mit dem Schwerpunkt Marktforschung zugeordnet. Des Weiteren betrifft es Unternehmen, die ein Wertversprechen im weitesten Sinne gegenüber ihren Kunden geben. In die Kategorie der Kundenbeziehungen ordnen sich Fallbeispiele der gleichen Branche wie die aus der Kategorie Wertversprechen ein, ebenfalls mit dem Schwerpunkt auf Marktforschung. Die Kategorie Kanäle wurde lediglich angedeutet, hier würden Unternehmen aus unterschiedlichen Sektoren fallen, die sich auf die Übertragung und Sensorik zur Erlangung von Daten konzentrieren. Dies könnten z.B. Telekommunikationskonzerne sein. Darüber hinaus Hersteller von Sensoren unterschiedlicher Arten, z.B. der Messung von Energie im weitesten Sinne oder Übertragungswege hier durch das Smartphone symbolisiert. Darüber hinaus ist der Big Data-Kanal jederzeit erreichbar bis auf wenige Ausnahmen im Jahr. In die letzte Kategorie fallen anhand der gegebenen Fallbeispiele im Wesentlichen Unternehmen aus der Branche der Industrie 4.0 und der Mobilitätsdienstleistungen. Es ist dabei entscheidend eine Balance zwischen den bestehenden Technologien (Exploit) und den neuen (Big Data) Technologien (Explore) herzustellen. In Abbildung 47 ist die genaue Zuordnung und der Zusammenhang zwischen Exploit und Explore dargestellt.

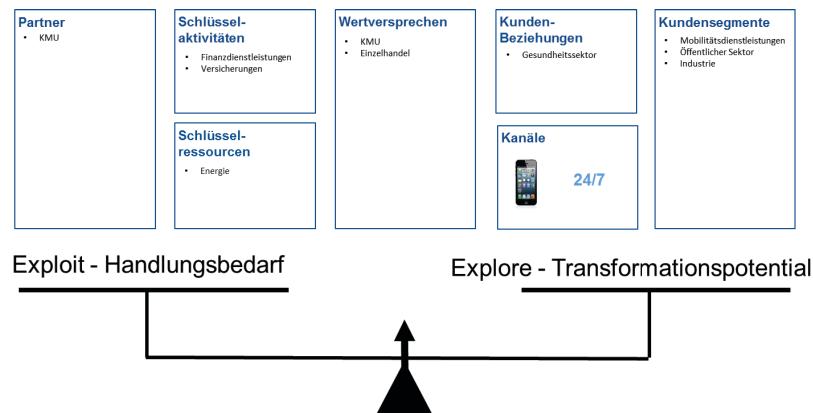


Abbildung 47: Business Model Canvas zur BITKOM Studie (In Anlehnung an Osterwald, Pigneur (2010))

2.6 Priorisierung der Schwerpunkte

Nachdem die einzelnen ausgewählten Branchen vertiefend untersucht wurden können Schwerpunkte identifiziert werden. Dazu werden aus entscheidenden Studien wesentliche Informationen noch einmal näher betrachtet. Anschließend werden unterschiedliche Potentiale anhand von Vorher-Nachher-Vergleichen mittels Kodierung aus Fallbeispielen der BITKOM Studie extrahiert und visualisiert.

In einem ersten Schritt gehen wir auf eine Studie von TNS Infratest ein, die zeigt, dass bei Versorgungs-, Dienstleistungs- und Kommunikationsunternehmen ein hoher Handlungsdruck im Bereich von Big Data besteht (182, S. 10). Darüber hinaus ist für Entscheider aus diesen Branchen in der IKT die Relevanz von Big Data Technologie überdurchschnittlich hoch (182, S. 10). In Abbildung 48 ist die Potentialbewertung für die Branchen aus der TNS Infratest Studie dargestellt.

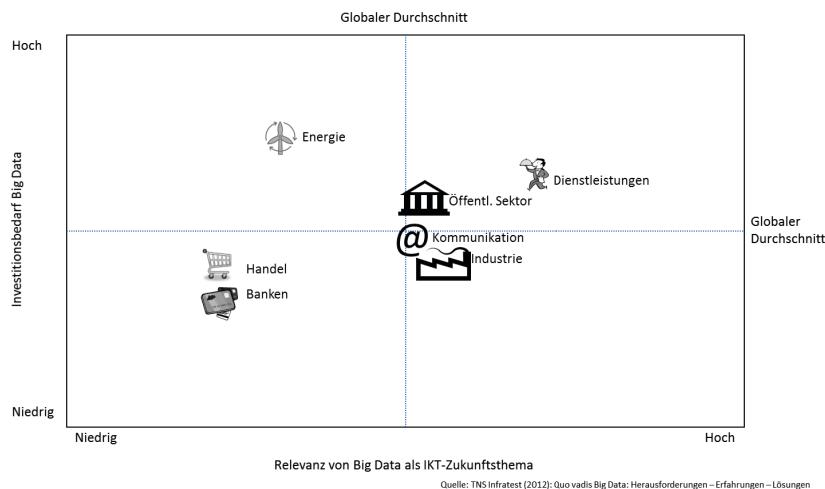


Abbildung 48: Handlungsdruck nach Branchen im Bereich Big Data (Quelle: TNS Infratest(182) (2012))

Erfolgsfaktoren für Wachstum und Wettbewerbsfähigkeit der deutschen IT-Branche. Die deutschen Infrastrukturvoraussetzungen für den Umgang mit Big Data liegen in der Breitband-Internetversorgung zu den Unternehmen. Ohne diese ist eine Übertragung der anfallenden Datenmengen nur bedingt möglich. Demzufolge ist der Ausbau der Infrastruktur eine entscheidende Ressource für den Einsatz von Big Data. Aber auch eine Dienstleistungskomponente für Unternehmen. Insbesondere falls der Prozess der Datenanalyse dem Outsourcing unterliegt. Dabei ist anzunehmen, dass der Transport des Datensatzes erheblich mehr Informationen und damit Volumen enthält als der Ergebnisdatensatz.

Außerdem zeigen die Untersuchungen mit Hilfe des Business Model Canvas aus dem Unterabschnitt 2.5, dass eine der Schlüsselressourcen der Datenschutz und die Datensicherheit ist. Dies gilt bei Big Data eher als Manko, allerdings kann Deutschland mit seinen hohen und angesehenen Kriterien und Richtlinien in diesem Bereich weltweit führend sein. Deutschland wäre dann in Lage die Nachfrage nach nachhaltiger und sicherer Big Data-Technologie zu bedienen. Insbesondere auch zum Schutz seines Mittelstandes vor möglicher Wirtschaftsspionage.

In einem zweiten Schritt wird nun zur Bewertung des Potentials von Big Data und zur Bestimmung von dessen wirtschaftlichen Anforderungen ein Vorher-Nachher-Vergleich anhand von Praxisbeispielen aus der Big Data Studie der BIT-KOM durchgeführt. Dabei wird anhand mehrerer Dimensionen untersucht, wie sich der Einsatz von Big Data in der Praxis auswirkt. Außerdem wird eine SWOT-Analyse durchgeführt, um die Stärken (Strengths), Schwächen (Weaknesses), Chancen (Opportunities) und Risiken (Threats) von Big Data aufzuzeigen.

Für die Auswahl geeigneter Dimensionen werden zunächst die Definitionen von Big Data genauer betrachtet.

Die Ursprungsdefinition von Gartner ordnet Big Data den drei Dimensionen, teilweise als Bewertungskriterien bezeichnet, Menge, Vielfalt und Geschwindigkeit zu (178). IBM nennt ein weiteres viertes Kriterium: den Wert der Daten. Das Kriterium des Datenumfangs sei bei den folgenden Untersuchungen als gegeben angenommen. Für eine erste Bewertung werden die Dimensionen Vielfalt der Daten und Geschwindigkeit mit welcher die Daten verarbeitet werden können herangezogen. Die Vielfalt der Daten wird in niedrige und hohe Vielfalt unterschieden. Daten mit einer niedrigen Vielfalt sind z.B. einfache Datentypen wie Integer, Double oder auswählbarer Text. Komplexere meist unstrukturierte Daten wie Texte oder Sprachen werden einer hohen Vielfalt zugeordnet. In der Dimension Geschwindigkeit wird zwischen niedriger und hoher Geschwindigkeit unterschieden, in welcher die Daten verarbeitet werden können. Es handelt sich dabei um die von den Praxisbeispielen angegebenen subjektiven Geschwindigkeiten, da eine objektive Bewertung der notwendigen Verarbeitungsgeschwindigkeit aus den vorhandenen Praxisbeispielen nicht möglich ist.

Die Dimensionen werden dabei auf die jeweilige Achse des Diagramms aufgetragen. Beispielsweise wird die Vielfalt auf der y-Achse und die Geschwindigkeit auf der x-Achse abgetragen (siehe Abbildung 48). Die Pfeile geben an, wie sich das Unternehmen mit Einführung von Big Data-Technologie entwickelt hat. Auf der Achse selbst liegen Punkte, die zum jeweiligen Zeitpunkt nicht angewendet werden konnten, z.B. weil es die Verarbeitungsgeschwindigkeit mit den bisherigen Methoden und Technologien nicht zuließ. In Abbildung 48 ist das jeweilige Potential der Unternehmen nach den soeben erörterten Dimensionen dargestellt. Die Pfeilrichtung gibt dabei das Potential an, d.h. die Änderung von Vorher (vor dem Einsatz von Big Data) zum Ist-Stand (nach dem Einsatz von Big Data) an.

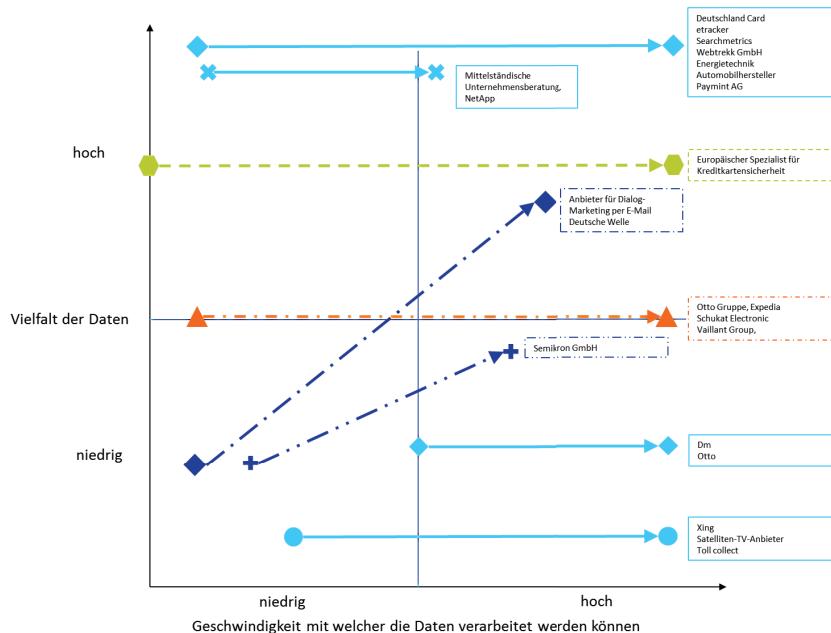


Abbildung 49: Vorher-Nachher-Vergleich der Dimensionen Datenvielfalt zu Verarbeitungsgeschwindigkeit zur Potentialbewertung (Eigene Darstellung, Quelle: BITKOM)

Die Auswertung der Analyse zeigt, dass es für einen überwiegenden Anteil an Unternehmen mit dem Einsatz von Big Data-Technologie möglich ist einen signifikanten Geschwindigkeitsgewinn in der Verarbeitung der Daten zu erzielen. In Bezug auf die Nutzung von Daten mit einer hohen Vielfalt besteht, wie aus der Abbildung 48 im Quadranten unten rechts ersichtlich ist, noch erhöhter Forschungsbedarf. In den gegebenen Fallbeispielen konnte kein Unternehmen identifiziert werden, welches vor dem Einsatz von Big Data-Technologie diesem Bereich zugeordnet werden konnte. Des Weiteren ist zu erkennen, dass alle Unternehmen nach dem Einsatz von Big Data-Technologie in der Lage waren die zur Verfügung stehenden Daten zu verarbeiten und zu beherrschen.

Des Weiteren lässt sich aus den Potentialanalysen der Unterabschnitte 2.1 bis 2.5 schließen, dass mit zunehmender Geschwindigkeit in welcher die Daten zur Verfügung stehen müssen auch das Potential steigt. Eine proportionale Abhängigkeit konnte dabei bisher nicht festgestellt werden. Dies wäre in zukünftigen Studien und Untersuchungen zu zeigen. Ob für die Qualität eine historische oder kurzfristige Datenbasis besser ist, ließ sich in der Literatur nicht feststellen. Demgemäß ist eine Prüfung auf Qualität im Einzelfall stets zusätzlich durchzuführen.

Insgesamt lassen sich vier Bereiche identifizieren, welche die untersuchten Unternehmen charakterisieren und in welche sie eingruppiert werden. Die Versteher unter den Big Data verwendenden Unternehmen kommen sowohl mit einer

hohen Vielfalt der Daten zu Recht, als auch einer hohen Geschwindigkeit mit welcher die Daten verarbeitet werden müssen. Die zweite Gruppe sind die Beschleuniger, das sind diejenigen, welche eine Beschleunigung durch den Einsatz von Big Data-Technologie erzielen. Eine dritte Gruppe bilden diejenigen, welche vorher ohne den Einsatz von Big Data-Technologie keine oder nur bedingte Möglichkeit besaßen die Datenmengen oder die Datengeschwindigkeit zu verarbeiten und nach dem Einsatz von Big Data-Technologie eine Verbesserung in mindestens einer der beiden erzielen konnten, wir bezeichnen diese als die Verbesserer. Die vierte und letzte Gruppe sind Unternehmen, die eine niedrige Datenvielfalt zu bewerkstelligen haben und eine hohe Geschwindigkeit zur Verarbeitung der Daten benötigen. Der vierten Gruppe konnten keine Unternehmen zugeordnet werden. Daher besteht ein Forschungsbedarf, zum einen ob es solche Unternehmen gibt und wenn ja wie diese Unternehmen zu den Verstehern in der jeweiligen Branche werden können. Dazu gehört auch die Erforschung der Rahmenbedingungen, welche vorhanden sein oder geschaffen werden müssen. Des Weiteren welche Faktoren einen nachhaltigen Effekt besitzen.

Ferner ist aus der zu erkennen, dass ein Teil der Unternehmen, ohne Big Data Lösung/Ansatz keine Möglichkeit der Handhabung dieser Daten besitzt. Die meisten dieser Unternehmen sind nach dem Einsatz von Big Data in der Lage zu den Verstehern aufzuschließen und beschleunigen damit nicht nur sondern erhöhen gleichzeitig die Datenvielfalt. Des Weiteren ist zu erkennen, dass ein hohes Potential in der Beschleunigung der zu verarbeitenden Daten liegt.

In einem nächsten Schritt wurden als weitere Dimensionen die Kontrolle der Wertschöpfung und die Transparenz der Wertschöpfung jeweils aus interner Sicht für das jeweilige Unternehmen untersucht. Auffällig ist, dass die Mehrheit der Unternehmensbeispiele nach dem Einsatz von Big Data-Technologie im hohen Bereich beider Dimensionen liegt. Die Fallbeispiele aus der Praxis zeigten kein Unternehmen, welches bereits vor dem Einsatz von Big Data-Technologien eine hohe Transparenz in der Wertschöpfung besaß. Alternative Ursachen für dieses Ergebnis könnten jedoch auch in dem kleinen Stichprobenumfang, den gewählten Praxisbeispielen oder den zu den Beispielen vorhandenen Informationen begründet sein. Aus diesem Grund wird empfohlen dieses Ergebnis in weiteren Untersuchungen zu validieren. In Abbildung 50 sind die Details zu den Unternehmen dargestellt.

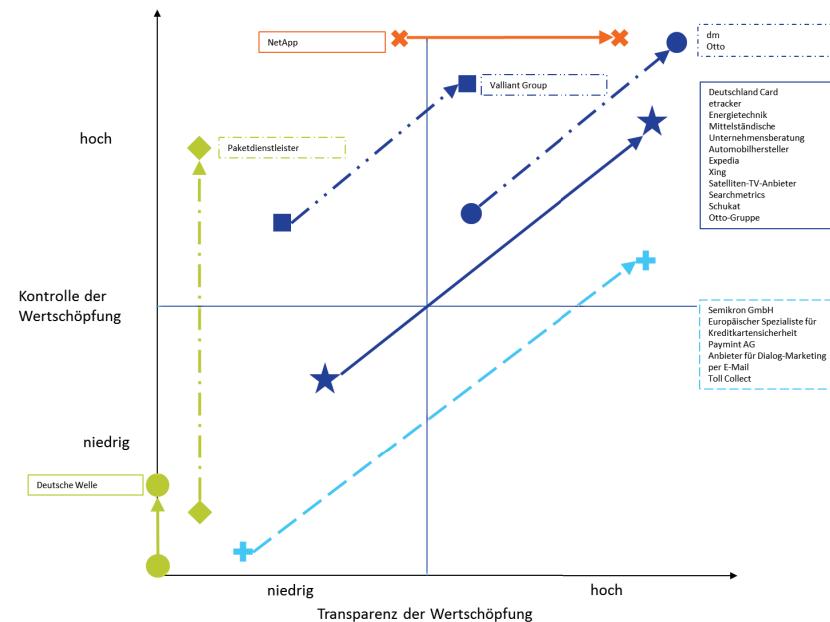


Abbildung 50: Vorher-Nachher-Vergleich der Dimensionen Kontrolle zu Transparenz der Wertschöpfung zur Potentialbewertung (Eigene Darstellung, Quelle: BITKOM)

Bei allen Fallbeispielen ist zu beobachten, dass mindestens die Kontrolle und/ oder mind. die Transparenz der Wertschöpfung zunimmt. Die Vielzahl der Unternehmen, wie aus Abbildung 50 zu erkennen ist, hat eine Steigerung in beiden Dimensionen.

In einem letzten Schritt wurden als Dimensionen die Gültigkeit der Empfehlung und die Datenlage gewählt. Dabei ist zunächst überraschend, dass die Gültigkeit der Empfehlungen, welche durch Big Data-Technologie entstehen nicht zunimmt. Bei einem Unternehmen, dem Anbieter für Dialog-Marketing per E-Mail, nahm die Gültigkeit sogar ab. Dafür konnten Entscheidungen jedoch ad hoc gefällt werden. Es ist nicht vollständig klar, warum bei der Wahl dieser Dimensionen kaum Veränderungen durch den Einsatz von Big Data-Technologie entstehen. Auf diesem Gebiet besteht ein erheblicher Forschungsbedarf. In Abbildung 51 ist die Entwicklung mittels Big Data-Technologie auf den beschriebenen Dimensionen zur Gültigkeit der Empfehlungen und zur Datenlage dargestellt.

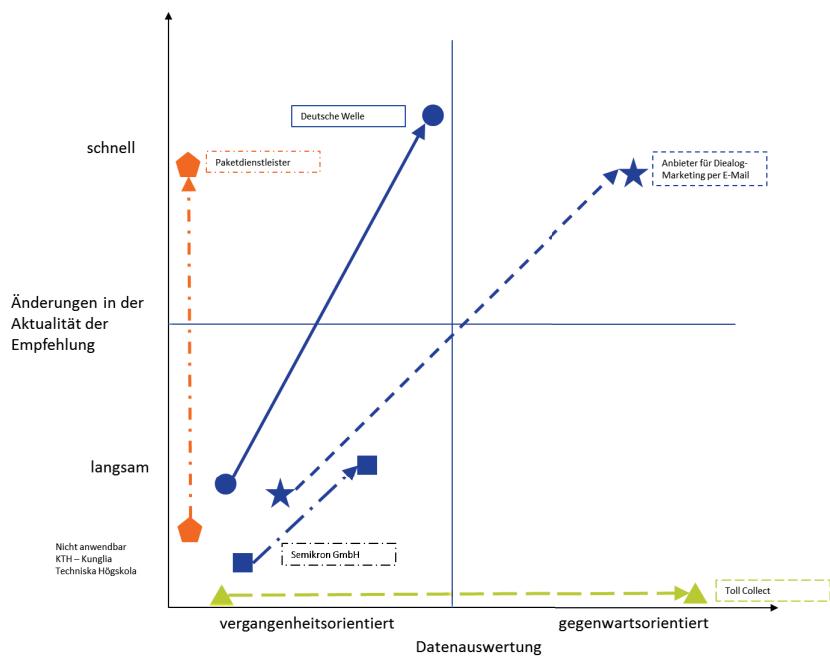


Abbildung 51: Vorher-Nachher-Vergleich der Dimensionen Gültigkeit der Empfehlung zu Zeitraum der Datenlage zur Potentialbewertung (Eigene Darstellung, Quelle: BITKOM)

Eine Verschiebung von strategischen hin zu operativen Entscheidungen konnte nicht festgestellt werden. Big Data ermöglicht jedoch neue Entscheidungen sowohl strategischer als auch operativer Art und Weise. Bei den Unternehmen Deutsche Welle, Europäischer Spezialist für Kreditkartensicherheit, Paketdienstleister und Anbieter für Dialog-Marketing per E-Mail sind nun strategische Entscheidungen bei kurzfristiger Verfügbarkeit der Daten, welche als Entscheidungsgrundlage dienen, anwendbar. Auf operativer Ebene sind neue Entscheidungen nun bei dm und Otto ebenfalls bei einer kurzfristigen Verfügbarkeit der Entscheidungsgrundlage möglich. Insgesamt ist die Verfügbarkeit der Entscheidungsgrundlage von langfristig zu kurzfristig zu erkennen. Big Data bietet somit ein erhebliches Potential bei der Verfügbarkeit der Entscheidungsgrundlage im Unternehmen an.

In einem letzten Vergleich schauen wir uns die Relevanz für Entscheidungsträger näher an. Dazu wird auf die innere Sicht des Unternehmens reduziert. Dies ermöglicht es herauszufinden, ob Big Data für Entscheidungsträger eine entscheidende Rolle spielt. Auf der x-Achse liegt dabei die Zeitrelevanz der Entscheidung. Dies bedeutet, ob die Entscheidung eher operativ und somit kurzfristig zu treffen ist oder strategisch. Die zweite Dimension ist das Zeitfenster, was den Entscheidungsträger zur Verfügung steht um eine Entscheidung zu fällen. Sollte für eine Entscheidung beliebig viel Zeit zur Verfügung stehen, so kann Big Data zwar eine Auswertung ermöglichen, jedoch ist die Geschwindigkeit (Velocity) dabei

nicht entscheidend. In Abbildung 52 ist die Zuordnung der Dimensionen mittels Vorher-Nachher-vergleichen dargestellt.

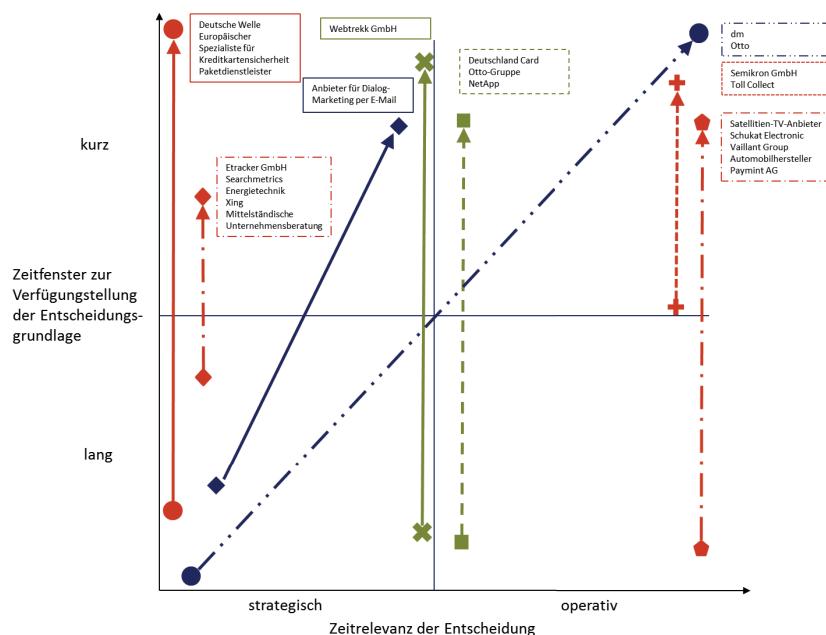


Abbildung 52: Bewertung des Potentials für die Dimensionen Verfügbarkeit der Entscheidungsgrundlage und seiner Zeitrelevanz (Eigene Darstellung, Quelle: BITKOM)

Insgesamt ist festzustellen, dass alle Unternehmen durch Big Data in der Lage sind nun kurzfristig ihre Entscheidungen zu treffen. Dies bietet einen erheblichen Vorteil gegenüber Konkurrenzunternehmen und stellt damit einen erheblichen Wettbewerbsvorteil dar. Lediglich die zwei Unternehmen dm und Otto waren in der Lage mittels ihrer Big Data-Lösung, welche stark BI getrieben ist von strategischen zu operativen Entscheidungen zu optimieren. Das bedeutet, dass kurzfristig nun auch auf operative und sehr kurzfristige Änderungen reagiert werden kann.

In einem nächsten Schritt werden die zu den einzelnen Stärken, Schwächen, Chancen und Risiken aus den im Unterabschnitt 2.4 durchgeföhrten SWOT-Analysen aggregiert und eine Bewertung mittels einer Nutzwertanalyse durchgeführt. Die folgende Tabelle 18 zeigt die identifizierten Stärken, Schwächen, Chancen und Risiken für die untersuchten Branchen im Überblick.

Branchen	Schwerpunkte	Öffentlicher Sektor					
		Industrie	Gesundheitssektor	Marktforschung, (Sozial-) Media und Entertainment	Mobilitätsdienstleistungen	Energie	Versicherungen
	Reaktion/Abwicklung	X	X		X		X
	Kostenreduktion	X	X	X			X
	Individuelle Kundenwünsche		X	X		X	X
	Genauere Analyse			X			X
	Neue Dienstleistungen			X		X	
	Bessere BI/Entscheidungen	X		X			X
	Neue BI-Ergebnisse		X		X		X
	Social Media				X		X
	Schnellere gegenwärtige Dienstleistungen der Regierung	X					
	Öffentliche Dienstleistungen	X					
	Politische Kampagnen	X					
	Gleichheit der Zugriffsmöglichkeiten und öffentlichen Dienstleistungen		X				
	Geschäfts- und Engineering Prozesse dynamisch		X				
	Durchgängige Transparenz	X					
	Neue Formen von Wertschöpfung	X					
	Neuartige Geschäftsmodelle	X					
	Bessere Erkennung von Krankheiten			X			
	Bessere Ärztliche Versorgung			X			
	Prävention			X			
	Genanalysierung und –visualisierung			X			
	Langfristigere Behandlung			X			
Volumen	Individuelle Patientenbehandlung						
	Wissen über Risikofaktoren						
	Nebenwirkungen			X			
	Reaktionen						
	Flexible Reaktionen auf Marktumfeld				X		
	Genauere und vielfältigere Angebote					X	
	Weniger Stau					X	
	Höherer Komfort durch mehr Service					X	
	Prognose						
	Qualitätssteigerung						
	Auswertungserstellung						
	Höhere Vielfalt						X
	Anomalieerkennung						X
	Criminal association rule mining und clustering						X
	Analyse von kriminellen Netzwerken						X
	Unstrukturierte Daten						X
	Verbesserte öffentliche Sicherheit						X
	Höhere Technologie- und Informationsverarbeitungseffizienz						X

Branchen	Öffentlicher Sektor	Industrie	Gesundheitssektor	Marktforschung, (Social-) Media und Entertainment	Mobilitätsdienstleistungen	Energie	Versicherungen
Schwerpunkte							
Kosten	X	X					
Kundenindividuelle Lösungen			X	X			
Datenvolumen	X		X				
Kosten für Strukturierung der Daten	X						
Verteilte Informationsquellen	X						
Abstimmung von Bund, Ländern, Kommunen	X						
Datenvolumen	X						
Transparenz	X						
elektronische Wahlen (e-polling)	X						
Investitionen		X					
Social Media Patientenplattformen analysieren			X				
Medikamentnebenwirkungen analysieren			X				
Patienten mehr einbinden/ mehr Verantwortung übertragen			X				
Kundenindividuelle Lösungen (gläserne Kunde)				X	X		
Langfristiges Marketing und gezielte, personalisierte Empfehlungen				X			
Reichhaltigen Informationsschatz nutzen				X			
Infrastruktur (Lademöglichkeiten, Car-Sharing)					X		
Von statischer zu dynamische Steuerung						X	
Internetsicherheit, Betrugssstrategien						X	
Velocity							
Distributive Daten	X	X					
Abstimmung von Bund, Ländern, Kommunen	X						
Unstrukturierte informelle Kommunikation mit dem Bürger	X		X				
Stark textueller Inhalt	X						
Value/varacity							
Vertrauen		X					
Verlässlichkeit		X					
Nutzung		X					
Konvergenz der Betreibermodelle		X					
Personenspezifischer Inhalt	X		X				
Verschiedener aber eng verwandter Inhalt/mehrsprachiger Inhalt			X	X			
Genauere Betrugserkennung (Fake Daten/Muster nicht sofort klar)					X	X	
Einsparpotential kann negativ sein oder zu einer erhöhten Nachhaltigkeit führen						X	
Echtzeit-Informationen							X

Branchen		Öffentlicher Sektor	Industrie	Gesundheitssektor	Marktforschung, (Social-) Media und Entertainment	Mobilitätsdienstleistungen	Energie	Versicherungen
Variety	Schwerpunkte							
	Datenschutz	X	X	X	X	X	X	X
	Datensicherheit	X	X	X	X	X	X	X
	Latenzen		X				X	
	Ethische Gründe			X				
	Gesellschaftliche Akzeptanz						X	
Summe		21	19	18	14	11	9	17

Tabelle 18: Potentialgewichtung nach ausgewählten Branchen (Eigene Darstellung)

Im nächsten Schritt sind die gewichteten Branchen zusammenfassend in Abbildung 53 dargestellt. Dabei zeigt der blaue Balken die Anzahl an Vorkommen aus der SWOT-Analyse und der rote Balken den prozentualen Anteil an Fallbeispielen

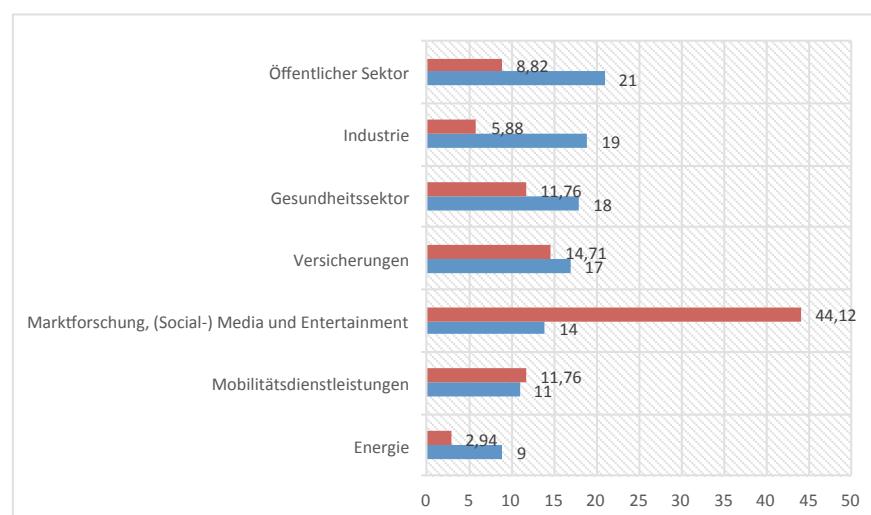


Abbildung 53: Potential nach Branchen und prozentualer Anteil an untersuchten Fallbeispielen (Eigene Darstellung)

Ein hohes Potential (Top 4) zeigen dabei der öffentliche Sektor, die Industrie, der Gesundheitssektor/ Life Science und das Risikomanagement und das Versicherungswesen. Auf den ersten Blick scheint der Bereich der Energie sehr niedrig, dies hängt jedoch auch damit zusammen, dass in der BITKOM Studie wie bereits in Abbildung 36 gezeigt gerade einmal 2,94 Prozent der untersuchten Fälle

liegen, bei Marktforschung zeigt sich ein gegenteiliges Bild. Hieraus lässt sich schließen, dass der Anteil an Energie deutlich höher sein müsste und der Anteil an Marktforschung niedriger. Dies wird unterstützt durch die bereits getätigten Aussagen. Die Big Data-Technologie wird in der Marktforschung bereits stark mittels BI & A verwendet, während auch TNS zu dem Ergebnis kommt, dass die Energie noch ein erhebliches Potential aufweist.

2.7 Zusammenfassung

In den Untersuchungen und den ausgewählten Branchen zeigt sich, dass das Datenvolumen über unterschiedliche Technologien, wie z.B. Hadoop, Stratosphere, in einigen Fällen auch über SAP HANA, ParStream, oder andere In-Memory-Datenbanken beherrschbar ist. Der Datenschutz und die Datensicherheit zeigen ein erhebliches Potential für eine Big Data-Technologie in Deutschland. Aufgrund dessen, dass in Deutschland ein ausgeprägtes Verständnis von vertrauenswürdigen Daten vorhanden ist, kann Deutschland in diesem Bereich zum Marktführer werden. Dazu besteht in dem Bereich des Datenschutzes und der Datensicherheit weiterer Forschungsbedarf, insbesondere im Hinblick auf Integration von Datenschutzfunktionalität in existierende oder entstehende Datenanalysesysteme bzw. Algorithmen. Insbesondere zur Wahrung von wirtschaftlichen Interessen ist der Schutz von Daten von erheblicher Bedeutung. Zum besseren Verständnis werden in Kapitel 3 die rechtlichen Rahmenbedingungen näher betrachtet.

Aus den einzelnen untersuchten Branchen lassen sich aufgrund der vorherigen Analysen aus den Kapiteln 2.1 bis 2.9 Schwerpunkte ableiten und wie folgt reihen.

Branchenschwerpunkte:

1. Entwicklung neuartiger Technologien, um eine skalierbare Verarbeitung von komplexen Datenanalyseverfahren auf riesigen, heterogenen Datenmengen mit hoher Datenrate zu realisieren
2. Senkung der Zeit und Kosten der Datenanalyse durch automatische Parallelisierung und Optimierung von deklarativen Datenanalysespezifikationen
3. Schaffung von Technologieimpulsen, die zur erfolgreichen weltweiten Kommerzialisierung von in Deutschland entwickelten, skalierbaren Datenanalysesystemen führen
4. Ausbildung von Multiplikatoren im Bereich der Datenanalyse und der skalierbaren Datenverarbeitung, welche die Möglichkeiten von Big Data in Wissenschaft und Wirtschaft tragen werden

5. Technologietransfer an Pilotanwendungen in Wissenschaft und Wirtschaft
6. Schaffung eines Innovationsklimas durch Konzentration von kritischem Big Data Know-how, damit deutsche Unternehmen und Wissenschaft nicht im Schatten des Silicon Valleys stehen
7. Interaktive, iterative Informationsanalyse für Text und Weiterentwicklung geeigneter Geschäftsmodelle zur Schaffung von Marktplätzen für Daten, Datenqualität und Verwertung von Daten
8. Datenschutz und Datensicherheit

Abbildung 54: Abgeleitete Schwerpunkte aus den verschiedenen Branchen (eigene Darstellung)

Die Entwicklung von Big Data kann neue Wettbewerbsvorteile schaffen und sichern. Dies setzt zum einen unternehmerische Kreativität und zum anderen systematische Geschäftsmodellentwicklung voraus. Big Data ist besonders bei Dienstleistungen von Relevanz. Allerdings ist es notwendig, dass Big Data eine Anwendung in der Praxis findet, um neue Dienstleistungen zu ermöglichen. Die einzelnen Branchen werden stärker, gemeinsamer und interdisziplinärer zusammenarbeiten müssen, da Big Data einen fließenden Übergang zwischen den Branchen schafft.

Big Data-Technologie ist bereits in Ansätzen, wie erste Anwendungen und Untersuchung zeigen, vorhanden. Erste Anwendungen können mit Big Data-Technologie bereits realisiert werden (vgl. BITKOM Studie). Jedoch beginnt die Diffusion der Big Data-Technologie gerade. Aufgrund dieser Tatsache können zwar die Erwartungen an die Technologie abgefragt werden, allerdings kann die Technologie noch nicht den gesamten Markt durchdringen, die Umsetzung muss in den einzelnen Branchen noch erfolgen. Vor dieser Umsetzung stehen zum einen die technische Verfügbarkeit, die ökonomische Sinnhaftigkeit und zum anderen die breite Akzeptanz samt rechtlicher Zulässigkeit. Die Erwartungen sind aufgrund des hohen erhofften Potenzials groß.

KMU Anbieter und Dienstleister sind zu fördern, damit diese die Datenmengen bewältigen können und einen Zugang zu Big Data Technologie erhalten. Referenzbeispiele sind bereits vorhanden. Dabei wird der Nutzerkreis verbreitert und neue Geschäftsmodelle entstehen. Geschäftsmodelle besonders in der Finanz- und Versicherungsbranche zeigen einen hohen Value und Veracity auf. So finden Big Data-Technologie einen experimentellen Einsatz. Großer Forschungs- und Entwicklungsbedarf ist in der Algorithmen Technologie, welche Sensorik, Logdaten, etc. beherrscht. Diese sind z.B. im Energie- und Industriebereich zu finden.

Die Vielfalt (Variety) der verschiedenen Datenarten und –typen steht dabei im Vordergrund. Abbildung 55 zeigt dabei einen Überblick über diese Untersuchung.

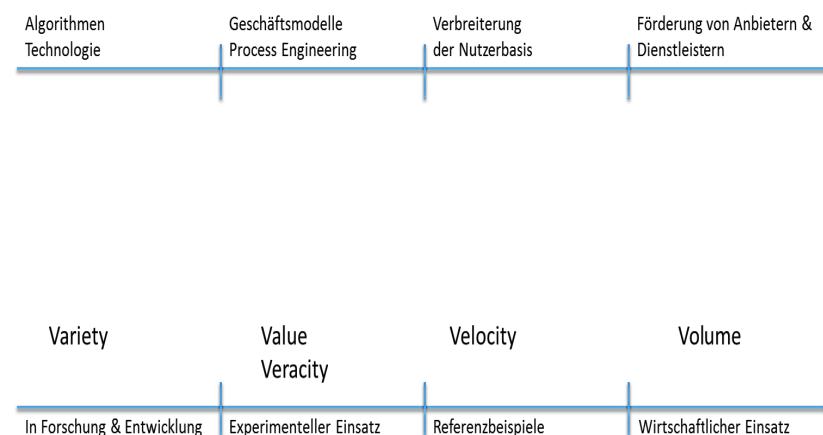


Abbildung 55: Überblick der Untersuchung (Eigene Darstellung)

3 Analyse von Big Data & Big Data Technologien

Aus den Anforderungen der Industrie (siehe Kapitel 2) können wir vier Kernanforderungen für das Management von Big Data ableiten:

1. **Umgang mit großen, heterogenen Datenmengen:** Datenanalysesysteme müssen Datenströme und große Mengen statischer Daten mit geringer Latenz analysieren können. Diese Daten liegen oft in unterschiedlichen Medientypen (Text/Bild/Audio/Video) und Datenmodellen (relational, hierarchisch, graph-strukturiert) vor. Dazu müssen statistische und mathematische Methoden des maschinellen Lernens, der Sprachverarbeitung und der Signalverarbeitung mit semantischen Methoden verbunden werden.
2. **Komplexe Datenanalysealgorithmen:** Die Algorithmen der Datenanalyse gehen durch den Einsatz von benutzerdefinierten Funktionen, Iterationen und den Zugriff auf einen potentiell verteilten Zustand über die klassischen Operationen der relationalen Algebra hinaus, welche in Standarddatenbanksystemen realisiert sind. Um diese neuartigen Algorithmen in Datenanalysesystemen zu spezifizieren, sind deklarative Sprachen erforderlich, die über die klassischen Möglichkeiten von SQL hinausgehen. Gleichzeitig sind existierende NoSQL-Systeme wie Map/Reduce unzureichend, da diese keine automatische Optimierung und Parallelisierung komplexer, iterativer Datenanalysealgorithmen ermöglichen und somit Datenanalysen in diesen Systemen den wenigen Experten vorbehalten sind, die gleichzeitig Systemprogrammierungskenntnisse und Kenntnisse in Datenanalyse und maschinellem Lernen besitzen.
3. **Interaktive, oftmals visuell unterstützte Datenanalyse:** Datenanalyse ist oftmals ein Prozess, in dem aufgrund von Annahmen ein Modell gebaut, validiert/interpretiert und iterativ verfeinert wird. Interpretation und Validierung werden oftmals visuell unterstützt, so dass ein derartiger Prozess üblicherweise als „Visual Analytics“ bezeichnet wird. Datenanalysesysteme (218) müssen die Ausführung von Analyseverfahren auf riesigen Datenvolumen mit ausreichend hoher Verarbeitungsgeschwindigkeit und geringer Latenzzeit ermöglichen. In einem interaktiven, also durch

menschliches Feedback gesteuerten Szenario wäre „ausreichend“ definiert durch das Erbringen erster relevanter Resultate durch einen Algorithmus in weniger als einer Sekunde.

4. **Nachvollziehbare Datenanalyse:** Das Big Data Management umfasst insbesondere auch Mechanismen zur Bewertung der Herkunft der Daten, zur Abschätzung der Korrektheit der benutzten Data-Mining Modelle bzw. Vorhersagemodelle. Wesentlich sind dabei insbesondere die Aspekte statistische Signifikanz (d.h. Verwendung einer angemessenen Stichprobe), Modellwahl (z.B. chaotische Datenverteilungen ohne statistische Momente gegenüber gauss-artigen Datenverteilungen mit wohldefinierten statistischen Momenten) sowie die messbare Bewertung von Qualitätseigenschaften der Daten.

Diese Herausforderungen werden von existierenden Datenmanagementsystemen bisher nicht erreicht. Um diese Herausforderungen zu meistern, müssen **skalierbare, einfach zu bedienende Datenanalysesysteme und neue Algorithmen bzw. Paradigmen zur Datenanalyse** entwickelt werden, die die verschiedenen Aspekte und Anforderungen gleichzeitig adressieren. Das folgende Kapitel beschreibt vier zu adressierende Herausforderungen: 1.) Die Entwicklung und Verbesserung von komplexen Analysemethoden entlang des iterativen Analyseprozesses auf großen, heterogenen Inhalten mit geringer Latenz 2.) die Entwicklung von Systemen und prozessübergreifenden Technologien 3.) die Integration und Kommerzialisierung der Prozesse und Technologien in Systemen und 4.) die Be- und Verwertung von Daten auf Informationsmarktplätzen.

3.1 Der Datenanalyseprozess

Data Science. Neben der Erschließung von neuen Datenquellen ist ein weiteres Merkmal des „Big Data“ Trends ein steigender Bedarf für komplexe Analysen. Während in der Vergangenheit hauptsächlich relationale OLAP und Reportinganfragen vorherrschten, wird nun die Nutzung von maschinellen Lern- und Data-Mining Verfahren und Signalverarbeitungsalgorithmen immer wichtiger. Diese Technologien werden zum Beispiel angewendet um Vorhersage- oder Klassierungsmodelle zu trainieren, komplexe statistische Zusammenhänge in Daten zu entdecken oder Informationsextraktoren zu erstellen. Angewendet auf neue Datenquellen, ergeben sich zahllose interessante Anwendungsmöglichkeiten. Für die Disziplin aus vielfältigen Datensätzen mit Hilfe von komplexen Datenanalysen neue wertvolle Informationen zu gewinnen wurde der Begriff „Data Science“ geprägt. Sogenannte „Data Scientists“ benötigen umfangreiches Wissen in den Bereichen Statistik, maschinelles Lernen, Informationsvisualisierung und Datenverarbeitung. Angesichts des riesigen Potentials, das in Data Science steckt, stehen wir gerade erst am Anfang dieser Entwicklung.

Der Datenanalyseprozess. Die Datenanalyse beschränkt sich nicht nur auf das effiziente Ausführen von maschinellen Lernverfahren oder Data-Mining Algorithmen. Vielmehr ist dieser Aspekt nur einer von mehreren Schritten eines Datenanalyseprozesses. Nachdem das Ziel eines Datenanalyseprojekts definiert worden ist, werden zunächst **geeignete Datenquellen identifiziert**. Anschließend müssen die Daten dieser Quellen in mehreren, oft aufwändigen, Schritten vorverarbeitet werden, da die meisten Datenanalyseverfahren strukturierte Daten als Eingabe voraussetzen. So müssen zum Beispiel aus Text-, Bild-, Audio- oder Videodaten zunächst **relevante Informationen extrahiert** werden, bevor diese analysiert werden können. Ein weiterer wichtiger Schritt ist die **Integration und Säuberung von Daten**. Dazu zählen die Konvertierung in ein einheitliches Datenformat und die Entfernung von Datendubletten. Viele Datenanalysealgorithmen wie das Trainieren von Vorhersage- oder Klassifikationsmodellen werden nur auf einer Teilmenge aller Dateneigenschaften ausgeführt. Diese Eigenschaften müssen besonders relevant und ausdrucksstark für das zu lösende Problem sein und werden „Features“ genannt. Die Identifikation von Features und ihre Extraktion erfordern häufig einen großen Rechenaufwand. Des Weiteren haben gängige Maschinenlern-Algorithmen zahlreiche Parameter, die eingestellt werden müssen, um gute Ergebnisse zu erzielen. Aus diesen Gründen muss der Schritt der **Datenanalyse und Modellbildung** häufig mehrfach mit unterschiedlichen ausgewählten Features und Parametern durchgeführt werden, um ein gutes Analyseergebnis zu erzielen. Nach jeder Ausführung findet eine Überprüfung des Ergebnisses eines Datenanalyseprogrammes statt. Zum Beispiel kann die Qualität eines Vorhersagemodells oder eines Informationsextraktors mittels eines Testdatensatzes validiert werden. Eine solche **Überprüfung lässt sich durch eine geeignete Visualisierung deutlich vereinfachen**. Die einzelnen Phasen des Datenanalyseprozesses sind in Abbildung 56 dargestellt. Da viele Schritte in diesem Prozess häufig angepasst und wiederholt ausgeführt werden müssen, sind kurze Verarbeitungszyklen sehr wichtig. Gleichzeitig werden typischerweise große Datenmengen verarbeitet. Um dennoch vertretbare Antwortzeiten des Analyseprozesses zu gewährleisten, ist es erforderlich, entweder durch geeignete Zerlegung des Problems in gut parallelisierbare Teilprobleme, durch geeignete Approximation oder Verwendung von Stichproben eine Implementierung des gegebenen Datenanalyseproblems zu finden, das effizient berechenbar ist. Die Anwendung von heutigen Data-Mining Algorithmen und Verfahren des maschinellen Lernens auf großen Datensätzen stellt jedoch eine große technische Herausforderung dar. Bisher lag der Fokus bei der Entwicklung von diesen Verfahren weniger auf parallelisierbaren und skalierbaren Verfahren sondern eher darauf qualitativ bessere Verfahren zu entwickeln. Tatsächlich setzen viele Verfahren Dirketzugriff (engl. random access) auf die Daten voraus und bieten folglich nur akzeptable Antwortzeiten, wenn alle Daten gemeinsam im Hauptspeicher vorgehalten werden. Deshalb stellt der Skalierungsaspekt (im Sinne von massiver Parallelisierung oder Verwendung von Stichproben) oftmals eine große Hürde dar. Ein Hauptproblem ist hierbei,

dass die entsprechenden Modifikationen des Grundalgorithmen, die für die Skalierbarkeit nötig sind, nicht trivial sind, und auch in dem üblichen Formalismus schwer abzubilden sind. Massive Parallelisierung hat sich für manche Lernmethoden dennoch als ein geeigneter Weg (80) herausgestellt, allerdings besteht hier noch großer Forschungsbedarf.

In diesem Abschnitt der Studie werden die einzelnen Phasen des Datenanalyseprozesses im Detail vorgestellt und die Herausforderungen und Möglichkeiten diskutiert.

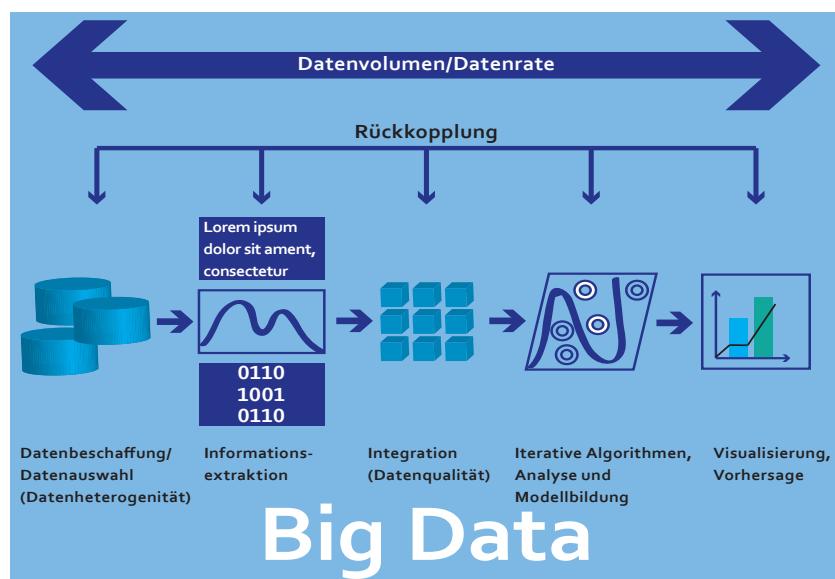


Abbildung 56: Der Datenanalyseprozess

Datenbeschaffung/Datenauswahl: Durch Webseiten, das Internet der Dinge, Sensordaten, unternehmensinterne Daten, Multimediadaten, Daten aus Maschinen-Maschinen-Kommunikation, etc. stehen heutzutage eine Vielzahl von Datenquellen für Big Data Anwendungen zur Verfügung. Die Speicherung großer Datenmengen stellt heutzutage aufgrund des technologischen Fortschritts und dem Preisverfall bei Datenspeichern keine Hürde mehr da.

Suchtechnologien für Daten werden notwendig: Die technologische Unterstützung bei der Suche nach Datensätzen ist gegenwärtig noch nicht hinreichend. Suchmaschinen und Webcrawlers indizieren in der Regel textuelle Information. Eine generelle Suche nach Datensätzen, wie z.B. DBpedia (265) Freebase (266), Wikidata (267) etc. gibt es jedoch noch nicht. Die Open Knowledge Foundation (OKF) stellt zwar mit Datahub eine Plattform bereit, um Daten zu suchen und zu veröffentlichen. Dies kann jedoch nur als erster Ansatz gelten, da für die Bereitstellung und Veröffentlichung der Daten, diese Daten aktiv veröffentlicht werden müssen, d.h. der Datensatz muss verlinkt und mit Metadaten versehen werden. Die Entwicklung von Suchmaschinen und

Webcrawlern, um strukturierte Datensätze auszuwählen, zu erkennen und unter Berücksichtigung der rechtlichen Rahmenbedingungen abzuspeichern, stellt daher eine wissenschaftliche und technologische Herausforderung dar, um auch wirtschaftlich das Potential von Big Data nutzen zu können.

Ausbau der Netze in Deutschland für Datenbeschaffung bei Big Data essentiell: Ein massiver Ausbau der Datennetze ist hingegen für die Beschaffung von Daten für Big Data Anwendungen notwendig, denn insbesondere bei Anwendungen, die eine hohe Datenaktualität erfordern oder Anwendungen, die sehr kurze Antwortzeiten von Analyseergebnissen benötigen ist dies essentiell. Im schlimmsten Fall kann durch die Latenz bei der Datenbeschaffung die Anwendung nicht realisiert werden. Zudem steigt laut einer Studie der EMC von 2010 (268) die pro Jahr erzeugte Datenmenge bis zum Jahr 2020 um den Faktor 44.

Datenmarktplätze können die Kosten für Datenbeschaffung und –auswahl signifikant reduzieren: Datenmarktplätze (siehe Abschnitt 3.4.1) können durch die zentrale Bereitstellung und Aufbereitung von strukturierten Daten und die Bereitstellung von Datensätzen für vielfältige Anwendungszwecke, die Kosten für Datenbeschaffung und –auswahl, die sonst die einzelnen Unternehmen tragen müssten, signifikant reduzieren.

3.1.1 Informationsextraktion

Durch das Internet steht heutzutage eine große Menge an Daten zu diversen Themengebieten in Form von Texten, Bildern und Videos zur Verfügung. Um diese jedoch in für Unternehmen nutzbare Informationen zu überführen sind maschinelle Verfahren unumgänglich. Die Informationsextraktion (204)(214) beschäftigt sich daher mit maschinellen Verfahren, die das Ziel haben die Identifikation und Extraktion von strukturierter Information aus Text-, Bild-, Audio- und Videodaten zu ermöglichen.

Sprachtechnologie und Technologien zur Extraktion von Informationen und Semantik aus Audio-, Video- und Bilddaten sind der Schlüssel zu Big Data. Laut dem „Strategischen Forschungsbericht für ein multilinguales Europa 2020“ der *Multilingual Europe Technology Alliance* ist insbesondere die Sprachtechnologie eine Schlüsseltechnologie für die nächste IT-Revolution und spielt insbesondere bei der Erschließung von Big Data eine zentrale Rolle (5).

Um Big Data sinnvoll nutzen zu können, reicht es daher nicht, die Daten einfach abzuspeichern und durchsuchbar zu machen, sondern diese müssen in der Regel erst erschlossen werden. Insbesondere bei der Erschließung von Textdaten, aber vielfach auch bei der Verarbeitung von Audio- und Videodaten, spielt die Sprachtechnologie eine zentrale Rolle:

Während es jedoch für die englische Sprache bereits eine gute Unterstützung für Textanalyse und Text und Sprachressourcen gibt, so gibt es für die deutsche Sprache nur eingeschränkte Unterstützung. Für die deutsche Sprache sind fast alle Texttechnologien weniger entwickelt als für die Englische (6).

Informationsextraktion aus Text

Ziel der Informationsextraktion (IE): Ziel der IE aus Text ist das gezielte Ableiten von potentiell relevanten Informationen, deren Strukturierung und deren semantische Analyse und Annotation. Informationsextraktion ist die Basistechnologie, um Informationen in Texten für die Analyse aufzubereiten, durchsuchbar zu machen oder mit anderen Datenquellen verknüpfen zu können. Die Informationsextraktion aus Dokumenten kann in drei aufeinander aufbauende Bereiche unterteilt werden: Dokumentenspezifische Verarbeitung, sprachspezifische Verarbeitung und domänen spezifische Verarbeitung. Zur dokumentenspezifischen Verarbeitung gehört die Analyse des Dokumentlayouts (Titel, Absätze etc.), die Identifikation und Separierung von Markup und Inhalt und die Erkennung von Strukturen, wie z.B. Tabellen oder Listen, um diese für die weitere Verarbeitung berücksichtigen zu können. Zur sprachspezifischen Verarbeitung gehört die Sprachenerkennung, die morphologische Analyse (wie z.B. Part-of-Speech-Tagging (POS), Wortstammanalyse, Lemmatisierung), die syntaktische Analyse (Parsing), das semantische Parsing („semantic role labeling“), die Erkennung von domänenübergreifenden Eigennamen (z.B. Zeitausdrücken, Datumsangaben, Orte) und die Extraktion von Relationen basierend auf sprachspezifischen Strukturen. Zur domänen spezifischen Analyse gehören u.a. die Erkennung von domänen spezifischen Eigennamen und Relationen und die Sentiment-Analyse.

OpenIE für n-äre Relationen: Informationsextraktionsverfahren sind inzwischen so weit fortgeschritten, dass diese Technologie auch zur Extraktion von komplexen Informationsstrukturen genutzt wird, wie zum Beispiel für die Extraktion von Protein-Protein-Interaktionen aus biomedizinischen Texten (vgl. BioCreative II Wettbewerb (269)) und die Extraktion von Eventinformation aus Nachrichtentexten (208). Bei solchen komplexen IE-Aufgaben steht das Schema der zu extrahierenden Information bereits fest, was für eine domänen spezifische Anwendung durchaus angemessen ist. Die aktuelle Forschung im Bereich IE geht jedoch darüber hinaus und beschäftigt sich inzwischen auch mit domänenübergreifenden IE-Verfahren, bei denen das Datenschema nicht bereits festgelegt ist, der sogenannten Open IE (209) (210) (211) (290). Open IE hat das Ziel domänenunabhängig und unüberwacht strukturierte Information (z.B. binäre Verbrelationen (209)) zu extrahieren. Open IE eignet sich daher insbesondere für große domänenübergreifende Textkorpora, wie z.B. Webseiten (210). Durch Clustering und Filtern kann dann spezifische Information identifiziert

werden (212). Aktuelle IE-Forschungsgebiete sind die Extraktion von komplexen (n-ären) und verschachtelten Relationen, vgl. (211) und die Extraktion von Informationen auf Basis von tiefer syntaktischer Analyse z.B. Dependenzbaumpfaden (278) oder Treekernels (290).

Generierung von Wissensbasen auf Basis komplexer IE: Ebenfalls ein aktuelles Forschungsthema ist die Transformation des extrahierten Wissens in komplexe semantische Informationsstrukturen, wie beispielsweise das automatische Erzeugen von Wissensbasen. Nicht nur in der biomedizinischen Domäne ist dies eine relevante Fragestellung (der aktuelle BioNLP shared task 2013 (270) beschäftigt sich z.B. mit der Extraktion von Events aus dem Genia-Korpus (207) in Verbindung mit der Erstellung einer Wissensbasis (siehe auch BioCreative IV (271)).

Seit der Einführung der Benchmark Task des National Institute of Standards and Technology (NIST) in den USA zum Thema „Knowledge Base Population (KBP)“ (291) ist dies auch eine zentrale Fragestellung in der Forschung zur Relationsextraktion allgemein geworden.

Verbesserung der Qualität durch Feedback zwischen NLP-Komponenten:

McCallum (272) weist darauf hin, dass bis dato viele NLP Verfahren eine Verarbeitungspipeline bilden, das heißt, dass z.B. die Ausgabe des POS-Taggers als Eingabe des Parsers dient und die Ausgabe des Parsers die Eingabe der Korefenzauflösung darstellt. Er macht zudem darauf aufmerksam, dass sich durch diese Pipelinearchitektur auch Fehler fortpflanzen und multiplizieren. Er schlägt daher vor, die Ergebnisse der einzelnen Schichten nicht gesondert zu betrachten, sondern als Feedback auf andere Schichten bzw. gemeinsame Schlussfolgerungen (engl. „joint inference“) zu ermöglichen. Die Idee der Joint Inference wird gegenwärtig von verschiedenen Forschern aufgegriffen und für unterschiedliche NLP-Bereiche angewendet, z.B. IE (273), Opinion Extraction (274), Ontology Matching (275), Temporale Relationsextraktion (294), Koreferenzresolution (295). Unter anderem wird hierbei auf probabilistische Verfahren, wie z.B. Markov-Ketten, Markov Logic Networks (296) oder Monte Carlo Methoden (276) zurückgegriffen.

Entwicklung skalierbarer NLP-Komponenten essentiell für Big Data: Bei der Entwicklung von NLP-Komponenten, wie sie für die IE die Basis darstellen, wurde lange Zeit hauptsächlich auf hohe Genauigkeit der Ergebnisse geachtet. Erst in jüngerer Zeit wird auch das Erreichen hoher Ausführungsgeschwindigkeit in den Vordergrund gestellt, siehe z.B. (277). Für Textanalysen auf großen Datenvolumen sind geringe Ausführungszeiten jedoch ein wichtiger Faktor um zeitnahe Analysen zu ermöglichen. Die meisten NLP-Komponenten sind zurzeit noch nicht für massiv parallele Ausführung optimiert.

Effiziente Verfahren zur Indizierung von annotierten Daten notwendig für Anwendungen: Die IE identifiziert nicht nur potentiell relevante Information

im Text, sondern erzeugt im Rahmen der sprachlichen Verarbeitung zusätzliche Information, beispielsweise Syntaxbäume von Sätzen, POS-Annotationen für Wörter oder Eigennamenbezeichner für Phrasen. Um diese Information gewinnbringend in Big Data Anwendungen nutzen zu können, bedarf es der Entwicklung von effizienten, skalierbaren Indizierungsverfahren, wie es beispielhaft für Frage-Antwortssysteme mithilfe von semantischem Parsing bereits für das Englische demonstriert worden ist (292).

Kostenreduktion durch gezielte Auswahl bei Erstellung von Trainingsdaten:

Bereits mit relativ einfachen Verfahren, wie z. B. dem Anwenden von manuell geschriebenen regulären Ausdrücken lassen sich gezielt bestimmte Informationseinheiten in Texten finden und extrahieren, z.B. die Erkennung von Zeitausdrücken (279) oder Proteinnamen (205). Der Nachteil dieser Verfahren ist der manuelle Aufwand, um die entsprechenden Regeln zu erstellen. Ansätze, die diese regulären Ausdrücke mit maschinellen Lernverfahren erlernen, existieren jedoch bereits (siehe z.B. (206)). Im Gegensatz zu regelbasierten Systemen werden für die Eigennamenerkennung aber auch überwachte maschinelle Lernverfahren eingesetzt, die nicht Extraktionsregeln (reguläre Ausdrücke) lernen, sondern auf Basis von Trainingsdaten Textmuster für die Extraktion erlernen. In der Regel erfordern diese jedoch zumeist Trainingsdaten, in denen die relevanten Informationen manuell annotiert worden sind. Um die Menge der manuell annotierten Daten gering zu halten, werden vielfach halbüberwachte Lernverfahren verwendet (301). Bei diesen Lernverfahren werden zusätzlich zu den manuell annotierten Daten ungelabelte Daten verwendet. Durch Bootstrapping wird versucht, Klassenlabel von den manuell annotierten Daten auf die ungelabelten Daten iterativ zu übertragen. So vergrößert sich die Menge der Trainingsdaten und idealerweise auch die Qualität der darauf trainierten Extraktionssysteme/Klassifikatoren. Eine Sonderform des halbüberwachten Lernens ist das sogenannte Active Learning (280), bei dem zwar Untermengen der ungelabelten Trainingsdaten manuell annotiert werden müssen, diese aber durch eine geschickte Methode sehr gering gehalten werden. Eine verwandte Lernmethode, die zunehmend an Bedeutung gewinnt, ist das sogenannte Distant Supervision (293). Dieses Lernverfahren, das aus dem Bereich der „Knowledge Base Population“ stammt, verlangt als Eingabe lediglich prototypische Relationstupel für die Relationstypen, für die Klassifikatoren gebaut werden sollen, z.B. <Wolfgang Amadeus Mozart, 27. Januar 1756> für den Relationstyp „geboren-am“. Die Sätze, in denen nun solche Tupel vorkommen, werden als (pseudo-) gelabelte Trainingsdaten verwendet, auf denen Klassifikatoren trainiert werden können, die auf neuen Textdaten weitere bislang unbekannte Relationstupel der gewünschten Relationstypen erkennen können. Mittlerweile sind auch diverse Verfahren erprobt worden, um das inhärente Rauschen dieser Trainingsdaten zu minimieren (297).

Informationsextraktion aus Audiodaten

Die Anwendungsgebiete der Informationsextraktion aus Audiosignalen sind vielfältig und umfassen u.a. die Komprimierung von Audiodaten durch gezielte Auswahl von Informationen, die Extraktion von Features z.B. für Spracherkennung oder Sprachsynthese oder die Verbesserung der Qualität des Signals durch intelligente Informationsfilterung.

Verbesserung der Qualität von IE auf einzelnen Signalen durch multimodale Signalverarbeitung: Bei der multimodalen Signalverarbeitung wird die IE auf einzelne Aspekte des Audiosignals angewendet. Die Zusammenführung der individuellen Ergebnisse kann unter Umständen die Qualität des Gesamtergebnisses verbessern bzw. die Verlässlichkeit der Ergebnisse erhöhen. Polzehl et al. (281) haben beispielsweise gezeigt, wie die Kombination aus linguistischen und akustischen Merkmalen die Erkennung von Wut/Ärger in Audiosprachsignalen eine verbesserte Erkennungsleistung liefert.

Schnellere Signalverarbeitung durch Verarbeitung komprimierter Signale: Die Informationsextraktion geschieht in der Regel auf den unkomprimierten Signalen. In Big Data Szenarien mit großem Datenvolumen ist die Dekomprimierung ein Zeitfaktor, der je nach Volumengröße des Datensatzes die Verarbeitungszeit nicht unerheblich verlängern kann. Eine Forschungsfrage ist daher, ob sich Informationsextraktionsverfahren auf komprimierten Audiodateien realisieren lassen, bzw. zu ermitteln, ob einige Komprimierungsalgorithmen spezielle Eigenschaften besitzen, die sich für die Informationsextraktion auf unkomprimierten Daten besonders eignen.

Kostenreduktion durch automatisches Erstellen sehr großer Audiodatensammlungen: Sammlungen von Audiodaten sind die Basis für vielfältige Anwendungen, z.B. Entwicklung von Sprachenerkennern, Sprachsyntheseprogrammen oder für die automatische Emotionserkennung. Bisher wurden diese Datensammlungen manuell erstellt und zum Teil aufwendig durch Transkriptionen und Metadaten erweitert. Durch das Internet stehen jedoch sehr große Mengen an Multimediadaten zur Verfügung, z.B. Radiosendungen mit Texttranskripten des Inhalts. Eine interessante Forschungsfrage ist nun, ob sich diese sehr großen Multimediadaten nutzbar machen lassen, um automatisch Datensammlungen erstellen lassen, die zuvor festgelegte Kriterien erfüllen.

Informationsextraktion aus Videodateien

Datenbanken mit Medieninhalten wie Video- und Audiosignalen sowie zugehörigen Metadaten befinden sich in einem rasanten Wachstum und umfassen bereits jetzt viele Petabytes.

Ausnutzung multimodaler Korrelationen: Videodateien setzen sich zusammen aus mehreren Modalitäten (Audioinformation, Bildinformation, Metadaten). Die Ausnutzung der verschiedenen Modalitäten und die Untersuchung und

Entwicklung von Verfahren für eine multimodale Videodatenanalyse (Text, Metadaten, Sprache, Bilder), die insbesondere Korrelationen zwischen den Modalitäten nutzt, ist eine Herausforderung und sollte daher für zukünftige Forschung in diesem Bereich einen Schwerpunkt bilden.

Skalierbare Verfahren für multimodale Datenanalyse: Da es sich bei Videodaten um multimodale Daten handelt, bietet es sich an die Informationsextraktion in den einzelnen Modalitäten zu parallelisieren. Insbesondere in Big Data Szenarien mit hohen Datenvolumen kann die parallele Ausführung der Datenanalysen signifikant die Verarbeitungsgeschwindigkeit steigern. Die Entwicklung von skalierbaren Verfahren für die multimodale Datenanalyse stellt daher einen weiteren möglichen Forschungsschwerpunkt dar.

Videomining auf komprimierten Daten: Auch bei der Informationsextraktion aus Videos ist die Untersuchung von Verfahren, die auf komprimierten Daten arbeiten können eine sinnvolle Forschungsfrage, da durch die zeitliche Ersparnis, gerade bei Verarbeitung von sehr großen Datenvolumina eine signifikante Reduktion der Verarbeitungszeit zu erwarten ist.

Automatische Erstellung von Videosammlungen für vielfältige Anwendungsbereiche: Videodaten lassen sich für vielfältige Forschungs- und Anwendungszwecke einsetzen. Die manuelle Suche nach geeigneten Videos ist jedoch kostenintensiv. Verfahren, die automatisch Videokorpora zusammenstellen, können daher eine Kostenreduktion erzielen. Darüber hinaus können auch qualitative Verbesserungen von Extraktionsverfahren erwartet werden, da wesentlich mehr Testdaten zur Verfügung stehen.

3.1.2 Datenintegration

Unter der Datenintegration versteht man die Zusammenführung von Daten aus verschiedenen Datenquellen (304).

Entwicklung von Methoden zur Bewertung der Korrektheit von Daten aus unterschiedlichen Quellen: Bei der Interpretation, Analyse und Aggregation von Daten kann die Bewertung der Korrektheit von Daten eine entscheidende Rolle spielen. Jedoch sind solche Bewertungen auch immer mit einer Unsicherheit versehen. Für die meisten Big Data Anwendungen ist es daher eine Notwendigkeit mit diesem Unsicherheitsfaktor umzugehen. Die Herausforderung bei Big Data besteht darin, dass potentiell eine sehr große Menge von relevanten Datenquellen zur Verfügung steht und ausgewertet werden muss. Welche Faktoren dabei helfen können die Korrektheit der Daten zu bewerten, ist jedoch noch unklar. Insbesondere durch Aggregation über viele Daten könnte die Qualität des Analyseergebnisses verbessert werden. Auch

Datencharakterisierung durch Profiling könnte hierbei hilfreich sein. Insgesamt besteht in diesem Bereich jedoch noch großer Forschungsbedarf.

Entwicklung von skalierbaren Algorithmen zur Duplikatserkennung: Ein wichtiger Schritt bei der Integration von Daten ist die Erkennung von Duplikaten (305), da diese nicht nur unnötigen Speicherplatz verbrauchen, sondern im schlimmsten Fall auch die Datenanalyseergebnisse verfälschen können. Für Big Data Anwendungen besteht eine große Herausforderung in der Entwicklung von skalierbaren Algorithmen zur Duplikatserkennung.

Kostenreduktion bei der Integration von Daten durch Datenmarktplätze: Datenmarktplätze (siehe Abschnitt 3.4.1) stellen extrahierte und integrierte Daten zentralisiert bereit und können somit die Kosten für einzelne Unternehmen, die diese Daten dann integriert und bereinigt beziehen können, signifikant senken.

3.1.3 Datenanalyse und Modellbildung

Big Data Analysen haben üblicherweise zum Ziel Informationen zu gewinnen, um bestehende Prozesse zu optimieren oder neue Prozesse zu ermöglichen. Dazu werden statistische Modelle mit maschinellen Lernverfahren berechnet oder versteckte Informationen mittels Data-Mining Algorithmen identifiziert und extrahiert. So können zum Beispiel Datenmodelle erstellt werden, die Produktempfehlungen für Kunden generieren, die Kunden identifizieren, die möglicherweise planen in naher Zukunft einen Vertrag zu kündigen, oder die Kreditanwärter gemäß eines möglichen Ausfallrisikos klassifizieren. Kunden können zu Marketingzwecken in Gruppen eingeteilt werden, um zielgerichteter zu werben, gute Standorte für beispielsweise Car-Sharing Parkplätze oder Ladestationen für Elektrofahrzeuge zu identifizieren oder kraftstoffsparende Routen für Logistiktransporte zu planen. Die hier vorgestellten Anwendungsfälle sind nur wenige Beispiele für die zahllosen Anwendungsmöglichkeiten solcher Analysen. Eine wichtige Voraussetzung ist jedoch immer die Verfügbarkeit der relevanten Daten in ausreichender Menge und Qualität, die für eine Analyse benötigt werden. Diese Daten können zum Teil selber gesammelt werden wie zum Beispiel GPS-Routenaufzeichnungen oder auf Datenmarktplätzen (z.B. demographische Daten) erworben werden. Der große Wert von vielen Datensätzen, der durch die richtigen Analysen verfügbar wird und so deutliche Optimierungen ermöglichen kann, wird derzeit in Deutschland häufig noch unterschätzt. Das begründet sich auch aus dem geringen Wissen über die Chancen und Möglichkeiten die komplexe Datenanalysen bieten.

Erforschung von maschinellen Lernverfahren für heterogene Daten: Die meisten verfügbaren Datenanalyseverfahren setzen strukturierte Daten als Eingabe voraus. Um diese Verfahren auf unstrukturierte und heterogene Daten

anwenden zu können, müssen zunächst strukturierte Informationen, sogenannte Features, aus diesen Daten extrahiert werden. Es ist wichtig, dass diese Features die Merkmale der Daten, die für die jeweilige Analyse relevant sind, besonders gut repräsentieren. Da die Ausführungszeit von maschinellen Lernverfahren und Data-Mining Algorithmen häufig stark von der Anzahl der betrachteten Features abhängt, sollten nur die ausdruckstärksten Features für eine Analyse genutzt werden. Es gibt Verfahren, die eine große Anzahl von Features in eine kleinere Menge transformieren und dabei die Aussagekraft der ursprünglichen Informationen nur minimal reduzieren (z.B. Chi-Square Feature Selection oder Principle-Component-Analysis (306,124)). Diese Vorverarbeitungsschritte sind ein essenzieller Bestandteil der Datenanalyse und sehr wichtig für ihren Erfolg. Die Auswahl der richtigen Features ist eine schwierige Aufgabe und erfordert häufig Domänenwissen. Der Auswahlprozess ist von zahlreichen manuellen Anpassungen und Versuchen geprägt.

Die direkte Analyse von heterogenen Daten, wie zum Beispiel die Analyse von Text und Audiosignalen, ohne initiale Extraktion von Features ist bisher noch wenig untersucht, bietet jedoch das Potential einer qualitativen Verbesserung der Ergebnisse. Ein weiterer Forschungsaspekt ist die Selbstdadaptation dieser Verfahren.

Skalierbare Verfahren für Maschinelles Lernen: Die Entwicklung von neuen maschinellen Lernverfahren oder Data-Mining Algorithmen hatte bisher üblicherweise eine qualitative Verbesserung der Ergebnisse oder eine Laufzeitoptimierung auf einem einzelnen Rechner zum Ziel. Parallelisierungsaspekte, insbesondere die Parallelisierung über Rechnergrenzen hinweg, spielen eine eher untergeordnete Rolle. So setzen derzeit die meisten Verfahren voraus, dass alle Daten im Hauptspeicher gehalten werden und dass auf jedes Datum jeder Zeit wahlfrei zugegriffen werden kann. Eine naive Übersetzung eines Algorithmus mit diesen Eigenschaften auf eine massiv-parallele Ausführungsumgebung in einem Rechencluster, würde keine befriedigenden Ergebnisse liefern. Da der Zugriff auf den Hauptspeicher eines anderen Rechners über das Netzwerk sehr viel langsamer ist, als der direkte Zugriff auf den eigenen Hauptspeicher, kann ein einzelner Rechner unter Umständen einen solchen Algorithmus schneller ausführen als mehrere zusammengeschaltete Rechner des gleichen Typs.

Entwicklung neuartiger Datenanalysesysteme: Zur Spezifikation und Verarbeitung von Datenanalysealgorithmen sind existierende Systeme nur bedingt geeignet. In Abschnitt 3.3. werden neuartige Datenanalysesysteme beschrieben und Forschungsthemen im Bereich dieser Systeme aufgezeigt.

Kostenreduktion durch neue Hardwaretechnologien: Für die Analyse von Big Data (große Datensätze, zeitkritische Analyse, heterogene Daten) sind insbesondere Verfahren des maschinellen Lernens prädestiniert, die durch massive

Parallelisierung effizient sind. Neue Entwicklungen bei Hardware-Technologien (siehe Abschnitt 3.3.1) unterstützen die massive parallele Berechnung, so dass es voraussichtlich zu vergleichsweise geringeren Verarbeitungszeiten kommen kann.

Im Zeitalter von Big Data sind einzelne Maschinen jedoch häufig aufgrund von hohen Datenraten und –volumen nicht mehr ausreichend. Sobald ein Datensatz die Größe des Hauptspeichers übersteigt, wird ein neuer oder aufgerüsteter Rechner benötigt. Im Gegensatz dazu können parallele Systeme einfach um einen oder mehrere Rechenknoten erweitert werden, unter Umständen sogar kurzfristig und vorübergehend mit Infrastructure-as-a-Service Angeboten wie zum Beispiel Amazon EC2. Mittlerweile gibt es einige Datenverarbeitungssysteme, die Schnittstellen anbieten um Datenanalyseprogramme zu definieren und diese massiv-parallel auszuführen. Diese Schnittstellen setzen allerdings voraus, dass die Algorithmen in einer bestimmten Form definiert werden, die sich gut parallelisieren lässt. Die Entwicklung von neuen maschinellen Lernverfahren oder Übersetzung von existierenden Algorithmen auf diese Schnittstellen ist eine große Herausforderung. Erste Ansätze für skalierbare maschinelle Lernverfahren und Data-Mining Algorithmen, wie beispielsweise die Mahout Bibliothek (230) für das Map/Reduce Datenverarbeitungssystem Hadoop, existieren zwar bereits, sind aber noch nicht ausgereift und durch die Fähigkeiten der momentan verfügbaren parallelen Ausführungssysteme beschränkt. So sind die Schnittstellen und Ausführungsstrategien der meisten Verarbeitungssysteme nicht im Hinblick auf komplexe maschinelle Lernverfahren entwickelt worden. In Abschnitt 3.3 werden neuartige Datenanalysesysteme beschrieben und Forschungsthemen im Bereich dieser Systeme aufgezeigt. Neue Entwicklungen bei Hardware-Technologien (siehe Abschnitt 3.3.1) können die massive parallele Berechnung unterstützen, so dass voraussichtlich eine deutliche Reduktion der Verarbeitungszeiten erzielt werden kann. Insgesamt gibt es in diesem Bereich ein sehr großes und vielfältiges Forschungspotential, sowohl auf Seiten der parallelen Datenverarbeitungssysteme als auch auf der Seite der Datenanalysealgorithmen, die auf die neuen Systeme und ihre Eigenschaften angepasst werden müssen.

3.1.4 Visualisierung von Analyseergebnissen und Vorhersagen

Die Visualisierung von Analyseergebnissen inklusive der Darstellung von Vorhersagen (wie z.B. Trends) durch geeignete Visualisierungsmethoden ist ein sehr wichtiger Aspekt der Datenanalyse. Da bei der Anwendung komplexer Analyseverfahren, wie Data-Mining oder maschinellem Lernen, die Auswahl der richtigen Features und Parametereinstellungen sehr schwierig ist, werden hierfür üblicherweise viele Versuche benötigt. Die gewählten Features und Parametereinstellungen können auf Basis des berechneten Analyseergebnisses bewertet werden. Eine gute Visualisierung des Ergebnisses kann dabei helfen eine

schnelle und genaue Bewertung der Ergebnisqualität vorzunehmen. So kann zum Beispiel die Qualität eines Vorhersage- oder Klassifikationsmodells gegen einen Testdatensatz validiert werden. Fehlerhafte Klassifizierungen oder Vorhersagen helfen dem Analysten die Featureauswahl oder Parametereinstellungen für das Analyseprogramm zu verbessern. Ohne unterstützende Visualisierung kann dieser Prozess sehr aufwändig und langwierig sein. Deshalb ist die Visualisierung von Analyseergebnissen ein wichtiger Bestandteil der des Datenanalyseprozesses. Doch auch für die spätere Anwendung des Analyseergebnisses sind Visualisierungstechniken wichtig, damit Analysten korrekte Schlussfolgerungen aus den Ergebnissen ziehen können. So können sehr große Analyseergebnisse ohne technische Hilfe gar nicht oder nur mit großem Aufwand ausgewertet werden. Gute Visualisierungen können diesen Prozess erheblich erleichtern. Abhängig von der Komplexität des Visualisierungsverfahrens, können sehr große Analyseergebnisse auch einen erheblichen Berechnungsaufwand bedeuten und so eine interaktive Bedienung erschweren.

Explorative Suche in Analyseergebnissen ist essentiell für Big Data: Auch wenn je nach Analyseanfrage die extrahierten Datenmengen unabhängig von der Menge der zu durchsuchenden Daten sehr klein sein können (z.B. wird ein Smart-Home-System für ein Einfamilienhaus in der Regel nur eine sehr kleine Anzahl von Fenstern zum Öffnen vorschlagen, auch wenn zur Berechnung des Vorschlags potentiell sehr viele Daten, z.B. Sensordaten, Daten von Wetterstationen, etc. verwendet werden können. Jedoch sind gerade im Big Data Kontext, z.B. Webmining auch Ergebnisse zu erwarten, die ein sehr hohes Volumen an Ausgabedaten produzieren und sich somit nicht mehr durch statische Visualisierungsverfahren darstellen lassen und eine interaktive explorative Datenanalyse erforderlich machen (217) (218). Shneiderman (219) definiert für explorative Analysen die folgenden 7 Aufgabentypen: overview, zoom, filter, details-on-demand, relate, history und extract.

Visualisierung von Datenmodellen: Viele maschinellen Lernverfahren berechnen Datenmodelle, die sich von den Ergebnissen traditioneller Reportinganfragen und Data-Mining Algorithmen deutlich unterscheiden. Datenmodelle können zur Klassifikation von Daten, Generierung von Empfehlungen oder für Vorhersagen eingesetzt werden. Um die Qualität eines Datenmodells zu bestimmen, werden diese üblicherweise auf Testdaten angewendet und die Ergebnisse validiert. Eine Visualisierung dieser Testergebnisse kann äußerst hilfreich bei der Konfiguration der Lernverfahren sein. Aufgrund der unterschiedlichen Anwendungszwecke von Datenmodellen, sind hierfür jeweils andere Visualisierungstechniken vorteilhaft. Eine weitere Herausforderung ist die Darstellung von vielen Datenpunkten, die sich aus großen Testdatensätzen ergeben.

Schnellere und komplexere Visualisierungen durch neue HW-Technologien: Die Entwicklung im Bereich der Visualisierungsmethoden der letzten Jahrzehnte hat gezeigt, dass die zur Verfügung stehenden Methoden und Verfahren stark abhängig sind vom technologischen Stand. Während in den frühen 70er Jahren Datenvisualisierung nur sehr rudimentär möglich war (z.B. Tabellen, Balkendiagramme.) und diese zumeist nur auf monochromen Bildschirmen dargestellt werden konnten, sind heutzutage detailreiche interaktive 3D-Visualisierungen möglich. Diese Entwicklung steht im Zusammenhang mit der technologischen Entwicklung im Bereich von Bildschirmen, schneller Verarbeitung durch verbesserte Prozessoren und insbesondere der technologischen Entwicklung bei Grafikprozessoren. Wie in Abschnitt 3.3.1 beschrieben wird, ist bei CPUs und GPUs zukünftig mit wesentlich mehr Kernen pro Einheit zu rechnen, so dass ausdrucksstärkere und aufwändigere Methoden zur Datenvisualisierung möglich werden. Ein weiterer Grund für höhere Effizienz bei der Berechnung von Visualisierungen ist auch die ansteigende Größe von Analyseergebnissen.

Vorteile durch präzisere Vorhersagemodelle: Big Data bietet die Chance durch die Zusammenführung von sehr vielen Datenquellen und die Auswertung von sehr großen Datenmengen präzisere Datenmodelle zu entwickeln. Für Unternehmen, die zum Beispiel Kaufempfehlungen generieren oder aufgrund von Vorhersage- oder Klassifikationsmodellen Entscheidungen treffen, kann dies zu erheblich effizienteren und effektiveren Prozessen und in Konsequenz zu reduzierten Kosten oder steigenden Einnahmen führen. Das steigende Interesse an Big Data Analytics lässt erwarten, dass die damit verbundenen Chancen immer häufiger genutzt werden.

3.1.5 Herausforderungen und Forschungsfragen

Wir gliedern und priorisieren Forschungsfragen der Datenanalyse nach „Erforschung neuer Funktionalität“, „Qualitative Verbesserung existierender Funktionalitäten“ und „Optimierung der Ausführungskosten“ bei gleicher Qualität einer existierenden Funktionalität.

Erforschung neuer Funktionalität

- Wie können Analysten bei der Suche und Auswahl von relevanten Datensätzen unterstützt werden?
- Wie können Audio-/Videodaten genutzt werden, um die Entwicklung von Systemen wie z.B. Spracherkennungs- oder Sprachsynthesesystemen zu unterstützen und zu automatisieren?
- Wie können Audio- und Videosammlungen automatisch für spezifische Anwendungszwecke erstellt werden?
- Wie kann die Korrektheit von Daten aus unterschiedlichen Quellen automatisch bewertet werden?
- Wie können Qualitätsbewertungen bei der Datenanalyse berücksichtigt werden?
- Wie können Datenanalysealgorithmen auf heterogenen Datensätzen angewendet werden?
- Welche Visualisierungstechniken können Analysten bei der explorativen Datenanalyse unterstützen?

Qualitative Verbesserung existierender Funktionalitäten

- Wie können syntaktische und semantische Informationen effizient indiziert werden?
- Wie können neue IE-Komponenten durch minimal überwachte Verfahren effizient erstellt und qualitativ verbessert werden?
- Wie kann multimodale Signalverarbeitung zur Informationsextraktion eingesetzt werden, um die Qualität der Informationsextraktion zu verbessern?
- Wie kann der iterative Prozess der Spezifikation von komplexen Datenanalyseprogrammen unterstützt, vereinfacht und beschleunigt werden?
- Wie kann der Prozess der Erstellung von Datenmodellen zur Vorhersage oder Klassifikation durch geeignete Visualisierung verbessert werden?

Optimierung der Ausführungskosten

- Wie kann die Ausführung von NLP-Komponenten und Pipelines effektiv parallelisiert werden?
- Wie kann IE direkt auf komprimierten Daten (Audio/Video) angewendet werden?
- Wie können maschinelle Lernverfahren (Clustering, Klassifikation, Regression, etc.) für die Anwendung auf sehr große Datenmengen optimiert werden?
- Welche Technologien werden benötigt, um sehr große Analyseergebnisse interaktiv zu visualisieren?

3.2 System- und prozessübergreifende Technologien

Der vorherige Abschnitt beschrieb Methoden und Algorithmen für die Analyse und Identifikation von häufig unbekannten Datenverteilungen in einem interaktiven, iterativen Prozess. Dieser Prozess erfordert die Lösung von drei technischen Herausforderungen: (1) Menschen müssen das gewünschte Ergebnis der Anfragen in einer Hochsprache beschreiben können, (2) die Technologie muss iterative Datenströme verarbeiten und (3) die Technologie muss auch unbekannte Programme dritter Anbieter, sogenannte benutzerdefinierte Funktionen (user-defined functions (UDFs)) ausreichend schnell verarbeiten können. In diesem Abschnitt beschreiben wir den Stand der Technik und die technologischen Herausforderungen.

3.2.1 Programmiermodelle für die breite Masse von Analysten

Die deklarative Spezifikation von Datenanalyseprogrammen bedeutet, dass die Beschreibung des Analyseproblems von seiner Ausführung entkoppelt wird (z.B. die Anfragesprachen SQL für relationale Daten, XPath und XQuery (79) für XML-Daten, oder SparQL für RDF-Daten, Linked-Open-Data im NoSQL-Bereich). Für Big Data gibt es derzeit keine derartige deklarative Sprache, die auf beliebigen Rechnersystemen skaliert und unabhängig von Daten, deren physischer Speicherung und statistischen Eigenschaften automatisch optimierbar und parallelisierbar ist. Dies wurde von den führenden US-amerikanischen Big Data Forschern als eines der größten Forschungsprobleme identifiziert (61). Die Spezifikation von Algorithmen zur Analyse von Big Data erfordert neben den klassischen Operationen der relationalen Algebra auch die Verwendung von komplexen benutzerdefinierten Funktionen sowie die Spezifikation von iterativen Algorithmen. Erst dadurch können beispielsweise Verfahren der linearen Algebra, des maschinellen Lernens, der Sprach-, Video- und Signalverarbeitung in die Datenanalyse integriert werden. Für die Spezifikation und Verarbeitung von benutzerdefinierten Funktionen über Datensätzen oder Gruppen von Datensätzen wurde das Map/Reduce Paradigma für Big Data eingeführt (85), ein funktionales Programmiermodell, das aus zwei Funktionen zweiter Ordnung besteht, wobei beliebige Funktionen erster Ordnung zur Transformation oder Selektion von Datensätzen (map) sowie zur Aggregation (reduce) verwendet werden können. Das im Big Data Bereich populäre System Hadoop (154) implementiert dieses Programmiermodell und parallelisiert Analysen automatisch auf großen Rechenclustern. Allerdings unterstützt das Map/Reduce-Modell nicht alle Operationen der relationalen Algebra (z.B. Verbünde). Ferner unterstützen weder Hadoop noch darauf aufbauende Systeme wie zum Beispiel Pig, Hive, Hadoop und iteratives Map/Reduce die deklarative Spezifikation oder die automatische Optimierung und Parallelisierung von iterativen, zustandsbehafteten

Algorithmen. Diese müssen bislang durch Einsatz spezialisierter Tricks wie verteilten Caches sowie außerhalb des Programmiermodells durch die Kopplung von Map/Reduce Phasen durch Treiberprogramme realisiert werden. Dadurch geht der größte Vorteil des Map/Reduce Modells verloren, da effektiv innerhalb einer map oder reduce Funktion nun verteilte Algorithmen oder sogar ganze Komponenten eines Datenbanksystems (z.B. Hash-Join und Merge-Join Verfahren und manchmal sogar Teile eines Anfrageoptimierers) implementiert werden müssen, was die Erstellung von Datenanalyseprogrammen aufwändig, fehleranfällig, datenabhängig und oftmals ineffizient und teuer macht.

In den letzten Jahren haben aus dem Data-Mining Bereich kommende Methoden zur **effizienten Beschreibung von Datenströmen** wie z.B. die Count-Min Sketches (82) zu interessanten Ansätzen wie den sogenannten Hash Kernels geführt (144), mit denen sich große Datenmengen und auch große Merkmalsräume geeignet kompakt beschreiben lassen. Hier lässt sich bereits ein zarter Trend zur Konvergenz von Datenbankmethoden und maschinellem Lernen erkennen. In komplexen Anwendungsfeldern zeigt sich ein Trend zur vermehrten Berücksichtigung von heterogenen Strukturen (Graphen, Bäume etc.) und multimodaler Methoden zur Fusion heterogener Daten (Multiples Kern Lernen (MKL)). Das verstärkte Einbeziehen solcher Strukturen führt zwar zu deutlich verbesserten Modellierungsergebnissen, jedoch ist die algorithmische Skalierbarkeit eingeschränkt, da Parallelisierungsaspekte bei der Entwicklung von Algorithmen zum maschinellen Lernen bislang kaum beachtet wurden. Ferner wird numerische Stabilität nicht immer bei der Implementierung von skalierbaren Algorithmen des maschinellen Lernens auf Hadoop betrachtet (z.B. Mahout; numerisch stabile Algorithmen werden z.B. in (150) diskutiert). So kann die Reihenfolge, in der Daten verarbeitet werden, oder die Wahl einer Ausführungsstrategie einen Einfluss auf Rundungsfehler des Endergebnisses einer Rechnung haben. Insbesondere wenn Datenanalyseprogramme automatisch optimiert werden, kann dies im Extremfall zu fehlerhaften Analyseergebnissen führen.

Forschungsziel: Deklarative Spezifikation und Optimierung iterativer Algorithmen. Es gibt neben Hadoop eine Vielzahl von interessanten Forschungsarbeiten zur massiv-parallelen Verarbeitung von datengetriebenen iterativen Algorithmen (70) (89) (103) (111) (116) (146). Allerdings adressiert keine dieser Entwicklungen die deklarative Spezifikation und automatische Optimierung von iterativen Algorithmen. Daher erfordert die Analyse von Big Data derzeit Kenntnisse in der Programmierung von (verteilten) Systemen zusätzlich zur Kenntnis der Analysedomäne sowie den Kenntnissen der maschinellen Lernverfahren. Personen, die diese Kombination an Kenntnissen mit sich bringen, sind rar (114). Die Überwindung dieses Engpasses wird der kritische Erfolgsfaktor, nicht nur für neue Big Data-Technologien, sondern auch für die breite Anwendung von Big Data Analytics sein.

3.2.2 Iterative Datenstromverarbeitung

Die effiziente, parallele Verarbeitung von iterativen Datenflüssen birgt viele technische Herausforderungen, von denen die Wichtigsten in diesem Abschnitt diskutiert werden.

Iterative Analysen, verteilter Zustand und geringe Latenz: iterative Datenanalyseverfahren berechnen üblicherweise in vielen einzelnen Schritten das Ergebnis der Analyse. In jedem Schritt wird dazu ein Zwischenergebnis oder Zustand berechnet und aktualisiert. Da diese Berechnungen aufgrund des Datenvolumens parallel ausgeführt werden müssen, muss der Zustand effizient über viele Rechner verteilt, gespeichert und verwaltet werden. Aus Effizienzgründen ist es notwendig den Zustand im Hauptspeicher zu halten. Viele Algorithmen benötigen außerdem sehr viele Iterationen, um das endgültige Ergebnis zu berechnen. Deshalb ist es sehr wichtig, dass einzelne Iterationen mit einer geringen Latenz berechnet werden können, um die gesamte Anfragezeit zu minimieren. In einigen Anwendungsfällen reduziert sich der Berechnungsaufwand erheblich von der ersten Iteration zur letzten Iteration. Batch-basierte Systeme wie Map/Reduce (84) und Spark (254) führen in jeder Iteration immer alle Berechnungen aus, auch wenn sich das Ergebnis einer wiederholten Berechnung nicht ändert. Im Gegensatz dazu können echte iterative Datenflusssysteme wie Stratosphere (146) oder spezielle Graphverarbeitungssysteme zum Beispiel GraphLab (103) und Google Pregel (111) diese Eigenschaft ausnutzen und reduzieren den Berechnungsaufwand mit jeder Iteration.

Pipelining und Materialisierung: Eine große Herausforderung bei der Verarbeitung von Datenströmen sind hohe Datenraten. Datenströme mit hoher Datenfrequenz erfordern, dass die eintreffenden Daten sehr schnell verarbeitet werden müssen, um einen Datenrückstau zu vermeiden. Dazu ist es häufig notwendig, dass Daten schnell weitergereicht werden können, ohne auf andere Daten zu warten. Sobald ein Verarbeitungsschritt abgeschlossen ist, werden die Daten zum nächsten Verarbeitungsschritt weitergereicht. Diese Eigenschaft eines Programms wird Pipelining genannt und hängt von den verwendeten Algorithmen ab. Alternativ können Programme auch Daten ansammeln und zwischenspeichern, um sie zu verarbeiten. In diesem Fall spricht man von Materialisierung. Niedrige Anfragelatenz ist häufig nur für Analyseprogramme möglich, die mit Pipelining arbeiten.

Fehlertoleranz für iterative Analysen: Bei allen parallelen Systemen ist der Umgang mit Fehlern ein sehr wichtiges Thema. Der Ausfall von Rechenknoten durch Hardware- oder Softwarefehler ist ein häufiges Ereignis und kann zur Folge haben, dass ein Teil des berechneten Ergebnisses verloren geht. Ein solcher Ausfall sollte von parallelen Systemen verkraftet werden können, ohne dass die Verarbeitung eines laufenden Analyseprogramms abgebrochen werden muss.

Ein übliches Verfahren hierfür sind regelmäßige Sicherungspunkte (Checkpoints) aus denen ein zuvor abgeschlossener Berechnungsstand im Fehlerfall wiederhergestellt werden kann. Dieses Verfahren ist vergleichsweise einfach umzusetzen und sehr häufig anwendbar. Nachteilig ist der hohe Aufwand zum Synchronisieren aller Rechenknoten und zum Erstellen eines Sicherungspunktes, da dieser repliziert auf ein persistentes Medium geschrieben werden muss. Dadurch entstehen im Regelbetrieb auch ohne Fehlerfall hohe Ausführungskosten. Alternative Verfahren sind optimistischer und können für bestimmte iterative Analysealgorithmen ein unvollständiges Zwischenergebnis wieder in einen gültigen Zustand überführen mit dem die Berechnung fortgeführt werden kann. Für dieses Verfahren entstehen nur im Fehlerfall zusätzliche Kosten. Allerdings ist es nicht für jeden Algorithmus anwendbar und erfordert algorithmen-spezifischen Programmcode.

3.2.3 Verarbeitung von Benutzerdefinierten Funktionen mit geringer Latenz

Benutzerdefinierte Funktionen ermöglichen die Integration von externer Ausführungslogik, beispielsweise für semantische Textanalyse in Datenanalyseprogrammen. Wünschenswert sind Erweiterungen für die Verarbeitung von benutzerdefinierten Funktionen und die Optimierung von Datenflüssen, insbesondere unter der Wiederverwendung von häufig verwendeten Teilausdrücken.

Stand der Technik. Benutzerdefinierte Funktionen, also Algorithmen und Funktionen, die nicht Bestandteil des Ausführungssystems sind, werden schon lange in relationalen Datenbanksystemen unterstützt. Allerdings sind die Schnittstellen für diese Funktionen in relationalen Datenbanken häufig zu restriktiv, um komplexe Algorithmen zu implementieren. Google's MapReduce (84), SCOPE (236), Stratosphere (146) und Spark (254) sind einige Systeme, die ausdruckstärkere, benutzerdefinierte Funktionen parallel verarbeiten können. Dabei ist der Grad der parallelen Verarbeitung weitestgehend bestimmt durch die Semantik der Programmierschnittstellen, also z.B. die Funktionen zweiter Ordnung wie map und reduce in Hadoop, bzw. weiterer Funktionen wie Match, CoGroup und Cross in Stratosphere.

Niedrige Latenz auch für UDFs: Benutzer-definierte-Funktionen gehören nicht zum Funktionsumfang des ausführenden Systems, sondern werden häufig extern ausgeführt. Aus diesem Grund sind mit ihrem Einsatz oft höhere Ausführungskosten verbunden. Diese Kosten können gesenkt werden, je stärker UDFs mit dem Ausführungssystem integriert werden können. Eine weitere Herausforderung ist der Umgang mit benutzer-definierten-Funktionen bei der Optimierung von Datenanalyseprogrammen. Da die Semantik einer UDF dem Ausführungssystem üblicherweise unbekannt ist, können Analyseprogramme mit

UDFs nur eingeschränkt optimiert werden, wenn der Optimierer über zusätzliche Informationen über die UDFs verfügt. Erste Ansätze versuchen diese Informationen mit Hilfe von manuellen Annotationen (68) (220) oder mittels statischer Codeanalyse (221) (222) zu erhalten.

3.2.4 Herausforderungen und Forschungsfragen

Wir gliedern und priorisieren Forschungsfragen nach „Erforschung neuer Funktionalität“, „Qualitative Verbesserung existierender Funktionalitäten“ und „Optimierung der Ausführungskosten“ bei gleicher Qualität einer existierenden Funktionalität.

Erforschung neuer Funktionalität

- Was ist ein einheitliches und deklaratives Verarbeitungskonzept für Datenströme, Batch-basierte und iterative Verarbeitung?
- Wie können iterative Algorithmen, Zustand und geordnete Kollektionen deklarativ spezifiziert werden? Welche mathematisch-algebraischen Konstrukte sind hierfür erforderlich?
- Welche Äquivalenzregeln und damit automatische Optimierungsmöglichkeiten ergeben sich durch eine solche deklarative Spezifikation?
- Wie kann Fehlertoleranz deklarativ spezifiziert werden?
- Können Konsistenz- und Fehlertoleranzanforderungen automatisch aus der deklarativen Spezifikation eines iterativen Datenanalysealgoritmus abgeleitet werden?
- Wie können wir verteilte Datenstrukturen während der Anfrageverarbeitung verwalten?
- Wie können Datenstrukturen spezifiziert werden, um sie möglichst domänenübergreifend wiederzuverwenden?

Qualitative Verbesserung existierender Funktionalitäten

- Wie können wir trotz optimierender Transformationen numerische Stabilität der Datenanalysealgorithmen sicherstellen?

Optimierung der Ausführungskosten

- Welche algebraischen Transformationen können Datenanalyseprogramme optimieren?
- Wie können iterative Datenflüsse mit einem hohen Anteil an benutzerdefinierten Funktionen effizient ausgeführt werden?
- Welche Teile eines deklarativen Datenanalyseprogramms sollten als Datenfluss und welche als Kontrollfluss optimiert und parallelisiert werden?
- Sollte ein gegebenes Datenanalyseprogramm Fehlertoleranz pessimistisch durch Checkpointing, optimistisch durch Kompressionsoperationen oder hybrid realisieren?
- An welchen Stellen eines deklarativen Datenanalyseprogramms sollte im Hinblick auf Latenz und Datenrate Pipelining eingesetzt werden und an welchen Stellen materialisiert werden?
- Wie können wir Algorithmen zur Datenbeschaffung aus heterogenen und verteilten Quellen effektiv ausführen ohne dabei die Ressourcen der Datenquellen zu überlasten? Wie können Synergieeffekte zwischen Analyseprogrammen, die ähnliche Operationen auf gleichen Daten ausführen, ausgenutzt werden?
- Wie können Synergieeffekte zwischen Analyseprogrammen, die ähnliche Operationen auf gleichen Daten ausführen, ausgenutzt werden?

3.3 Datenanalysesysteme

In den letzten Jahren wurde eine große Anzahl von Datenanalysesystemen entwickelt und vorgestellt. Diese Systeme unterscheiden sich teilweise erheblich in den Technologien, die sie einsetzen und den Anwendungsgebieten für die sie entwickelt wurden. Auf der Technologieseite sind derzeit folgende Trends zu beobachten: die Entwicklung und der Einsatz von dynamisch konfigurierbarer, paralleler oder heterogener Hardware, Nutzung von großen Hauptspeicher mengen und die Parallelisierung von Datenanalyse über viele Rechner hinweg. Die Einsatzgebiete reichen von eher traditioneller relationaler Datenbankverarbeitung wie klassische OLAP Anfragen (Data Warehousing, Reporting) bis zu speziellen Anwendungen wie der Verarbeitung von graph-strukturierten Daten, Rechenoperationen auf großen Matrizen oder iterativer Datenverarbeitung, wie sie für viele Verfahren des maschinellen Lernens oder Data-Mining benötigt werden. Einige Systeme haben den Fokus auf die Unterstützung von möglichst vielen Anwendungsfällen gelegt. Das vielfältige Angebot von Datenverarbeitungssystemen wird von einer großen Zahl unterschiedlicher Anbieter bereitgestellt und entwickelt. Viele Systeme werden von der Open Source-Gemeinschaft entwickelt und sind frei verfügbar. In diesem Bereich ist das Apache Hadoop Projekt mit seinen zahlreichen Unterprojekten hervorzuheben, das eine große Menge von unterschiedlichen Komponenten und Systemen bereitstellt. Weiterhin sind sowohl viele junge Unternehmen und Start-Ups in diesem Segment aktiv, als auch große Konzerne wie zum Beispiel IBM, Microsoft, SAP und Amazon. Schließlich werden Systeme auch von Forschergruppen entwickelt wie zum Beispiel Asterix von der UC Irvine (259), Spark von der UC Berkeley (254), Hyper an der TU München (261) und Stratosphere an Universitäten in Berlin und Potsdam (146). Teilweise werden diese Systeme auch unter Open Source Lizzenen frei bereitgestellt. Aufgrund der Vielfalt von verwendeten Technologien, unterstützten Anwendungsgebieten und Anbietern ist ein direkter und umfassender Vergleich von Systemen zur Datenanalyse schwierig. Darüber hinaus werden derzeit kontinuierlich neue Systeme und Projekte vorgestellt. Seine große Dynamik und Heterogenität macht diesen Markt derzeit sehr unübersichtlich. Abbildung 57: Schichtenmodell ausgewählter „Big Data Systeme“ zeigt eine selektive Übersicht einzelner Systeme die für spezielle Anwendungsanforderungen von einzelnen Organisationen entwickelt wurden. Zum Beispiel ergänzt das System BigTable (233) von Google das verteilte Dateisystem GFS (227) mit Strukturen für den schnellen punktuellen Zugriff auf Daten bzw. für Anfragen die Wertebereiche erfordern. Durch Zugriff auf den Datenspeicher BigTable ermöglicht das Anfrageausführungssystem Dremel (246) die schnelle Beantwortung von Aggregationsanfragen (z.B. im Reporting). Das System Pregel (111) ergänzt BigTable für den Spezialfall der Verarbeitung sehr großer Graphen (Webgraphen, Klickgraphen bzw. Graphen in sozialen Netzwerken). Die Firma Google stellt diese Technologien nur intern zur Verfügung. Jedoch arbeiten derzeit verschiedene Forschungseinrichtungen an

Erweiterungen oder eigenen Implementierungen dieser Funktionalitäten (siehe Hadoop (154), Asterix (259), GraphLab (103), Spark (254) oder Stratosphere (146)). Auf der linken Seite zeigt die Abbildung auch die Schichtung eines „klassischen“ relationalen Datenbankmanagementsystems. Derzeit werden viele Systeme entwickelt, die diese Architektur auf die Nutzung von großen Hauptspeichermengen anpassen. So entwickeln Microsoft, IBM, Vertica, Vectorwise und SAP HANA Technologien für spaltenorientierte Speicherung und Kompression im Hauptspeicher. An der TU München wird der klassische Aufbau im Rahmen des Hyper Projekts (261) zur gleichzeitigen Verarbeitung von OLTP und OLAP Anfragen erweitert. Das größte Manko der existierenden Systeme ist ein Fehlen einer deklarativen Spezifikation von komplexen Datenanalyseprogrammen, welche komplexe benutzerdefinierte Funktionen und iterative Algorithmen beinhalten, sowie die damit verbundene automatisch ableitbare Parallelisierung und die dynamische Adaption auf eine Zielrechnerarchitektur und Systemumgebung. Erst dadurch wird ein erheblich größerer Personenkreis Analysen auf Big Data durchführen können, da die deklarative Spezifikation keine Kenntnisse in Systemprogrammierung von verteilten Systemen oder parallelen Algorithmen mehr erfordert. Dennoch wird durch die Automatisierung eine effiziente, von der Problemgröße, Workload und Rechnerarchitektur abhängige optimierte Verarbeitung ermöglicht, wie sie im Umgang mit heterogenen, großen Datenmengen mit hoher Datenrate erforderlich ist. Der Entwicklung eines Datenanalysesystems mit deklarativer Spezifikation und automatischer Optimierung käme dieselbe wissenschaftliche und wirtschaftliche Bedeutung zu wie die Entwicklung der relationalen Algebra im Jahre 1970. Diese einheitliche, deklarative und mathematisch fundierte Basis ermöglichte eine deklarative, optimierbare Spezifikation von Anfragen und begründete aus einem Wildwuchs von vielen kaum wartbaren, teuren, schlecht integrierten, nur von Experten nutzbaren Datenverarbeitungssystemen (hierarchisch, CODASYL, Netzwerkdatenbanken) den Milliardenmarkt der relationalen Datenbanksysteme. Derzeit befindet sich der Big Data Markt in einem ähnlichen Zustand wie der Markt der Datenbanksysteme vor 1970.

In diesem Kapitel beschreiben wir aktuelle Trends in der Entwicklung und Forschung von Systemen zur Verarbeitung und Analyse von Big Data. In den folgenden Abschnitten stellen wir den Einsatz von paralleler und heterogener Hardware zur Datenverarbeitung, Hauptspeicherdatenbanken, parallele Datenbanksysteme, parallele Systeme zur stapelbasierten Verarbeitung und parallele Systeme für Maschinenlernverfahren vor.

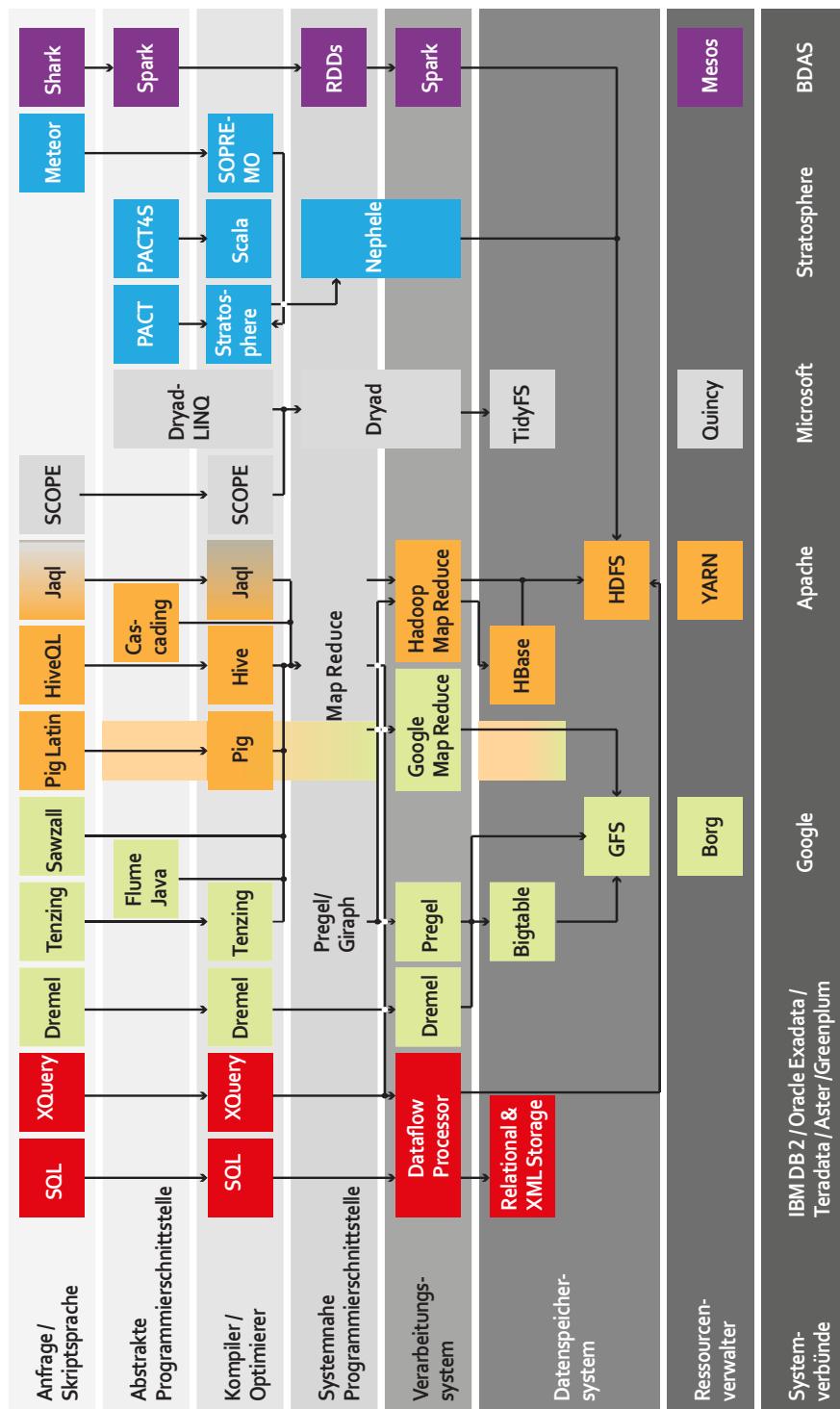


Abbildung 57: Schichtenmodell ausgewählter „Big Data Systeme“

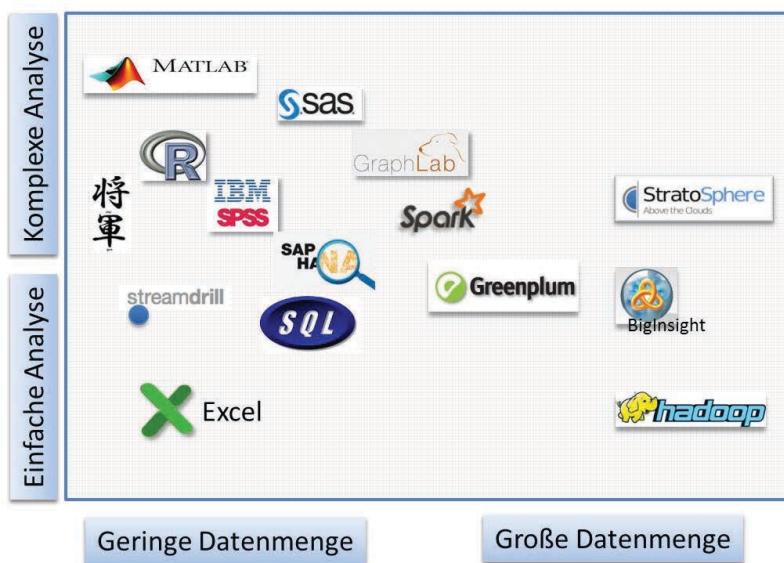


Abbildung 58: Landschaft ausgewählter Anbieter

3.3.1 FPGAs, GPUs, MultiCore-CPUs und PCM für die Datenanalyse

„Big Data“ erfordert die parallele Verarbeitung von Daten, um auch in akzeptabler Zeit Datenanalyseanfragen auf großen Datenmengen und hohe Datenraten ausführen zu können. Das Gesetz von Amdahl (163) beschreibt die Grenzen der Skalierbarkeit durch Parallelverarbeitung in Abhängigkeit des parallelisierbaren Anteils eines Programms. So begrenzt bereits ein 5%-iger Anteil von sequenziellen Arbeitsschritten in einem Analyseprogramm die mögliche Laufzeitreduzierung auf den Faktor 20 bei einer 4000fachen Parallelisierung. Deshalb ist es wichtig, möglichst vollständige Parallelisierung von Analyseprogrammen zu erzielen. Aktuell sind einige Hardwaretrends zu beobachten, die in diesem Kontext relevant sind. Grafikprozessoren und Multi- und ManyCore CPUs erhöhen die Anzahl von Rechenkernen und somit den möglichen Parallelisierungsgrad innerhalb eines Rechners. FPGAs sind Rechenprozessoren, die dynamisch umkonfiguriert werden können, um bestimmte Aufgaben sehr schnell direkt in Hardware auszuführen. Heterogene Hardware ermöglicht es verschiedene Aufgaben jeweils auf der am besten geeigneten Hardware zu verarbeiten. In diesem Abschnitt stellen wir aktuelle Hardwareentwicklungen vor und beschreiben ihren möglichen Einfluss auf die Zukunft des Big Data Managements.

Vielfältige Hardware: Bis in die späten 90er Jahre konnte die sequenzielle Verarbeitungsgeschwindigkeit von Operationen auf Mikroprozessoren (engl. Central Processing Unit, CPU) stetig gesteigert werden. Obwohl die Verarbeitungsgeschwindigkeit eines einzelnen Rechenkerns seit dieser Zeit nahezu konstant geblieben ist, gilt immer noch das vom Intel Mitgründer Gordon

Moore postulierte Gesetz des exponentiellen Wachstums der Transistoren auf einem Chip. Um weitere Leistungssteigerungen zu erzielen, bestehen Mikroprozessoren heutzutage aus vielen parallelen Rechnerkernen, die unabhängig (MIMD, multiple instruction multiple data, multicore) oder abhängig (SIMD, single instruction multiple data, häufig auf Grafikprozessoren, sogenannten GPUs) mehrere Daten parallel verarbeiten und darauf Berechnungen durchführen.

Aufgrund von Problemen mit der Abwärme von Prozessoren können dabei bei höchster Integrationsdichte nicht mehr alle Prozessoren gleichzeitig betrieben werden. Daher zeichnet sich die Entwicklung von sogenannten heterogenen CPUs ab, welche mehrere unterschiedliche Funktionen in Hardware durchführen können (SIMD und MIMD kombiniert, zusätzliche Recheneinheiten für Verschlüsselung, Datenstromverarbeitung, etc., teilweise auch als Hardware programmierbar, als sogenannte free programmable gate arrays (FPGA)). All diese Architekturen haben zum einen ihre massive inhärente Parallelität gemein, zum anderen ein zur klassischen von-Neumann CPU unterschiedliches Programmiermodell. Solche Hardwarearchitekturen bieten die Möglichkeit die Arbeit zwischen der CPU und den anderen Prozessoren aufzuteilen und damit größere Arbeitspakete zu verarbeiten. Allerdings müssen heutige Datenanalysesysteme auf derartige neue Technologien erst angepasst werden, beziehungsweise neue Systeme entwickelt werden, die die Eigenschaften von moderner Hardware berücksichtigen und ausnutzen. In diesem Bereich gibt es viele Möglichkeiten für Unternehmen. Auch deutsche Unternehmen (z.B. ParStream) sind in diesem Bereich aktiv und können durch innovative Nutzung der neuen Hardwaredmöglichkeiten Marktanteile im Bereich der Datenanalysesysteme erobern. Im Folgenden beschreibt die Studie einige der wesentlichen technologischen Entwicklungen im Detail.

Field Programmable Gate Arrays (FPGAs): FPGAs sind Hardwarechips mit Ein- und Ausgängen, die mit geringem Aufwand umkonfiguriert werden können. Reguläre CPUs verarbeiten Daten indem sie zunächst Verarbeitungsvorschriften einlesen und diese danach auf die Eingabedaten anwenden. Im Gegensatz dazu werden die Verarbeitungsvorschriften bei FPGAs fest verdrahtet, so dass einfache Algorithmen direkt in die Hardware implementiert werden und sehr effizient ausgeführt werden. Die Eigenschaft von konfigurierbarer Hardware wird durch den besonderen Aufbau von FPGAs erreicht. FPGAs sind Felder von konfigurierbaren Logikblöcken (CLB), die durch Routingkanäle verbunden sind und die durch programmierbare Input-/Outputblöcke ergänzt werden, d.h. Schnittstellen zwischen dem FPGA und externen Ressourcen. Mehrere Logikzellen und eine Switchbox ergeben ein CLB. Switchboxen verbinden die Logikzellen mit dem Verbindungsfeld. Eine Logikzelle enthält Logikgatter, Carry-Elemente und ein Speicherelement. Ein Logikgatter ist implementiert als eine Lookuptabelle (LUT).

FPGAs ermöglichen kurze Produkteinführungszeiten: Die Entwicklung von FPGAs hat die Art und Weise, wie digitale Schaltungen entwickelt werden, verändert. Die Fähigkeit Hardwarechips im Betrieb so oft wie nötig umzaprogrammieren, ermöglicht es Hardwareentwicklern schnell Prototypen zu entwickeln und resultiert somit in kurzen Produkteinführungszeiten. Zusätzlich bedeuten Fehlerkorrekturen nicht mehr, dass Platinen neu entworfen werden müssen, da die Schaltungen einfach umkonfiguriert werden können.

FPGAs können bereits im Labor für die Datenanalyse verwendet werden:

Die Autoren von (164)(165)(166)(168) haben gezeigt, dass FPGAs für die Datenverarbeitung verwendet werden können. Als Anwendungsfall haben sie zum Beispiel den Medianoperator für Sortiernetzwerke realisiert. Wenn mehrere Datenströme parallel verarbeitet werden sollen, haben FPGAs eine bessere Leistung im Vergleich zu CPUs. Zusätzlich verbrauchen FPGAs weniger Strom als CPUs, was von immer stärkerer Bedeutung für die Energieoptimierung der Datenanalyse wird.

Graphics Processing Units (GPUs) waren ursprünglich reine Grafikbeschleuniger, z.B. für anspruchsvolle Computerspiele. Im Vergleich zu CPUs haben GPUs wesentlich mehr Rechenkerne, die parallel betrieben werden können. Jedoch ist der Funktionsumfang dieser Kerne deutlich geringer als die von CPU Kernen. Deshalb eignen sich GPUs nur für bestimmte Operationen, die dann jedoch sehr schnell ausgeführt werden können. Seit dem Jahr 2000 wurden GPUs zunehmend einfacher zu programmieren und Forscher begannen damit Grafikkarten für Anwendungen abseits von Bildverarbeitung zu verwenden. Hieraus entwickelte sich ein Trend, der unter dem Namen „General-Purpose Computing on Graphics Processing Units“ (GPGPU) immer populärer wurde. Mittlerweile wird diese Technologie auch für die Verarbeitung von Daten eingesetzt.

Das Programmieren für eine GPU war für viele Programmierer noch zu kompliziert und erforderte detaillierte Kenntnisse der Architektur. Im Jahre 2004 haben Ian Buck et.al. von der Stanford Universität das System Brook veröffentlicht, ein System für die Programmierung allgemeiner Anwendungen auf Grafikhardware. „Brook“ (169) wurde als eine Erweiterung von C entwickelt und machte die GPU Verarbeitung wieder einem größeren Publikum zugänglich. Buck arbeitete seit 2006 für NVIDIA und veröffentlichte dort die erste Version von CUDA (Compute Unified Device Architecture) einer parallelen Berechnungsplattform und einem Programmiermodell für allgemeine Algorithmen (169)(170). Ende 2008 veröffentlichte die Khronos-Gruppe ein Konkurrenzprodukt zu CUDA namens OpenCL 1.0, ein offenes Cross-Platform-Programmiermodell. Im selben Jahr stellte Microsoft DirectCompute auf der Gamefest 2008 vor. DirectCompute ist Teil von DirectX 11 und ist eine API für GPGPU für Microsoft Windows Vista und Windows 7.

GPUs sind noch nicht ausreichend vorbereitet auf die Datenanalyse haben aber großes Potenzial: GPUs bieten viel Rechenleistung für die Verarbeitung von „Big Data“. Engpässe dieser Hardware sind aber derzeit direkt adressierbare Speicher und Bandweite. So haben GPUs typischerweise nur auf vergleichsweise kleine Speichermengen direkten Zugriff und ihre Anbindung zum deutlich größeren Hauptspeichers ist wesentlich schlechter als dessen Verbindung zur CPU. Deshalb dauert es verhältnismäßig lange um Daten von außerhalb der GPU verfügbar zu machen. Daher eignen sich GPUs vorwiegend für aufwendige, oder wiederholte Berechnungen auf kleinen Datenmengen, die in den internen Speicher passen, um den langsamen Datentransfer über den PCIe-Bus zu amortisieren.

Die Rolle der „klassischen“ CPU: In der Zukunft werden CPUs mit immer mehr Rechnerkernen (engl. CPU cores) ausgestattet und damit verstärkt umgewandelt in parallele Systeme. Entweder geschieht dies durch Verbindung unterschiedlicher Rechnerkerne – wie z.B. bei AMDs APUs – wobei zwei Schnittstellen bereitgestellt werden, eine Schnittstelle für sequentielle und kontrollstrukturlastige Codeteile und eine andere für parallele Datenverarbeitungsoperatoren. Oder es wird Intels SCC-Idee verfolgt (223), die viele Rechnerkerne auf einem einzigen Chip integriert. Bei dieser Architektur sind Rechnerkerne in einem Netzwerk auf dem Chip organisiert, so dass die Zugriffsgeschwindigkeit von Rechnerkernen auf den Speicher von anderen Kernen nicht einheitlich ist, sondern von der Position der Kerne auf dem Chip abhängt. Der Aufbau dieses Netzwerks muss im Rahmen der Optimierung und Verarbeitung von Datenanalyseprogrammen durch die Datenanalysesysteme berücksichtigt werden.

Software-defined Networking: Software-defined Networking (SDN) bietet die Möglichkeit, durch die dynamische Umkonfiguration von Routern die Datenübertragung in einem Netzwerk zu optimieren. Dazu können Teile der Routing-Verfahren programmiert werden, beispielsweise durch die Programmierschnittstelle OpenFlow (224). Derartige Router können helfen, die Kommunikation in „Big Data“ Analysesystemen zu verbessern, z.B. durch eine effizientere Implementierung der Shuffle-Phase in Map/Reduce-Systemen. Auf diese Weise könnten Router die Zeit der Datenübertragung und damit die Latenz bei der Verarbeitung von Algorithmen zur Analyse von Big Data reduzieren.

Phase Change Memory (PCM): Phase Change Memory ist eine neue Technologie Speicherzellen. Sie bietet wie Hauptspeicher sehr schnelle Lese- und Schreiboperationen auf beliebige Speicherbereiche, speichert Daten jedoch nicht flüchtig. Das heißt, auch im Falle einer Stromunterbrechung bleiben die gespeicherten Daten erhalten. Derzeit ist PCM noch in der Entwicklungsphase, erste Speichermodule werden jedoch schon hergestellt und verkauft. Wann sich diese Technologie für den Massenmarkt durchsetzt ist noch nicht abzusehen. Mit PCM ergeben sich aber viele neue interessante Fragestellungen im Kontext

von Datenverarbeitungssystemen. Viele Komponenten aktueller Systeme (z.B. Buffer Pools, Caches, Logging und Recovery Methoden) adressieren die unterschiedlichen Zugriffsgeschwindigkeiten von persistentem (Festplatte, SSD) und flüchtigem Speicher (Hauptspeicher) und müssten im Falle von PCM sicher neu überdacht werden.

3.3.2 Hauptspeicherdatenbanken

In den letzten Jahren ist Hauptspeicher sehr viel preiswerter geworden. Einzelrechnerlösungen gelangen im Rahmen von Business Intelligence (BI) Anwendungen schnell an ihre Grenzen. Aufgrund der gestiegenen Speicherkapazität von Festplatten lassen sich riesige Datenmengen zwar auf einem Rechner bündeln. Die Auswertung der Daten nimmt dann jedoch immer mehr Zeit in Anspruch, so dass das System die gewünschte Analyse nicht rechtzeitig ermöglicht und damit letztendlich unbenutzbar wird. Der Grund dafür ist, dass bei gestiegener Speicherkapazität die Datenübertragungsrate von Festplatten nahezu konstant geblieben ist. Durch die Verfügbarkeit von immer größeren Hauptspeichermodulen, gepaart mit fallenden Preisen pro GB Speicherkapazität (siehe Abbildung 59) ergeben sich allerdings vollkommen neue Möglichkeiten für die schnelle Auswertung dieser riesigen Datenmengen. Es ist nun vertretbar große Mengen an Daten im Arbeitsspeicher vorzuhalten. So kann auf mehr Daten viel schneller zugegriffen werden. Hauptspeicherzugriffe können 10 – 100.000 schneller sein als entsprechende Zugriffe auf Festplatten oder Solid State Disks (SSDs) (Abbildung 60). Damit können auch Anwendungen, die auf diesen Daten arbeiten, deutlich beschleunigt werden.

Entwicklung der Hauptspeicherpreise pro Gigabyte

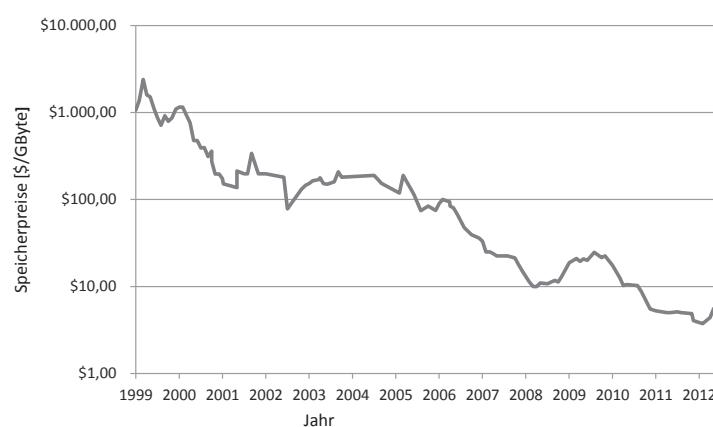


Abbildung 59: Fallende Entwicklung der Preise pro Gigabyte in USD für Arbeitsspeicher

Hinweis: Die Festplattentests wurden auf einer frisch gebooteten Maschine (WIN 2003 Server mit 64 GB RAM und acht 15.000-PPM SAS Platten mit RAID Konfiguration) ausgeführt, um den Effekt des Betriebssystem-Festplatten-Caches zu eliminieren. Der SSD Test nutzte Intel Hochleistungs SATA der neuesten Generation

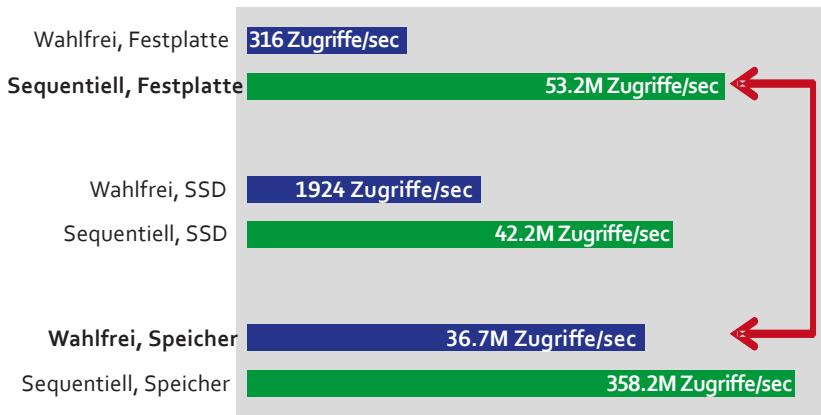


Abbildung 60: Zugriffsgeschwindigkeiten auf Festplatte, SSD und Hauptspeicher, (nach: <http://queue.acm.org/detail.cfm?id=1563874>)

Hauptspeicherdatenbanksysteme (Main Memory Database Systems): In der Vergangenheit nutzten Datenbanksysteme Festplatten aufgrund ihrer großen Kapazität um Daten zu speichern. Hauptspeicher wurde als Pufferspeicher (Cache) verwendet, um häufig angefragte Daten zwischenzuspeichern und auf Anfrage schnell verfügbar zu haben. Hauptspeicherdatenbanksysteme drehen diese Rolle um. Sie speichern alle oder einen Großteil der Daten komplett im Hauptspeicher und die Festplatte wird nur als Sekundärspeicher benutzt um Daten gegen Stromausfälle zu sichern. Hauptspeicherdatenbanken haben den Vorteil, viel schnellere Antwortzeiten zu haben, als traditionelle festplattenbasierte Systeme. Dies begründet sich aus den viel schnelleren Zugriffszeiten auf die Daten im Hauptspeicher. Der entscheidende Vorteil ist die Beschleunigung der Lese- und Schreibzugriffe. Die zunehmende Bedeutung von Hauptspeicherdatenbanken im Rahmen der BI belegen bereits einige kommerzielle Systeme, wie SAP HANA, EXASolution und ParStream aus Deutschland und andere Systeme von internationalen Anbietern.

Kompressionsverfahren und Multi-Core-Architekturen: Da Hauptspeicherplatz begrenzt ist, benutzen die meisten Systeme Kompressionsverfahren, um den Speicher effizient zu nutzen. Moderne Hardware kann hierbei die Kompression und Dekompression direkt unterstützen, so dass dadurch kaum Geschwindigkeitseinbußen entstehen. Bestimmte Berechnungen können auch, abhängig vom Kompressionsverfahren, direkt auf komprimierten Daten ausgeführt werden. Weiterhin unterstützen viele dieser Systeme neben dem traditionellen zeilenbasierten Speicherverfahren auch ein spaltenbasiertes Speicherverfahren. Hier werden Daten, die zu einer Zeile gehören, nicht mehr „nebeneinander“

innerhalb eines Datums gespeichert, sondern die Daten einer Zeile werden aufgeteilt und gemeinsam mit allen anderen Daten einer Spalte gespeichert. Dies hat mehrere Vorteile. Es reduziert den Aufwand für den Datenzugriff, wenn Anfragen nur einen Teil einer Datenzeile lesen. Spaltenorientierte Daten lassen sich meist besser komprimieren, da sie das gleiche Format haben und eine höhere Datenredundanz aufweisen. Schließlich können spaltenorientierte Daten direkt vektorisiert von Rechenkernen mittels SIMD-Anweisungen (Single Instruction, Multiple Data) verarbeitet werden. Wenn Anfragen eine große Selektivität haben, also nur einen geringen Teil der verfügbaren Daten lesen sollen, sind zusätzlich Indexe vorteilhaft. Damit kann die Zugriffszeit signifikant reduziert werden, da die relevanten Daten schnell gefunden werden können. Hier unterscheiden sich die verfügbaren Systeme. SAP HANA erlaubt das manuelle Anlegen von Indexen. EXASolution erstellt die Indexe automatisch basierend auf dem Anfrageprofil. ParStream benutzt hoch komprimierte Indexe, welche als Bitvektoren gespeichert werden, um Anfragen mit Selektionen in vielen Attributen (z.B. beim Zugriff auf Teile einer riesigen Faktentabelle in einem Datawarehouse) effizient mit sehr geringer Latenz beantworten zu können.

Insgesamt ermöglichen Main Memory Datenbanken extrem schnelle Antwortzeiten auf großen Datenmengen, vorausgesetzt der Datenbestand passt in den Hauptspeicher. Alle diese Systeme basieren auf dem relationalen Speicher- und Verarbeitungsmodell und unterstützen die SQL-Anfragesprache. Damit lassen sie sich in ein konventionelles Reporting „Ökosystem“ integrieren (171).

Optimierungsmöglichkeiten: Durch die Wahl der Speicherorganisation ergeben sich unterschiedliche Optimierungsmöglichkeiten. Systeme, die nur lesen zur Berichtsgenerierung oder zur Analyse auf Daten zugreifen, profitieren eher von einer spaltenorientierten Organisation. Spaltenorganisierte Daten lassen sich meist besser komprimieren. Für Anfragen, die sich nur auf einen Teil der Daten beziehen, kann auch die Anzahl der Datenzugriffe reduziert werden, was die Latenz der Datenanalyse verbessert. Transaktionselle Systeme in denen Daten überwiegend eingefügt, geändert oder gelöscht werden, profitieren oftmals von einer zeilenorientierten Organisation. Zeilenbasierte Speicherorganisation konzentriert Schreiboperationen auf weniger physische Speicherorte und vereinfacht den Schutz der Daten vor konkurrierenden Zugriffen damit die Konsistenz der Daten sichergestellt ist. Dieser Unterschied wird in der Systemarchitektur von Data Warehouse Systemen deutlich. Transaktionen und Änderungen der Datenbasis werden von den vielen Systemen entweder nicht unterstützt – es handelt sich dann um sogenannte Read-Only- bzw Append-Only-Systeme wie zum Beispiel ParStream. Andere Systeme wie SAP HANA verwalten Transaktionen in einem besonderen differenziellen Speicher und überführen diese regelmäßig in einen spaltenorientierten Datenspeicher.

Key-Value Hauptspeicher-Caches: Nicht immer bieten Systeme eine SQL-Schnittstelle um auf Daten zuzugreifen. Eine häufige Alternative ist eine sogenannte Key-Value-Schnittstelle, bei der nur einzelne Datensätze über einen

festen Schlüssel abgerufen werden können. Der Vorteil ist, dass diese Abrufoperation sehr effizient und schnell ausgeführt werden kann. Jedoch ist das häufig die einzige effiziente Operation, die Systeme mit Key-Value Schnittstelle ausführen können. Komplexe Anfragen werden entweder gar nicht oder nur sehr ineffizient unterstützt. Auch diese Systeme wollen von verteiltem Hauptspeicher profitieren. In diesem Zusammenhang gibt es Lösungen, die es ermöglichen, den Hauptspeicher einer großen Anzahl von Rechnern quasi als Pool anzusprechen. Traditionell wird diese Lösung für verteilte Caches benutzt, um Zwischenergebnisse vorzuhalten, oder häufig benutzte Daten schneller verfügbar zu machen. In diesem Bereich gibt es neben Open Source Lösungen (memcached) ebenfalls kommerzielle Anbieter, wie das derzeit von der Software AG entwickelte System Terracotta BigMemory, welches neben einer verbesserten Konsistenzgarantie zusätzlich eine Volltextsuche über die im verteilten Speicher verfügbaren Daten anbietet. Systeme dieser Art bieten einen schnellen, punktuellen Zugriff auf einzelne Dateneinträge. Deshalb werden sie eher für transaktionselle Anwendungen oder zum Zwischenspeichern von Daten verwendet. Zur Verarbeitung oder Analyse von großen Datenmengen sind sie weniger gut geeignet.

Gegenstand aktueller Forschung sind sogenannte hybride Speicherorganisationen, welche die Daten abhängig von den Anforderungen spaltenbasiert, zeilenbasiert oder in einer Mischform speichern um die schnellsten Ausführungszeiten zu garantieren (171).

Weitere Verbesserungsmöglichkeiten ergeben sich aus der Art und Weise, wie Daten abgespeichert werden. Wenn Daten durch statistische Modelle beschrieben werden können, kann es sinnvoll sein, diese Modelle direkt in die Datenbanken zu integrieren. Die statistischen Modelle können mit neuen Daten schnell erweitert und verfeinert werden, wenn sie im Hauptspeicher gehalten werden. Dadurch werden Trendvorhersagen aufgrund der tatsächlichen Daten möglich und ein gutes Modell kann sogar die eigentlichen Daten ersetzen, da diese daraus hergeleitet werden können. Somit können Modelle theoretisch noch größere Datenmengen abbilden.

Ein Nachteil von Hauptspeicherdatenbanken ist ihr flüchtiger Speicher. Bei einem Stromausfall gehen alle Daten im Hauptspeicher verloren. Hier bieten sich unterschiedliche Strategien an, um Datenverlust auszuschließen. Hybride Systeme schreiben die Daten ebenfalls in nicht-flüchtigen Speicher. Alternativ können die Daten auch über mehrere Rechner verteilt, repliziert und in deren Hauptspeicher gespeichert werden. Interessant ist hier die Verbindung von Hauptspeicher und ebenfalls schnellem Flash-Speicher in einem so genannten **Appliance Model**, das spezielle Hardware und darauf angepasste Software bündelt. Durch die Verbindung von dedizierten Servern mit großem Hauptspeicher und schnellem Flash-Speicher in einer Appliance ergeben sich

Möglichkeiten, die Datenspeicherung und Verarbeitung zu beschleunigen. Verbesserte Daten- und Dateistrukturen, die den Flash-Speicher optimal ausnutzen und angepasste Formate für das Ablegen der Daten im Hauptspeicher, können den Systemdurchsatz bei der Verarbeitung von Datenanalyseprogrammen deutlich verbessern. Gleichzeitig kann solch ein System komplett optimiert werden, da es für einen bestimmten, im Vorfeld bekannten Anwendungsfall konstruiert wird. Die CPU erledigt die Datenkompression, die Daten werden entsprechend der Speicherbusbreite und der L1/L2 Caches auf der CPU im Speicher verwaltet und angepasste Datenstrukturen enthalten nicht nur Rohdaten, sondern Aggregationen, Indexe und/oder Statistiken über die Daten, um Antworten weiterhin zu beschleunigen. Wie in Abschnitt 3.3.1 beschrieben, kann die Software dahingehend optimiert werden, die Fähigkeiten neuester Rechnerarchitekturen zu benutzen (SIMD, MIMD, FPGA). Derartige spezialisierte, speziell auf Kundenanforderungen abgestimmte Systeme existieren derzeit noch nicht, könnten aber ein wichtiger Bestandteil des Markts für Datenanalysesysteme werden.

3.3.3 MPP-fähige Datenbanksysteme

Neben dem Ausnutzen von neuer Hardware und viel Hauptspeicher ist massiv parallele Datenverarbeitung eine andere Technologie um große Datenmengen schnell zu analysieren. Dafür wird ein Analyseprogramm parallelisiert, auf mehrere Rechner verteilt und dort gleichzeitig ausgeführt. Je nach Größe der Daten können hierfür bis zu einige tausend Rechner zum Einsatz kommen. Ein wichtiger Bereich der Datenanalyse sind nach wie vor klassische relationale Datenbank- und Reportinganwendungen, die eine SQL Schnittstelle voraussetzen.

Parallele Datenbanksysteme: Schon sehr lange bevor der Big Data Trend in Schwung kam, sind die ersten parallelen Datenbanksysteme entwickelt worden. Damit bilden parallele relationale Datenbanksysteme die erste Generation von Systemen, die sehr große Datenmengen verarbeiten können. Erste Forschungsprototypen und kommerzielle Systeme entstanden um 1984 (225) (226). Im Laufe der Zeit haben viele Hersteller wie zum Beispiel Oracle, IBM, Microsoft, Greenplum und Asterdata ebenfalls parallele Datenbanksysteme entwickelt oder bestehende Systeme entsprechend erweitert. In den letzten Jahren wurden außerdem zahlreiche neue analytische parallele Datenbanksysteme entwickelt, die auch neue Technologien und Hardwareeigenschaften ausnutzen. Darunter sind Vertica (262), EXASolution (252), HANA (171) und ParStream (263). Vertica setzt spaltenorientierte Speicherung und Kompression ein, um die Menge der zu verarbeitenden Daten stark zu reduzieren. EXASolution, HANA und ParStream sind Hauptspeicherdatenbanken und wurden bereits in Abschnitt 3.3.2 kurz vorgestellt. Neben dem Speichern von Daten im Hauptspeicher können diese Systeme auch parallel auf mehreren Rechnern betrieben

werden. Da parallele relationale Datenbanksysteme in der Regel relativ aufwändig zu konfigurieren und zu betreiben sind, wird qualifiziertes Personal für diese Aufgabe benötigt. Alle in diesem Abschnitt vorgestellten Datenbanksysteme sind kommerziell und haben hohe Betriebskosten. Diese setzen sich hauptsächlich aus den Lizenzkosten (häufig abhängig von der Anzahl der Rechner, Menge des Hauptspeichers oder der Größe der gespeicherten Daten) und Personalkosten zusammen.

3.3.4 Batch und Hadoop-basierte Systeme

Parallele relationale Datenbanksysteme sind in der Lage sehr große Datenmengen schnell und effizient zu analysieren. Allerdings erfordert ihr Einsatz, dass die Daten in einem relationalen Datenschema abgebildet und die Analyseprogramme in SQL ausgedrückt werden können. Für viele Anwendungsfälle sind diese Anforderungen zu restriktiv. Da relationale Datenbanksysteme außerdem aufwändig zu konfigurieren sind, verursacht ihr Betrieb häufig neben Lizenzkosten auch hohe Personalkosten. Aufgrund des limitierten Einsatzbereiches und der hohen Betriebskosten, wurden neue Systeme entwickelt, um große Datenmengen flexibel parallel auf vielen Rechnern verarbeiten zu können. Vorreiter bei der Entwicklung von Big Data Systemen der zweiten Generation war das Unternehmen Google.

Google GFS und Map/Reduce: In den Jahren 2003 und 2004 publizierte Google die Architektur seines verteilten Dateisystems GFS (227) und das Map/Reduce Programmiermodell und Ausführungssystem (85). Google benutzte diese Systeme unter anderem um aus allen Internetseiten einen Suchindex zu erstellen. GFS ist ein Speichersystem, das riesige Datenmengen beliebigen Formats verteilt auf tausenden Rechnern speichern kann. Das System ist äußerst fehler-tolerant, so dass einzelne Speicherknoten ausfallen können ohne den Betrieb zu gefährden. Damit entfällt die Anforderung zuverlässige Hardware einzusetzen und günstige herkömmliche Hardware kann für den Betrieb genutzt werden. Allerdings haben verteilte Dateisysteme auch Einschränkungen. So können einmal geschriebene Daten nicht mehr verändert werden. Damit eignen sie sich für Anwendungsfälle in denen Daten einmal geschrieben oder importiert werden und danach häufig gelesen werden. Map/reduce ist ein Programmiermodell, das die Spezifikation von parallelen Datenanalyseprogrammen ermöglicht. Normalerweise ist die Entwicklung von parallelen Anwendungen sehr aufwändig und erfordert umfangreiches Spezialwissen. Die wichtigste Eigenschaft von Map/Reduce ist, dass es die komplexen Aspekte von paralleler Programmierung wie automatische Fehlertoleranz, effektiven Lastausgleich und verlässliche Netzwerkkommunikation transparent für den Benutzer behandelt. Map/reduce erhält diese Eigenschaften durch sein Programmiermodell, das die

Spezifikation von Programmen in einem bestimmten Schema erfordert und dem Zusammenspiel von Programmiermodell und Ausführungssystem. Trotz des relativ strikten Programmiermodells können sehr vielfältige Datenverarbeitungsprogramme implementiert werden. So können Map/Reduce Programme beliebig strukturierte Daten aus allen Speichersystemen wie zum Beispiel verteilten Dateisystemen, relationalen Datenbanken oder auch Key-Value Stores lesen und verarbeiten. Analyseprogramme können flexibel skaliert und parallel auf mehreren tausend Rechnern ausgeführt werden. Map/reduce findet üblicherweise Anwendung, wenn sehr große Datenmengen analysiert werden müssen. Da Map/Reduce nicht im Hinblick auf interaktive Datenanalyse entwickelt wurde, sind dabei Antwortzeiten im Bereich von Minuten oder Stunden zu erwarten.

Apache Hadoop und sein Ökosystem: Google nutzt GFS und Map/Reduce nur intern und hat es nicht öffentlich zur Verfügung gestellt. Basierend auf den wissenschaftlichen Publikationen über diese Systeme wurde jedoch das Apache Open Source Projekt „Hadoop“ (154) gestartet. Hadoop besteht aus einem verteilten Dateisystem (HDFS) das sich an GFS anlehnt und einer Map/Reduce Programmierschnittstelle und dem entsprechenden Ausführungssystem. Die freie Verfügbarkeit eines Systems, das mit geringem technischen Aufwand und niedrigen Kosten die Auswertung von großen Datenmengen ermöglicht, hat dem „Big Data“ Trend einen starken Impuls gegeben. Basierend auf oder ergänzend zu Hadoop haben wurden in den letzten Jahren zahlreiche weitere Apache Open Source Systeme und Komponenten entwickelt. Zu diesem Hadoop Ökosystem gehören unter anderem die Anfrage- und Skriptsprachen Pig (228) und Hive (229), die Bibliothek für Data-Mining und maschinelles Lernen Mahout (230), das Graphverarbeitungssystem Giraph (231) (basierend auf Googles Pregeg (111)), der Key Value Datenspeicher HBase (232) (basierend auf Googles BigTable (233)) und das Workflow Ausführungssystem Oozie (234). Aufgrund der freien Verfügbarkeit des Systems und seines Quellcodes, der große Anzahl an Werkzeugen und ergänzenden Systemen und des großen Benutzerkreises hat sich Hadoop mittlerweile zu einem de-facto Standard für massiv parallele Datenanalyse entwickelt. In den letzten Jahren wurden im Hadoop Umfeld viele Firmen gegründet, die eigene Softwarepakete und Dienstleistungen anbieten. Dazu gehören unter anderem die US-amerikanischen Firmen Cloudera, Hortonworks und MapR. Da Hadoop immer wichtiger wird und sich als eine wichtige Komponente in der IT-Infrastruktur von Firmen etabliert, haben viele Hersteller von Datenverarbeitungssystemen Schnittstellen zu Hadoop Map/Reduce und HDFS entwickelt.

Parallele Datenanalysesysteme: Die Popularität von Map/Reduce basiert zu einem großen Teil auf der freien Verfügbarkeit von Apache Hadoop und seiner einfachen Bedienung. Hadoop ist unkompliziert zu installieren und das Map/Reduce Programmiermodell ist sehr flexibel, so dass vielfältige

Analyseprogramme umgesetzt werden können. Allerdings gibt es immer wieder Kritik an dem ineffizienten Ausführungsmodell (235) und dem Programmiermodell, dass es zum Teil sehr umständlich macht bestimmte Algorithmen zu implementieren. Aufgrund dieser Schwächen wurden Systeme entwickelt, die effizientere Ausführung und intuitivere Spezifikation von Analyseprogrammen zum Ziel haben. Microsoft hat mit Dryad (98) ein sehr flexibles Datenverarbeitungssystem vorgestellt, das auf parallelen Datenflüssen basiert. So können viele Aufgaben wesentlich effizienter als mit Hadoop ausgeführt werden. Zur einfachen Spezifikation von Datenverarbeitungsaufgaben entwickelt Microsoft die Analysesprache SCOPE (236). Beide Systeme werden jedoch ausschließlich intern von Microsoft verwendet. Aus dem akademischen Umfeld sind das Asterix Projekt der University of California in Irvine und das Stratosphere System von der TU Berlin, HU Berlin und dem HPI Potsdam entwickelt worden. Während Asterix einen Schwerpunkt auf die Verarbeitung von semi-strukturierten Daten legt, generalisiert und erweitert Stratospheres PACT-Programmiermodell das Map/Reduce Modell und bietet bessere Effizienz bei gleicher Ausdrucksstärke. Vor kurzem wurde Hadoop um die Komponente „YARN“ ergänzt. YARN ist kein eigenes Datenverarbeitungssystem, stellt aber Infrastrukturdiensste für parallele Datenverarbeitungssysteme bereit. Diese Systeme können bei einem YARN Dienst Rechenressourcen aus einer Gruppe von Rechnern bestellen. Zum einen müssen Datenverarbeitungssysteme, die diesen Dienst in Anspruch nehmen, diese Funktionalität nicht selber implementieren. Zum anderen können auf diese Weise unterschiedliche Datenverarbeitungssysteme auf einem Rechencluster ausgeführt werden, der von einem YARN Dienst verwaltet wird. Microsoft arbeitet zurzeit an dem Infrastrukturdienst „REEF“, der auf YARN aufbaut und die Entwicklung von Datenverarbeitungssystemen weiter vereinfachen soll. REEF soll in Kürze unter einer Open Source Lizenz veröffentlicht werden.

SQL auf Hadoop: Das Apache Hadoop wurde als eine Open Source Implementation von Googles Map/Reduce Systems und seinem verteilten Dateisystem GFS gestartet (227). Ein großer Unterschied zu klassischen relationalen Datenbanksystemen ist, dass die Speicherung und Verarbeitung von Daten nicht auf das relationale Daten- und Verarbeitungsmodell beschränkt sind sondern sehr viele Freiheiten erlauben. Mit der steigenden Popularität des Open Source Systems Hadoop wuchs jedoch der Bedarf relationale Anfragen auch auf dieser Plattform auszuführen. Gründe hierfür sind zum einen der hohe Aufwand und das erforderliche Expertenwissen das nötig ist um Datenverarbeitungsprogramme mit dem Map/Reduce Programmiermodell zu implementieren. Im Gegensatz dazu ist SQL ein sehr weit verbreiteter Standard mit dem komplexe Anfragen in kurzer Zeit erstellt werden können. Hive ist Teil des Apache Hadoop Open Source Frameworks und stellt eine Data Warehousing Lösung für Hadoop dar (229). Hive speichert relationale Daten in HDFS oder HBase (232) und übersetzt Anfragen aus einem SQL-Dialekt in Hadoop Map/Reduce Anfragen.

Google hat mit Tenzing eine äquivalente Komponente für sein eigenes Map/Reduce System implementiert (237). In den letzten Jahren wurden auch viele Forschungsarbeiten für Ansätze publiziert, um die Speicherung und Abfrage von strukturierten Daten aus verteilten Dateisystemen wie dem HDFS zu verbessern. Dazu wurden viele Techniken, die bereits für (parallele) Datenbanksysteme entwickelt wurden, portiert. Zu diesen Techniken zählen Indizierung, spaltenorientierte Speicherformate und lokationsbewusste Speicherung von verknüpften Daten (238)(239)(240)(241). Andere Systeme, wie Hadapt (242), das aus dem universitären Forschungsprototypen HadoopDB (243) hervorgegangen ist, speichern die Daten in vielen unabhängigen relationalen Datenbanken und nutzen Hive und Hadoop zur parallelen Abfrage dieser Daten. Microsofts Polybase (244) verfolgt den umgekehrten Ansatz und ermöglicht die Verarbeitung von Daten, die im HDFS gespeichert sind, mit seinem relationalen Datenbanksystem SQL Server. IBMs BigSQL stellt eine SQL-Schnittstelle für ihr BigInsights System (245) dar, das ebenfalls auf Hadoop basiert. Mit Ausnahme von Polybase haben alle zuvor genannten Systeme gemeinsam, dass sie auf einem Map/Reduce-basierten Ausführungssystem gründen.

Vergleich von Map/Reduce-basierten SQL Systemen und parallelen Datenbanksystemen: Parallele Datenbanksysteme und Map/Reduce-basierte Systeme unterscheiden sich in ihren Hardwareanforderungen. Parallele Datenbanksysteme werden typischerweise auf besonders ausfallsicheren Rechnern mit viel Rechenleistung installiert. Ein nachträgliches Ändern der Konfiguration wie zum Beispiel das Hinzufügen oder Aufrüsten von Rechnern ist normalerweise nicht vorgesehen. Map/reduce-basierte Systeme können auf handelsüblicher Hardware ohne besondere Anforderungen hinsichtlich Ausfallsicherheit oder Leistung installiert werden. Ein Anpassen der Rechneranzahl ist in der Systemarchitektur explizit vorgesehen und ohne weiteres möglich. Aufgrund der geringeren Hardwareanforderungen und den entfallenen Kosten für Softwarelizenzen haben Hadoop-basierte Systeme geringere Anschaffungs- und Betriebskosten als parallele Datenbanksysteme. Darüber hinaus unterstützen Map/Reduce Systeme wegen der ausdrucksstarken Programmierschnittstelle und Dateisystem als Datenspeicher wesentlich mehr Anwendungsfälle als rein relationale Datenbanksysteme. In Bezug auf Verarbeitungseffizienz sind parallele Datenbanksysteme Map/Reduce-basierten Systemen überlegen. Das liegt zum einen an der typischerweise leistungsfähigeren Hardware als auch an der speziell auf diesen Anwendungsfall zugeschnittenen Systemarchitektur und Implementierung. Map/reduce Systeme können diesen Nachteil zwar mit ihrer guten Skalierung auffangen, benötigen jedoch mehr Hardwaredressourcen um die gleiche Ausführungsleistung zu erreichen.

Neue Systeme für analytische SQL-Anfragen: Aufgrund der eher hohen Latenz in der Verarbeitung von SQL-Anfragen auf Map/Reduce-basierten Systemen wurden Systeme entwickelt, die einige Vorteile von parallelen

Datenbanksystemen mit denen von Map/Reduce-basierten Systemen vereinigen sollen. Der erste Kandidat dieser Gruppe ist Googles System „Dremel“, das schnelle Aggregationen über sehr große Datenmengen ermöglicht (246). Dremel bietet eine reduzierte SQL-Schnittstelle, eine sehr gute Skalierung und nutzt ein verteiltes Dateisystem als Speicher. Im Gegensatz zu parallelen Datenbanken unterstützt Dremel auch verschachtelte Datenformate. Zusammenfassend, vereint Dremel die Latenz und Verarbeitungsgeschwindigkeit von parallelen Datenbanken und die Skalierungseigenschaften und Unabhängigkeit von Datenspeichern, die Map/Reduce auszeichnen. Impala (247) und Apache Drill (248) sind zwei Open Source SQL-Systeme, die ähnliche Eigenschaften wie Dremel aufweisen. Während Impala von der Firma Cloudera entwickelt wird, ist Drill ein Apache Projekt. Beide Systeme verfügen über die standardisierten Schnittstellen von klassischen Datenbanksystemen (ODBC/JDBC), so dass auch konventionelle BI Reporting-Werkzeuge zur Datenanalyse verwendet werden können. Im Vergleich zu Hadoop, sind die Implementierungen dieser Systeme deutlich hardwarenäher. So werden explizit Multi-Core CPUs unterstützt und die Charakteristiken des Hauptspeichers ausgenutzt, um geringe Latenz bei der Verarbeitung von sehr großen Datenmengen zu ermöglichen. Der Unterschied zu parallelen relationalen Datenbanksystemen besteht in zahlreichen Schnittstellen zu Speichersystemen. Während Datenbanksysteme in der Regel Datenimport erfordern und Daten intern speichern, bieten Systeme wie Impala oder Drill Schnittstellen zu zahlreichen Speichersystemen wie HDFS, HBase, Key-Value-Datenspeichern oder auch relationalen Datenbanken. Daten müssen nicht zunächst importiert werden und können aus unterschiedlichen Speicherformaten gelesen werden.

3.3.5 Systeme für maschinelles Lernen auf dem HDFS und im Hauptspeicher

Systeme für komplexe Datenanalysen auf großen Datenmengen: In den letzten Jahren sind zahlreiche Systeme entwickelt und eingeführt worden, die das Ziel haben maschinelle Lernverfahren und Data-Mining Algorithmen auf großen Datenmengen auszuführen. Mahout (230) und MADlib (249) sind Bibliotheken von implementierten Analysealgorithmen. Während Mahout Bestandteil des Apache Projekts ist und auf dem Map/Reduce System Hadoop ausgeführt wird, ist MADlib eine Bibliothek von Skripten und benutzerdefinierten Funktionen für die relationalen Datenbanksysteme PostgreSQL (250) und Greenplum (251). Beide Bibliotheken sind frei als Open Source verfügbar. SAP bietet für seine Hauptspeicherdatenbank HANA ebenfalls Bibliotheken und Funktionssammlungen für bestimmte Anwendungsfälle wie zum Beispiel das Trainieren von Vorhersagemodellen. IBMs BigInsights System (245) unterstützt die deklarative Definition und Ausführung von komplexen maschinel- len Lernverfahren. BigInsights nutzt Apache Hadoop als Ausführungsschicht.

Das analytische parallele Datenbanksystem EXASolution (252) des deutschen Unternehmens Exasol unterstützt die Ausführung von Skripten der Programmiersprache R (253), die für statische Analyse entwickelt wurde. Neben Bibliotheken und Sprachunterstützung für Map/Reduce und relationale Datenbanksysteme gibt es auch parallele Ausführungssysteme, die die Anforderungen von komplexen Datenanalyseverfahren erfüllen. Spark (254) wird an der UC Berkeley für interaktives Data-Mining und iterative Algorithmen auf großen Datensätzen entwickelt. Spark nutzt Hauptspeicher und verteilt Berechnungen über mehrere Rechnerkerne oder auch mehrere Rechner in einem Cluster. Spark ist Open Source und seit kurzer Zeit Teil des Apache Projekts. GraphLab (103) von der Carnegie Mellon Universität ist ein weiteres paralleles System für maschinelles Lernen auf graphisch-strukturierten Daten. In Deutschland wird das Stratosphere System (146) entwickelt, das effizient iterative Datenflüsse ausführen kann und damit ebenfalls die Voraussetzungen für viele Data-Mining Algorithmen und Verfahren des maschinellen Lernens erfüllt. Ebenso wie Spark und GraphLab, ist Stratosphere ein reines Datenverarbeitungssystem und stellt keinen eigenen Datenspeicher bereit. Diese Systeme können von unterschiedlichen Speichersystemen Daten einlesen, wie zum Beispiel Hadoops verteiltem Dateisystem HDFS. Eine weitere Gemeinsamkeit der drei Systeme ist ihr Ursprung aus akademischen Forschungsprojekten und die freie Verfügbarkeit des Quellcodes. In all diesen Prototypen ist eine Kommerzialisierung absehbar.

3.3.6 Datenstromverarbeitungssysteme

Die Datenbankforschung hat das Thema Datenstromverarbeitung aktiv in den frühen 2000er Jahren bearbeitet und erste akademische Prototypen (258) (282) (283) erstellt. Zu den kommerziellen Produkten zählen unter anderem Microsofts StreamInsight (255), SAPs Galaxy und IBM Infosphere Streams (256). Viele dieser Systeme erweitern das Modell der relationalen Algebra mit Operatoren zur Ausführung auf Fenstern in Datenströmen. Seit 2010 wurden eine Reihe neuer paralleler Datenstromsysteme vorgestellt. S4 (127) wurde ursprünglich von Yahoo entwickelt und ist mittlerweile ein Open Source Apache Projekt. Konzeptuell besteht ein S4 Stromverarbeitungsprogramm aus einer Anzahl von Verarbeitungsknoten die verteilt auf mehreren Rechnern ausgeführt werden. Datenelemente, sogenannte Events, werden nach einer Verarbeitungsvorschrift von Knoten zu Knoten geschickt. Dabei findet die Navigation von Events teilweise ähnlich wie in Map/Reduce anhand eines Schlüsselwertes statt. Fensteroperationen werden ebenfalls unterstützt. Damit lassen sich Map/Reduce-ähnliche Programme auch auf Datenströme anwenden. Zwei weitere Systeme, die ein ähnliches Verarbeitungsmodell wie S4 haben sind Storm (257) und Spark Streaming (161). Beide Systeme sind ebenfalls frei unter Open Source Lizzenzen verfügbar. Spark Streaming basiert auf dem Stapelverarbeitungssystem Spark und wird wie dieses an der Universität von Kalifornien, Berkeley entwickelt. Um

einen Datenstrom auf Spark auszuführen, wird der Datenstrom in viele sehr kleiner Stapel zerlegt und diese individuell verarbeitet. Dieser Ansatz ermöglicht die gemeinsame Verarbeitung von Datenströmen und statischen Daten, erlaubt aber nicht die Abarbeitung iterativer Algorithmen (elementar für maschinelles Lernen und die Berechnung von Vorhersagemodellen). Das Stratosphere System ist in der Lage Datenströme mit bestimmten Servicegarantien, wie zum Beispiel Verarbeitungslatenz oder Durchsatz, zu verarbeiten (106) (110). Das Naiad System (116), das von Wissenschaftlern aus der Forschungsabteilung von Microsoft entwickelt wurde, verarbeitet Iterationen auf Datenströmen mittels der Technik der differentiellen Berechnung (d.h., der Verarbeitung sehr kleiner Deltas in einer durch die iterativen Algorithmen bestimmten Ordnung). Allerdings ist Naiad nur ein Prototyp, dessen Kommerzialisierung noch bevorsteht.

3.3.7 Herausforderungen und Forschungsfragen

Wir gliedern und priorisieren Forschungsfragen nach „Erforschung neuer Funktionalität“, „Qualitative Verbesserung existierender Funktionalitäten“ und „Optimierung der Ausführungskosten“ bei gleicher Qualität einer existierenden Funktionalität.

Erforschung neuer Funktionalität

- Wie kann eine deklarative Datenanalysesprache in ein Datenverarbeitungssystem integriert werden, das Iterationen, Zustände und benutzer-definierte Funktionen unterstützt?
- Wie können Programme, die in so einer Sprache definiert sind, automatisch übersetzt und optimiert werden?
- Wie können diese Programme transparent auf einem Mehrkernrechner oder einem Rechnercluster parallelisiert und ausgeführt werden?
- Welche Eigenschaften braucht ein System um explorative Datenanalyse zu unterstützen?
- Wie kann man Analysespezifikation und Ergebnisvisualisierung effektiv kombinieren, um den Prozess der explorativen Datenanalyse zu vereinfachen und zu unterstützen?
- Wie kann man parallele Datenverarbeitungssysteme möglichst nahtlos in bestehende IT-Infrastruktur integrieren?
- Welche neuen Hardwareinfrastrukturen ermöglichen die massiv parallelen Verarbeitung von Datenströmen mit geringer Latenz?

Qualitative Verbesserung existierender Funktionalitäten

- Wie können bestehende Algorithmen zur Datenanalyse und des maschinellen Lernens in einer deklarativen Sprache spezifiziert werden und effizient parallelisiert werden?

Optimierung der Ausführungskosten

- Wie können die Eigenschaften von neuer Hardware optimal für Datenverarbeitung und Datenanalyse ausgenutzt werden?
- Gibt es spezielle Datenstrukturen oder Verarbeitungsmodelle, die die Eigenschaften neuer Hardware optimal ausnutzen?
- Welche neuen Hardwaretrends ergänzen sich besonders gut bei der Analyse von großen Datenmengen oder haben sogar Synergieeffekte?
- Wie müssen benutzerdefinierte Funktionen mit Ausführungssystemen integriert werden, um sie besser optimieren und effizienter auszuführen zu können?
- Wie muss man ein paralleles verteiltes Datenanalysesystem entwerfen, um sehr schnelle verteilte Iterationen ausführen zu können?

3.4 Datenbewertung und -verwertung auf Informationsmarktplätzen

Bedarf an großen Data Warehouses, z.B. für Competitive Intelligence und Marktforschung: Zurzeit lässt sich ein zunehmender Bedarf nach Analysen von sowohl frei verfügbaren Web- als auch von In-House-Daten beobachten. Eines von vielen Beispielen hierfür ist die Analyse von Internetdaten unter Zuhilfenahme von Sprachverarbeitungstechnologien und statistischen Methoden, z.B. für das Marketing, die Finanzindustrie oder zur Unternehmenssteuerung (Business Intelligence). Insbesondere mittelständische Unternehmen und Fachabteilungen in Konzernen verfügen oft über einzigartiges Wissen für die Übersetzung von Daten in monetär verwertbare Informationen, beispielsweise Fachwissen mit Bezug auf eine spezielle Nische oder Beziehungen zu potentiellen Kunden, die bereit sind, für diese Nischeninformationen zu bezahlen. Zur Analyse der Daten sind IT-Infrastrukturen notwendig, wie zum Beispiel sehr große Data Warehouse Systeme, die Web-Daten einsammeln und mit Informationen über ihre Herkunft, ihren Inhalt und ihre sprachliche Struktur anreichern.

Mehrere kleinere Unternehmen teilen sich die Kosten über einen Datenmarktplatz. In der Regel erfordern derartige Unterfangen erhebliche Infrastrukturstinvestitionen und erzeugen erhebliche weitere Kosten für den Betrieb und die Aktualisierung von Daten. Daher sind solche Lösungen insbesondere für Mittelständler und Fachabteilungen oft zu teuer. Seit kurzem beginnen Datenverkäufer, Data Warehouse Entwickler sowie Entwickler von domänenpezifischen Analysealgorithmen, ihre Verfahren auch als Plattform-, Software- und Data-as-a-Service-Modelle auf „Datenmarktplätzen“ anzubieten. Der zentrale Einstiegspunkt eines Datenmarktplatzes erleichtert den Zugang zu diesen Dienstleistungen und Daten. Zusätzlich fungiert der Datenmarktplatz als Kunden- und

Anbieter-übergreifende Datenintegrationsplattform, insbesondere für das kollektive Speichern, Analysieren und Wiederverwenden von Daten.

Beispielsweise kann ein Anwendungsprogrammierer für die Nischen-Domäne „Elektroautos“ auf Basis von offenen und kommerziellen Daten einem Markt-platz eigene Tools für die Marktforschung anbieten. Zur verbesserten Strukturierung der Rohdaten aus dem Web könnte der Anwendungsprogrammierer zusätzliche Spezialsoftware, wie Sentiment-Tagger, auf dem Marktplatz ein-kaufen und damit die Rohdaten für seine Kunden besonders gut nutzbar machen. Das heißt, der Anwendungsprogrammierer kann ein relevantes Problem für die Marketing-Fachabteilung lösen, ohne selbst die nötige Basistechnologie zu entwickeln oder sogar erst eine passende technische Infrastruktur aufbauen zu müssen. Somit werden nicht nur für Mittelständler und Fachabteilungen in Konzernen erhebliche Markteintrittsbarrieren eliminiert. Vielmehr ermöglichen Datenmarktplätze völlig neue Geschäftsmodelle mit Informationen und Analysewerkzeugen als elektronisch handelbare Güter.

3.4.1 Was ist ein Datenmarktplatz?

In diesem Abschnitt abstrahieren wir das Konzept des Datenmarktplatzes basierend auf Studien von (172)(173). Interviewpartner waren leitende Angestellte oder Geschäftsführer von Marktplatzbetreibern oder Unternehmen im Vertrieb von Daten, wie Thomson Reuters, SAP AG, IBM Deutschland, factual.com, Kasabi.com, Vico Research & Consulting GmbH, TEMIS Deutschland GmbH, Neofonie GmbH, Datamarket.com, mendeley.com, inovex GmbH und Herr Prof. Dr. Thomas Ruf, selbständige Fachkraft im Bereich Business Intelligence (vormals Managing Director GfK Retail).

Infrastruktur: Ein Datenmarktplatz stellt potentiellen Kunden sowohl Daten als auch datennahe Dienstleistungen (kurz: DdD) bereit. Oft können die Kunden die Daten oder datennahen Dienstleistungen ohne den Datenmarktplatz nicht oder nur zu unverhältnismäßig hohen Kosten, beziehen. Datenmarktplätze integrieren häufig Daten von öffentlichen Quellen (Web, UNO, Regierungen), mit exklusiven, nicht-öffentlichen Quellen (Börsendaten) und In-House Quellen (ERP-Systemen).

Häufig stellt der Datenmarktplatz auch die Infrastruktur zur Verarbeitung der Daten bereit. Kunden können mit der Basisfunktionalität Daten bereinigen, mit zusätzlichen Merkmalen anreichern, mit anderen Daten integrieren und in ein gemeinsames Schema überführen. Dieser Prozess folgt daher dem traditionellen Extract-Transform-Load Prozess (ETL), teilweise mit mehreren Transform-Load Iterationen. Oft wünschen die Kunden jedoch unterschiedliche und

diverse Cleansing- und Anreicherungstransformationen für die gleichen Rohdaten. Diese, oft hochgradig domänen spezifischen, Transformationen werden dynamisch von Dritt-Anbietern in Form von User Defined Functions (UDFs) auf dem Marktplatz zur Verfügung gestellt. Abbildung 1 zeigt eine für viele Märkte gültige Abstraktion des Datenmarktplatzes.



Abbildung 61: Sieben Anbietergruppen für Daten und datennahe Dienstleistungen teilen sich eine gemeinsame Plattform. Diese Zusammenarbeit ermöglicht die Berechnung des ausschlaggebenden Informationsvorsprungs für Entscheider. Quelle: Umfrage mit Experten und CEOs (172) (173).

Akteure und Ökosystem: Im Rahmen dieser Interviews konnten wir sieben Akteure identifizieren, die von einem Datenmarktplatz profitieren. Diese lassen sich in direkte Profiteure (Datenanbieter, Analysten, Anwendungsentwickler, Algorithmen-Entwickler, Datenmarktplatzbetreiber) und indirekte Profiteure (Berater und Zertifizierungsagenturen) unterteilen:

1. **Datenanbieter:** Wir unterscheiden zwei Gruppen von Datenanbietern, kommerzielle und nicht-kommerzielle Anbieter. In die erste Gruppe fallen „klassische“ Datenanbieter, wie zum Beispiel Reuters oder Bloomberg, die eine lange Historie im Bereich Finanzdaten (u.a.) haben. Zum anderen gehören in diese Gruppe auch Suchmaschinenbetreiber, wie Google und Bing, aber auch Anbieter von Linked Data und Web-Forenbetreiber, die ihre Inhalte monetarisieren wollen. Nicht-kommerzielle Datenanbieter sind vor allem Regierungen und Behörden, wie die UNO oder die Weltbank, die Ihre Daten meist kostenlos anbieten wollen oder per Gesetz die Daten zur Verfügung stellen müssen. Datenanbieter können dem Marktplatzgeiger einen Geldtransfer anbieten, um ihre Daten anbieten zu können. In der Regel nutzen Datenanbieter eher Marktplätze, um ihre Daten zu speichern und zu bewerben und sie so einem größeren Publikum zugänglich zu machen, was meist günstiger als Eigenvertrieb ist. Daher fließen meist Daten vom Anbieter zum Marktplatz und die Datenanbieter bekommen Geldflüsse in Form von Umsatzbeteiligungen. So entsteht eine Win-Win-Situation, da die Datenanbieter ein größeres Publikum und

eine höhere Monetarisierung erreichen, während der Marktplatzbetreiber durch ein besseres Angebot attraktiver für Kunden ist.

2. **Entscheider/Analyst/Professional:** Typischerweise besteht diese Gruppe aus Domain-Experten, wie M&A-Experten, Produktmanagern, Brokern, Marketing-Managern oder Business-Analysten. Vom Angebot des Datenmarktplatzes erhoffen sich Analysten, durch die Fülle an zur Verfügung stehenden Daten und deren Kombinationsmöglichkeiten häufig genauere Analyseergebnisse. Mitglieder dieser Gruppe stellen – je nach Expertise und technischen Möglichkeiten – strukturierte und textbasierte, ad-hoc-Anfragen an den Datenmarktplatz.
3. **Anwendungsentwickler /-verkäufer:** Häufig sind Analysten nicht in der Lage, ihre Anforderungen in einer formalisierten Sprache auszudrücken, so dass diese von einem Datenverarbeitungssystem ausgeführt werden können. Da auch das Finden und Erkunden von Datenquellen und -sätzen eine zeitintensive Arbeit ist, hat sich hierfür ein Markt für Anwendungsentwickler etabliert. Diese entwickeln Applikationen, die Anfragen an den Datenmarktplatz vereinfachen, in dem sie 'Wissen' über Datenschemata, Datenintegration und Quellenauswahl in vorkompilierte Anfragen übersetzen. Anwendungsentwickler können so Daten einkaufen und diese Daten für ein Problem einer nicht technischen Kundengruppe aufbereiten; auf der anderen Seite erhält der Marktplatzbetreiber Nutzungsentsgelte für die bereitgestellten Daten, die Nutzung der Infrastruktur und die Vermittlung von möglichen Komponenten zur Datenanreicherung und -bereinigung.
4. **Entwickler von Technologien (UDFs) für Datenveredelung:** Die beiden vorgenannten Akteure, Analysten und Anwendungsentwickler müssen häufig große Datenmengen aus meist volatilen und heterogenen Datenquellen integrieren. Dafür werden Algorithmen, auch UDFs, für Data-Mining, Matching, Cleansing, Relevance Computation und Lineage eingesetzt. Diese UDFs sind häufig Nischenprodukte für eine bestimmte Domäne, Sprache, Datenquelle oder Funktionalität. Ein Datenmarktplatz bietet den Entwicklern der UDF die Möglichkeit, die Algorithmen als externes Modul zu hinterlegen. Auf diese Weise können andere Akteure diese Algorithmen kaufen oder mieten, um sie für ihre Zwecke einzusetzen, was das Angebot des Marktplatzes aufwertet. Im Gegenzug werden Entwickler abhängig von der Nutzung der durch sie erstellten UDFs entlohnt.
5. **Marktplatzbetreiber:** Der Betreiber ist das Wirtschaftssubjekt, das die Aufwendungen und das Risiko für den Betrieb der technischen Infrastruktur trägt, dafür aber die Entgelte der Marktteilnehmer vereinnahmt. Dabei sehen sich Betreiber mit einer Vielzahl an technischen, ethischen, rechtlichen und wirtschaftlichen Herausforderungen konfrontiert. Die wahrscheinlich größte ökonomische Herausforderung ist der Aufbau einer vertrauenswürdigen Marke und die Etablierung einer genügend großen Community. Eine wichtige technische Herausforderung ist die Entwicklung einer leistungsfähigen Verarbeitungs- und Angebotsplattform,

die es zum einen ermöglicht, Daten und Dienste zu speichern, zu durchsuchen und auszutauschen, und zum anderen die hierdurch entstehenden Kosten (Ausführungszeit) zu minimieren. Aufgrund der derzeitigen unsicheren Rechtssituation in Europa (siehe Anhang) können rechtliche Herausforderungen bzgl. des innereuropäischen Datenhandels ein Innovationshemmnis darstellen, da hierdurch unkalkulierbare Risiken entstehen. Häufig erfordern die Anfragen der Analysten die Integration mehrerer Datenquellen. Auf diese Weise formalisieren sie das häufig domänenspezifische Wissen über Datenintegrationschritte für bestimmte Nischen-Datenquellen. Dieses Wissen kann vom Marktplatzbetreiber gesammelt und Kunden angeboten werden. Daneben profitiert der Marktplatzbetreiber direkt durch die Entgelte, die Analysten für die Daten und die Nutzung von Anwendungen auf dem Marktplatz entrichten.

6. **Berater.** Die Interviewpartner wiesen häufig auf den Beratungsbedarf im Umfeld eines Marktplatzes hin. Die Beratung richtet sich dann vornehmlich an die Nutzer der Plattform und soll diese unter anderem dabei unterstützen, Quellen auszuwählen oder eigene Produkte auf Basis der vorhandenen Daten zu entwickeln. Der Datenmarktplatz generiert also indirekt Kunden für Beratungsgesellschaften. Allerdings profitiert auch der Marktplatzbetreiber, da so einerseits eine Nutzung sichergestellt wird und andererseits die Kunden eine technische Betreuung erfahren, für die der Betreiber nicht aufkommen muss.
7. **Zertifizierungsagenturen.** Die Aufgabe dieser Gruppe ist, die Sicherheit der Plattform und die Qualität der Daten und Prozesse zu zertifizieren. Durch entsprechende Zertifikate, für die in der Regel die Kosten vom Betreiber des Marktplatzes an die Kunden weitergegeben werden, kann die Plattform das Vertrauen bei den Kunden erhöhen und Marktanteile gewinnen. Verschiedene Akteure können mehr als eine Rolle annehmen; beispielsweise kann ein Analyst zunächst Rohdaten kaufen, bereinigen, anreichern und aggregieren und diese Daten wieder verkaufen, oder Datenanbieter entwickeln eigene Integrationsalgorithmen und bieten diese an.

3.4.2 Preismodelle für Daten und Datenaustausch

Die Autoren in (172) (173) leiten drei grundlegende Prinzipien ab, die den Preis für Daten und datennahe Dienstleistungen beeinflussen:

Daten ermöglichen einen verwertbaren Informationsvorsprung: Das subjektive Empfinden eines Kunden für einen durch Daten oder Dienste, erhaltenen Informationsvorsprung ist abhängig von dessen Domänenwissen. Oft fehlt jedoch einem Datenanbieter das Verständnis des „Hebels/Wertes“ dieser Daten für den Kunden. Die Diskrepanz zwischen dem Wissen des Anbieters und des Nachfragers über den Wert der Daten ist ein wichtiger Bestandteil des

Preisbildungsprozesses. Die Verwertbarkeit der Daten impliziert auch eine hohe Datenqualität, dazu gehören insbesondere berechenbare Kriterien, wie Vollständigkeit und Dichte, Nachvollziehbarkeit, Integrierbarkeit etc.

Exklusiver Zugang zu den Daten innerhalb der relevanten Peer-Gruppe des Kunden:

Der Kunde kann einen Informationsvorsprung häufig nur realisieren, wenn er exklusiv über die Daten verfügen kann. Aus den Interviews konnten wir exemplarisch folgende Mechanismen identifizieren, die Exklusivität sicherstellen: Ein Anbieter sichert durch hohe Infrastrukturausgaben eine hohe Qualität, z.B. Datenfrische für schnell alternde Daten und reduziert dadurch den Anreiz der Kunden die Daten (illegal) zu kopieren und weiterzugeben. Ein anderes Beispiel: Ein Anbieter verfügt über eine möglichst vollständige Kopie des Webs und über eine Infrastruktur zur Erkennung von Dubletten. Anhand der geschätzten Vollständigkeit der Kopie des Webs kann der Anbieter dem Käufer aufzeigen, ob eine Information bereits im Web kopiert wurde.

Möglichst umfängliche Nutzungs- und Wiederverwertungsrechte: Viele Datenanbieter schränken die Nutzungsrechte vertraglich ein. Dazu gehören Einschränkungen bezüglich der zeitlichen Nutzungsdauer oder der Weitergabe der Daten an Kunden des Käufers.

Die Autoren identifizieren auch sechs atomare und zwei hybride (zusammengesetzte) Preismodelle.

Atomar: Kostenlos-Modell. Statistische Daten von Behörden, wie zum Beispiel <http://data.gov.uk/> oder Bundesministerien und statistische Daten, stehen oft kostenlos im Internet zur Verfügung. Es ist unwahrscheinlich, dass diese Daten für Geld verkauft werden können, nur weil sie auch auf einem Markt-platz angeboten werden. Dennoch können kostenlose verfügbaren Daten einem Marktplatz helfen, Kunden zu gewinnen, was wiederum hilft, Anbieter von kommerziellen Daten zu gewinnen. Darüber hinaus können kostenlose Daten mit In-house oder privaten Daten integriert werden und so neue Werte schaffen. Ein Spezialfall ist, dass manche Behörden oder Non-Profit-Organisationen ihre Daten veröffentlichen müssen. Oft kann es für diese Organisation preiswerter sein, die Publikation auf den Marktplatz auszulagern und dafür den Betreiber zu entlohnern.

Atomar: Tauschgeschäft. Es kommt in der Praxis vor, dass ein Datenhändler Daten kostenlos zur Verfügung gestellt bekommt, wenn der Datenanbieter im Gegenzug die Daten in veredelter Form zurückbekommt. Das ist unter anderem immer dann der Fall, wenn Daten aus verschiedenen Quellen aggregiert als Bericht veröffentlicht werden. Es ist offensichtlich, dass dieses System nur funktioniert, wenn es neben den Datenlieferanten noch weitere Abnehmer für das Endprodukt gibt.

Atomar: Festpreis/Subskription/Abonnement. Dieses Preismodell hat als einzigen Parameter die Nutzungsdauer, für die ein (häufig unbegrenztes) Nutzungsrecht eingeräumt wird. Bei unseren Gesprächspartnern beobachteten wir dieses Preismodell vor allem in Bezug auf Software-Lizenzen und Software-Hosting. Das Festpreismodell bietet sowohl Nachfragern auch als Anbietern eine hohe Planungssicherheit (letzteren nur bzgl. der Einnahmen, nicht bzgl. der Nutzung). Das Modell ist nicht flexibel und ermöglicht Overselling, also das Verkaufen durch den Anbieter von mehr Leistung, als eigentlich vom Nachfrager benötigt wird.

Atomar: Paketpreise/Tiered Pricing. In diesem Modell werden dem Kunden unterschiedlich gestaffelte Pakete (bspw. API-Aufrufe, Anzahl der Datenquellen, Menge an Daten) zu einem festen Preis und für eine bestimmte Nutzungsdauer angeboten. Im Unterschied zum Festpreis wird neben der Zeit auch die Menge festgelegt. Paketpreise ermöglichen Arbitrage. In der Regel sind größere Pakete insgesamt teurer, aber relativ pro Einheit günstiger für den Nachfrager. Einige Interview-Partner bevorzugen dieses Modell anstelle des Festpreismodells aufgrund der Flexibilität und der Planungssicherheit. Wir beobachten dieses Preismodell auch bei Microsoft mit dem Kerngeschäft des Vertriebs von Office-Software/Betriebssystemen und als Nebengeschäft der Zugriff auf den Bing-Web-Index.

Atomar: Zahle-nur-was-Du-nutzt/Pay-per-Use (PPU). Dieses Preismodell setzt einen Preis für jede verbrauchte Einheit fest und bemisst den Endpreis anhand der Aggregation der verbrauchten Einzeleinheiten. Unsere Interviewpartner verwenden dieses Preismodell vorwiegend für die Abrechnung von Beratungszeit. Dieser volumenbasierte Ansatz wird dann uninteressant, wenn die Grenzkosten der Datenbeschaffung und -verarbeitung gegen Null konvergieren. Wir erwarten, dass der Trend fallender Grenzkosten anhält und somit volumenbasierte Preise weiter an Akzeptanz verlieren. Dieses Preismodell ist auch aufgrund der inhärenten Planungsunsicherheit (in Bezug auf erwartete Absätze) nur bedingt anwendbar für Betreiber von Marktplätzen.

Atomar: Progressiver Preis. In diesem Preismodell richtet sich der Preis eines Datensatzes nach dem Zeitpunkt des Erwerbes. Der Preis erhöht sich, je mehr Kunden eine nicht verkäufliche Lizenz zur Nutzung des Datensatzes erwerben. Dieses Preismodell wird vor allem dann angewandt, wenn die Verbreitung limitiert werden soll (künstlerische Werke) oder der Anbieter des Datensatzes einen Anreiz für einen frühzeitigen Erwerb setzen möchte, zum Beispiel in der kommerziellen Bilddatenbank <http://www.dreamstime.com>.

Hybrid: Zweigeteilter Tarif/Two-Part-Tarif. Der zweigeteilte Tarif ist eine Kombination aus einer Grundgebühr (Festpreis, ggf. progressiv) und entweder einem direkten (Pay-per-Use) oder indirekten (Paketpreis) bzw.

verbrauchsabhängigen Anteil. Dieses Preismodell wurde in der Vergangenheit häufig von Stromanbietern oder von Mobilfunkbetreibern verwendet und repräsentiert die hohen Investitionskosten dieser Branchen für Infrastrukturen. Wir beobachten dieses Preismodell für Software-Lizenzen, in dem der Basispreis die Kosten für die Software deckt und dann Daten nutzungsabhängig berechnet werden. Häufig wird der feste Teil genutzt, um einen sicheren Deckungsbeitrag zu erzielen, und die variable Vergütung dient der Kundendiskriminierung nach Zahlungsbereitschaft.

Hybrid: Freemium. Freemium ist ein Portmanteau aus free, also kostenlos, und Premium. Die Idee dabei ist, dass Benutzer grundlegende Produkte und Dienstleistungen kostenlos benutzen können, aber für den vollen Funktionsumfang ein Entgelt bezahlen müssen. Das Preismodell für die Premiumleistungen kann dabei jede der diskutierten Formen annehmen. Einige unserer Gesprächspartner verfolgen genau diesen Ansatz, um eine zahlreiche Teilnahme an ihren Diensten zu gewährleisten.

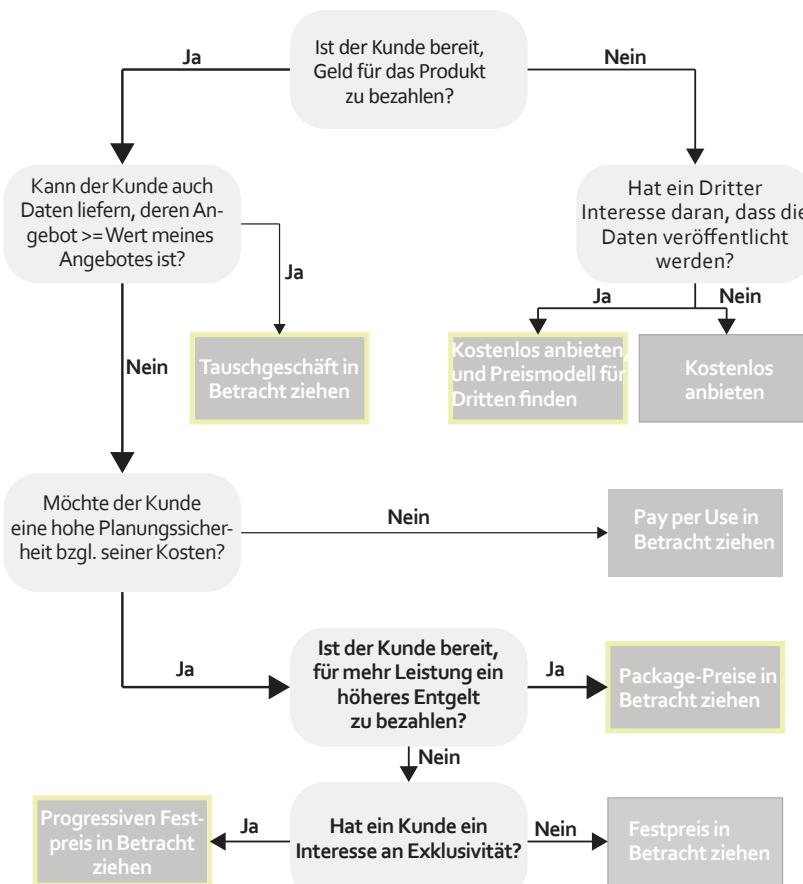


Abbildung 62: Das Diagramm zeigt eine Preismodellwahl, die sich an der Nachfrage des Kunden orientiert.

Durchläuft man diesen Entscheidungsbaum von oben nach unten, so erhält man eines der atomaren Preismodelle als Ergebnis. Dabei ist zu beachten, dass dieses Diagramm für jedes Produkt erneut durchlaufen werden sollte. Wir stellen derzeit einen Bedarf nach einer Test-getriebenen Preisfestsetzung fest, insbesondere in häufig monopolistischen

Märkten, für Daten: Der Anbieter von Daten versucht durch Hypothesen einen Preis herauszufinden, der am Markt durchsetzbar ist. Im Rahmen unserer Interviews konnten wir Parameter für die Beeinflussung der Preise und Preismodelle identifizieren. Die Preismodelle kostenlos, Tauschgeschäft, progressiver Festpreis und Freemium (je nach Ausgestaltung) haben keinen variablen Anteil und sind einfacher zu implementieren als die drei Modelle mit einem variablen Anteil. Letztere sind nur unter der Voraussetzung möglich, dass die Form des Angebotes, also wie die Daten zugänglich gemacht werden (API, Web-Interface, etc.), eine zeitnahe Abrechnung der anfragespezifischen Nutzung durch Kunden ermöglicht.

3.4.3 Nachvollziehbare Daten: Qualität, Ursprung und Verfälschung von Daten

Aufgrund der vielfältigen „Kundschaft“ auf einem Datenmarktplatz werden Prinzipien des Nachweises des Ursprungs und zur Messbarkeit der Qualität besonders wichtig.

Stand der Technik

Systeme innerhalb eines Unternehmens, wie Data Warehouses, werden für die Analyse eines eingeschränkten Kreises von Datenproduzenten optimiert. Ein klassisches Data Warehouse im Handel verarbeitet beispielsweise Transaktionen zwischen Verkäufern und Kunden. In diesem Szenario kann ein einzelnes Datum auf den Erzeuger der Daten, z.B. die Verkäuferin am Point-of-Sale, zurückgeführt werden. Analog können auch häufig Daten auf den Erzeuger zurückverfolgt werden, wenn diese durch **autorisierte Maschinen**, z.B. aus einem **Produktionsprozess**, erzeugt werden (siehe Abschnitt 2.4.2) oder durch autorisierte Stellen veröffentlicht werden, z.B. **Daten aus bundes- bzw. landeseigene Organisationen** (siehe Abschnitt 2.4.1). Die Nachvollziehbarkeit von strukturierten Daten aus diesen autorisierten Quellen in einem Relationem DBMS bzw. einem Data Warehouse wurde in der Vergangenheit schon recht gut untersucht, siehe auch Referenzen (1) (2) (3) (4) und Verweise innerhalb dieser Arbeiten.

Anforderungen

Nachvollziehbarer Ursprung und Weg für Daten aus unsicheren Quellen: Verfahren die, im Rahmen des Datenschutz erlaubte, Nachweise der Urheberschaft für unsichere Quellen erbringen, z.B. für Webdaten, sind besonders erforschungswürdig. Besonders gravierende Szenarien der **Betrugserkennung** stellten wir in Abschnitt 2.4.5 dar, z.B. zur Erkennung von urheberwidrig verwendeten Produktbezeichnungen, Urheberrechtsverletzungen von Textbeiträgen von Verlagen oder dem Diebstahl von Identitäten zur Verleumdung im Web, zur Vermeidung von Steuerzahlungen auf Verkaufsportalen oder sogar

zur Planung von terroristischen Akten. Daher ist in diesen Anwendungsbereichen die Nachvollziehbarkeit von Daten bis zum Urheber/Erzeuger der Daten deutlich schwieriger. In der **Marktforschung** (siehe Abschnitt 2.4.4) und dem **Gesundheitswesen** (siehe Abschnitt 2.4.3) werden Erzeuger von Daten aber auch Menschen häufig ihre Identität jedoch nicht offenlegen wollen (siehe auch Konzept des Erlaubnisvorbehalts im Datenschutz in Abschnitt 4.3.3) bzw. absichtlich ihre Identität verschleiern oder verfälschen. Unternehmen interessieren sich in diesen Szenarien, und ausschließlich im Rahmen der datenschutzrechtlichen Grundlagen, mehr für den **Weg der Verbreitung einer Information (Informationskaskaden)**, auch über verschiedene Medien (Blogs, Foren, News, Bild, etc.) und einzelne Communities hinweg.

3.4.4 Herausforderungen und Forschungsfragen

Das Internet hebt den Handel mit Informationen auf eine neue Ebene. Niemals zuvor war die Informationsbasis bei hoher Verfügbarkeit derart umfangreich. Dies gilt insbesondere für soziale Netzwerke. Menschen sind bereit kleinste Details ihres Verhaltens preiszugeben. In der Folge entstehen sehr große Datenmengen aus welchen anhand geeigneter Extraktionsverfahren Informationen gewonnen werden können. Dies stellt insbesondere für Marktforschungsunternehmen eine Chance, gleichwohl Herausforderung dar. Datenmarktplätze sind dabei der Ort, an dem Angebot und Nachfrage nach Informationen in Form von Daten überein gebracht werden. Sie sind Ergebnis eines gesteigerten Bedarfs an zeit- und sachgerechter Bündelung von Ressourcen.

Betreiber von Datenmarktplätzen stellen in unserer Betrachtung die notwendige Infrastruktur bereit. Sie ermöglichen den Handel mit Daten, Extraktoren oder Inferenzmaschinen, indem sie Lagerung, Transport und Produktionsmittel zur Verfügung stellen. Auf diesen Marktplätzen stellen Anbieter Daten oder Veredelungsverfahren zur Verfügung deren Nachfrager mit Geld oder wiederum durch Abgabe von Informationen bezahlen. Nicht selten tauschen Anbieter und Nachfrager hierbei ihre Rollen; auch ist es denkbar, dass der Marktplatzbetreiber zusätzlich als Anbieter oder Nachfrager auftritt. Neben der technischen Herausforderungen der Datenextraktion, Datenveredelung, Datenspeicherung und des Datentransports, ist unseres Erachtens eine ökonomische Analyse und Evaluation der Mechanismen notwendig, welche einen Datenmarktplateau überhaupt erst ausreichend attraktiv machen, um eine breite Teilnahme zu gewährleisten. Namentlich sind dies die Preisbildungsmechanismen.

Die Erwartung der Marktteilnehmer an eine transparente und intersubjektiv nachvollziehbare Preisbildung wird durch die Eigenarten des Gutes Information verschärft. Denn (1) konkretisiert sich die abstrakte Wertvorstellung einer bestimmten Information meist erst durch das Ergebnis der Entscheidung, deren Grundlage sie war, (2) ist der funktionale Zusammenhang von Nutzen und Zeit

bei Informationen atypisch. (3) gehen die Grenzkosten und somit mittelfristig die Grenzpreise ein und derselben Information gegen Null und (4) ist die Nachfrage nach Informationen meist heterogen wodurch sich die Dichtefunktion der Longtail-Verteilung nach rechts verschiebt.

Wir gliedern und priorisieren Forschungsfragen nach „Erforschung neuer Funktionalität“, „Qualitative Verbesserung existierender Funktionalitäten“ und „Optimierung der Ausführungskosten“ bei gleicher Qualität einer existierenden Funktionalität.

Erforschung neuer Funktionalität

- Welche Operatoren auf Daten sind besonders wünschenswert für die Bereiche Überwachung oder Vorhersage, z.B. Vergleich, Ausschnitt/CUBE, Aggregation oder Anreicherung?
- Welche neuen messbaren Datenqualitätsmerkmale erfordern diese Operatoren für welche Daten (z.B. Text, Relationale Daten), z.B. die Messung von Exklusivität oder Integrierbarkeit für Textdaten?
- Welche typischen Wertschöpfungsketten finden in der Praxis schon jetzt Anwendung und in welchen Branchen? Welche Operatoren beinhalten diese Wertschöpfungsketten?
- Welche transaktionalen Daten aus dem Marktfluss fließen in die Preisbildung mit ein?
- Können wir aus den transaktionalen Daten eine maschinelle Abschätzung von Preisober- und untergrenzen, i.B. auch in Monopolmärkten bzw. Oligopolmärkten, möglichst automatisiert ermitteln?

Qualitative Verbesserung existierender Funktionalitäten

- Wie können wir den Ursprung auch von unsicheren Daten, z.B. aus dem Web bzw. dem Social Web, nachvollziehen? Können wir automatisiert ermitteln, wer und warum, Daten weitergab, verfälschte, löschte, ergänzte oder auch nicht weitergab?
- Wie können wir genauer die Exklusivität und Integrierbarkeit von Daten messen?
- Können wir Feedback des Konsumenten bei der Berechnung der Expertise des Datenproduzenten bzw. „Datenweitergebers“ der Datenwertschöpfungskette einbeziehen?

Optimierung der Ausführungskosten

- Verbesserung von Dublettenerkennungsverfahren (als wichtige Vorarbeit zur Erkennung der Nachvollziehbarkeit), i.B. auch für Social Media-, Text-, aber auch Audio- und Videodateien.
- Effizientere Speicherung und rekursive, iterative Verarbeitung besonders hochdimensionaler und „dünner“ Datenräume, z.B. mit Operatoren auf Tensoren und Matrizen, sowohl in Hauptspeicherdatenbanken als auch in einem verteilten Dateisystem auf Festplattenbasis.

3.5 Zusammenfassung

Explorative Datenanalyse für verteilte heterogene und volatile Daten. Informationen sind häufig auf viele Quellen verteilt (z.B. Nachrichtenportale, Reiseblogs, soziale Netzwerke, Webservices, etc.). Zur Beantwortung komplexer Fragen (z.B. „Welche Astronauten von Apollo-Missionen waren bereits auf dem Mond?“, „Wo ist das nächstgelegene italienische Restaurant, das überdurchschnittliche Bewertungen erhalten hat?“, „Welche Sehenswürdigkeiten sollte ich in welcher Reihenfolge besuchen?“) müssen oft mehrere Anfragen vom Nutzer an verschiedene heterogene Quellen (Texte, Bilder, Videos, etc.) gestellt werden. Die Ergebnisse müssen dann manuell zusammengesetzt werden.

Unterstützung des menschlichen „Trial and Error“ Vorgehens bei der Analyse: Ein wirklicher Mehrwert für den Nutzer entsteht durch intelligente Methoden zur automatischen Informationsextraktion und –aggregation, um komplexe Fragen zu beantworten. Der Datenanalyseprozess ist in diesem Fall explorativ und iterativ (engl. „Trial and Error“). In einer Suchphase werden dabei zunächst die relevanten Daten identifiziert. In einer Lernphase wird der Kontext zu diesen Daten beschrieben, typischerweise durch voneinander abhängige Datenverteilungen. In der Explorationsphase ermöglichen verschiedene Operationen das Ableiten von Entscheidungen aus den Daten bzw. Umformen und Anreichern der Daten.

Wesentlich für die Analyse von „Big Data“ sind daher **Algorithmen für skalierbare Computerlinguistik, Signalverarbeitung, modellbasierte Analyse und maschinellem Lernen**. Diese Verfahren müssen „schnell“ genug sein, um menschliche Interaktion und visuelle, interaktive Analysen auf riesigen Datensätzen zu ermöglichen.

Programmiermodelle für die „Benutzung“ dieser Algorithmen müssen für die breite Masse der Analysten alltagstauglich sein, offen für vorkomplizierte Programme dritter Anbieter sein, sogenannter benutzerdefinierter Funktionen (engl. user-defined function, UDF), sowie der Verarbeitung von iterativen Algorithmen und der Verwaltung und den Zugriff auf einen (für „Big Data“ potentiell verteilten) Zustands.

Ein wichtiges Beispiel ist die **Sprachverarbeitung** und Techniken der Computerlinguistik. Diese korrigieren in vorberechneten Korrelationen widersprüchliche, fehlende oder unvollständige Informationen durch den gezielten Einsatz menschlicher Interaktion. Die Maschine sucht dabei diese qualitativ minderwertigen Informationen aus einem „Big Data Pool“ heraus und zeigt nur diese Informationen einem Menschen zur Korrektur. Häufig erfolgt die Korrektur durch den Menschen implizit, beispielsweise in seiner gewohnten Anwendungsumgebung wie einem Spreadsheet. Aus den Interaktionen lernt die Maschine

zusätzliche Regeln, die eventuell in den maschinell bereits erfassten und vorberechneten Daten noch nicht enthalten sind.

Datenmärkte und Preisbildung: Die Pionierarbeit (173) (172) an deutschen Hochschulen und der deutschen Industrie an der rechtskonformen Bewertung und Verwertung von Daten ist einzigartig in der Europäischen Union und kann für den Dialog mit Gremien und Entscheidern der Bundesregierung und der Europäischen Union herangezogen werden. Die Ergebnisse möglicher Projekte im Bereich der Bewertung und Verwertung von Daten können bei der Erstellung von Musterverträgen, Checklisten und Leitfäden der Wirtschaft zur rechtskonformen Ausgestaltung ihrer Geschäftsbeziehungen dienen. Zusammen mit den aufgestellten Preismodellen wird es möglich sein, einen sicheren und fairen Datenhandel zu etablieren. Dafür müssen vor allem technische Probleme gelöst werden, wie die **Berechnung von Domänen-unabhängigen und abhängigen Datenqualitätsmerkmalen**. Diese Merkmale und die Merkmale der Nachfrage potentieller Käufer der Daten sollten eine Preisberechnung auch in monopolistischen Märkten, der Standardmarkttyp für Daten, ermöglichen und auch Daten-gegen-Daten Tauschgeschäfte stärker ermöglichen.

4 Darstellung rechtlicher Rahmenbedingungen in Deutschland

Die technische Entwicklung der letzten Jahre hat die verfügbaren Datenmengen erheblich vergrößert. Der Einsatz der in den vorherigen AP beschriebenen Technologien und deren Anwendungsfelder ermöglichen nicht nur die quasi unbegrenzte Langzeitspeicherung von Daten, sondern auch deren detailgenaue Analyse im Hinblick auf Nutzerverhalten und -interessen. Diese Form der Datenspeicherung und –nutzung betrifft dabei alle Bereiche des täglichen Lebens und erlangt zunehmend wirtschaftliche Bedeutung. Der Datenumgang wirft allerdings auch vielfältige rechtliche Fragen auf, die vor allem die Rechtsdisziplinen des Datenschutz-, Urheber- und des Vertragsrechts betreffen.

Allen voran ist bislang ungeklärt, ob ein **Eigentumsrecht an Daten** besteht und wie die rechtliche Zuordnung von Daten erfolgt. Daten haben sich zu einem bedeutsamen Handelsgut entwickelt, so dass auch die Frage nach der Inhaberschaft an Daten im Hinblick auf das Big Data-Management der Klärung bedarf. Daneben stellt die Wahrung der Grundsätze des deutschen **Datenschutzrechts** eine zentrale Herausforderung dar. Beachtlich ist hierbei zunächst die Frage nach der grundsätzlichen Anwendbarkeit der deutschen Rechtsordnung. Daten sind heute, besonders in der Cloud, hochgradig standortflexibel und ihre Speicherung und Verarbeitung kann an verschiedenen Orten gleichzeitig stattfinden. Soweit deutsches Datenschutzrecht anwendbar ist, steht die Datenverarbeitung unter einem Verbot mit Erlaubnisvorbehalt. Es gilt, die Erlaubnisstatbestände auf ihre Vereinbarkeit mit bzw. ihre Anwendbarkeit auf Big Data zu untersuchen. Besondere Bedeutung erlangen hier die Anonymisierung und Pseudonymisierung personenbezogener Daten, die den Datenumgang legitimieren können. Gleichermassen von Interesse ist in diesem Zusammenhang auch der grenzüberschreitende Datentransfer, etwa im Rahmen der Auftragsdatenverarbeitung. Die systematische Speicherung großer Datenmengen im Rahmen von Big Data wirft auch **urheberrechtliche Fragen** auf, vor allem nach den Rechten des Datenbankherstellers gem. §§ 87 ff. UrhG. Datenbanken können sowohl einen urheberrechtlichen Schutz als auch einen sui-generis-Schutz genießen. Die Auswertung von Datenbanken im Rahmen des Big Data-Managements kann folglich auch das Urheberrecht verletzen. Schließlich ergeben sich beim

Umgang mit großen Datenmengen **vertragsrechtliche und haftungsrechtliche Probleme**, denn die Qualität einer Information spielt für das Big Data-Management eine bedeutende Rolle. Falsche Informationen und Daten können in der Praxis erhebliche Konsequenzen haben.

Die aufgeworfenen Rechtsfragen und damit verbundene Sonderprobleme können nicht im Rahmen dieses Kurzgutachtens vollständig beleuchtet werden. Eine ausführliche Analyse der Berührungspunkte von Big Data und Recht sowie konkreter Anwendungsfelder, die bereits in den vorherigen AP angesprochen wurden, befindet sich im Anhang zu dieser Studie. Auf diese sei an dieser Stelle verwiesen.

4.1 Eigentum an Daten

Täglich werden unbegrenzte Datenmengen im Internet generiert, verarbeitet oder vermittelt. Der Umgang mit Daten stellt einen erheblichen Wirtschaftsfaktor dar. Dabei fließen die Datenströme, ohne dass eine zentrale Rechtsfrage geklärt ist. Denn nach heutigem Stand der Rechtswissenschaft ist noch unklar, ob Daten eigentumsfähig sind und, falls dies so ist, wem das Eigentum daran zu steht. Die Klärung dieser Frage ist für die weitere Diskussion über die Möglichkeiten und Grenzen des Big Data-Managements jedoch von zentraler Bedeutung. Hieraus ließe sich ableiten, ob Big Data-Unternehmen eine gesicherte Rechtsposition im Hinblick auf die von ihnen verarbeiteten Daten haben. Auch auf insolvenzrechtliche Fragestellungen hätte die Klärung der Eigentumsfähigkeit Einfluss.

4.1.1 Eigentumsfähigkeit

Eigentumsfähigkeit setzt voraus, dass Daten nach der Rechtsordnung einem Rechtssubjekt zugeordnet werden können. Der Datenbegriff setzt sich aus zwei Elementen zusammen: Inhalt und Darstellung (58).² Rechtlich sind also drei Ge genstände getrennt zu betrachten: Erstens das Medium, als physisch vorhandene Sache oder Eigenschaft einer Sache (Darstellung), zweitens die Information, also der Inhalt, und drittens das davon abzugrenzende Datum als „Gebilde“. Be grifflich scheinen Daten von Sache und Information trennbar zu sein. Die Frage nach der Eigentumsfähigkeit von Daten setzt aber voraus, dass diesen auch eine rechtliche Eigenständigkeit zukommt.

² Welp, IuR 1988, 443 (444).

Ein Medium ist als Sache unabhängig von der darauf oder darin enthaltenen Information, fällt zivilrechtlich unter den Begriff der Sache nach § 90 BGB³ und damit auch unter den strafrechtlichen Sachbegriff der Vermögensdelikte (36, 54, 45, 30).⁴ Es kann somit dem Eigentum einer Person nach § 903 BGB unterliegen. Inhalte können ihrer vielfältigen Gestalt entsprechend von unterschiedlichen Rechten erfasst sein. Ihrem Schaffensprozess oder der Investition nach schutzwürdige Inhalte unterliegen dem Urheberrecht oder den gewerblichen Schutzrechten, personenbezogene Daten unterfallen hingegen dem Datenschutzrecht. Teilweise können diese Rechte zwar unter den Eigentumsbegriff des Art. 14 Abs. 1 GG fallen,⁵ ihre Existenz wird aber bereits durch die entsprechenden Gesetze (UrhG, GeschmacksmusterG etc.) gewährleistet und unterliegt nicht dem zivilrechtlichen Schutzregime für Sacheigentum (32). Die Beziehung einer Person zu einem Medium hat keinen Einfluss auf die Zuordnung von Daten zu Personen. Zwar müssen Werke, um Urheberrechtsschutz zu genießen, wahrnehmbar sein. Eine körperliche Fixierung ist jedoch dafür nicht zwingend notwendig.

4.1.2 Eigentumsrechtliche Zuordnung

Ausgangspunkt der weiteren Überlegungen ist das Eigentum an dem jeweiligen Medium. Dessen Eigentümer kann, soweit nicht Rechte Dritter entgegenstehen, nach § 903 BGB mit dem Medium nach Belieben verfahren und Dritte von jeder Einwirkung ausschließen. Gleiches gilt für die physikalische Materialisierung von Daten auf dem Speichermedium. Der Eigentümer kann somit grundsätzlich jeden Dritten von der Nutzung von Daten und Datenspeichern ausschließen, soweit nicht das Gesetz oder die Rechte Dritter entgegenstehen. Insofern stellt sich die Frage, ob Dritte ein dem Eigentumsrecht am Speichermedium (§ 903 BGB) entgegenstehendes Recht geltend machen können. In Betracht käme hier ein „Eigentumsrecht“ an den Daten. Eigentum zeichnet sich durch die rechtliche Zuweisung einer Sache zu einer Person im Sinne eines umfassenden Herrschaftsrechts aus (54, 36).⁶ „Dateneigentum“ muss also zumindest auch dieses Merkmal aufweisen.

³ BGH, NJW 1988, 406 (408); Stresemann, in: MüKo, § 90 Rn. 25; Jickeli/Stieper, in: Staudinger, § 90 Rn. 12.

⁴ Eser/Bosch, in: Schönke/Schröder, § 242 Rn. 9; Kühl, in: Lackner/Kühl, StGB, § 142 Rn. 2.

⁵ Papier, in: Maunz/Dürig, Art. 14 GG Rn. 197-200.

⁶ Klinck, in: Staudinger, Eckpfeiler des Zivilrechts, Abschn. W Rn. 1-4; Wendehorst, MüKo, Art. 43 EGBGB Rn. 41.

4.1.3 Zuordnung über §§ 90 ff. BGB

Fraglich ist, ob eine Zuordnung von Daten über § 90 BGB erfolgen kann. Das konstitutive Merkmal ist dabei die Sacheigenschaft. Eine direkte Anwendung des § 903 BGB kommt ausschließlich in Betracht, wenn körperliche Gegenstände im Sinne des § 90 BGB betroffen sind (38, 54).⁷ Daten sind als durch Zeichen vermittelte Informationen nicht eigenständig, sondern körperlich abhängig. Anders ausgedrückt kann nicht das Datum, sondern nur der Datenträger körperlich sein (36).⁸ Daten können also nicht als Sachen im Sinne des § 90 BGB qualifiziert werden. Weiterhin stellt sich die Frage, ob gespeicherte Daten eventuell als wesentliche Bestandteile im Sinne von § 93 BGB angesehen werden können. Nach § 93 BGB sind Bestandteile einer Sache, die voneinander nicht getrennt werden können, ohne dass der eine oder andere zerstört oder in seinem Wesen verändert wird (wesentliche Bestandteile), nicht Gegenstand besonderer Rechte. Insofern würde dann das Eigentum am Datenträger auch das Eigentum an den darauf bezogenen Daten umfassen. Dies erscheint aber insoweit problematisch, als dann „Dateneigentum“ und Sacheigentum nicht auseinander fallen könnten, was im Hinblick auf die tatsächlichen Verhältnisse kaum praktikabel ist.

4.1.4 Eigentum am Datenträger

Die Rechtsprechung geht davon aus, dass ein Datenträger mit den darauf verkörperten Daten eigentumsfähig sei. So hat das *Oberlandesgericht Karlsruhe*⁹ die Eigentumsfähigkeit von Daten im Rahmen von § 823 Abs. 1 BGB bejaht (18).¹⁰ Jedenfalls der Datenträger mit den darauf verkörperten Daten sei eine körperliche Sache (54, 40).¹¹ Zwar könne die Information, die in diesen Daten repräsentiert ist, wegen ihrer immateriellen Natur nicht Schutzgut einer Norm gegen die Beschädigung der materiellen Substanz sein. Es sei jedoch deswegen von einer Verletzung des Eigentums auszugehen, weil der Eigentümer durch den Eingriff gehindert werde, mit der Sache seinem Wunsch entsprechend zu verfahren.

⁷ Bassenge, in: Palandt, § 903 Rn. 2; Seiler, in: Staudinger, Vorbemerkung zu § 903 Rn. 3.

⁸ Stresemann, in: MüKo, § 90 Rn. 25.

⁹ OLG Karlsruhe, NJW 1996, 200.

¹⁰ Anderer Ansicht LG Konstanz, NJW 1996, 2662; siehe auch Gerstenberg, NJW 1956, 540.

¹¹ Jickeli/Stieper, in: Staudinger, § 90 Rn. 12; Redeker, NJW 2008, 2684 (2685).

4.1.5 Recht am eigenen Datenbestand

Zum Teil wird vertreten, dass sich aus § 823 Abs. 1 BGB als sonstiges Recht auch ein Recht am eigenen Datenbestand herleiten lasse (35).¹² Ein solches Recht benötigt für seine Legitimation eine gewisse eigenständige Bedeutung neben anderen Rechten. Werden Daten auf einem Datenträger unbrauchbar gemacht, so ist – sogar weitergehend als nach der Ansicht des *OLG Karlsruhe*¹³ – von einer Substanzverletzung an dem Datenträger auszugehen, da dieser nachhaltig physisch verändert wurde (35).¹⁴ Mit dem bloßen Rückgriff auf das Eigentum am Datenträger werden nämlich viele Probleme, die bei vernetzten Datenbeständen auftreten, nicht gelöst (35).¹⁵ Bedeutsam sind besonders die Fälle, in denen der wirtschaftliche Schaden gar nicht bei dem Eigentümer des Datenträgers entsteht, weil nicht er, sondern ein Anderer die Daten wirtschaftlich nutzt. Für solche Fälle hat ein Recht am eigenen Datenbestand durchaus eine Berechtigung. Allerdings zeichnen sich „sonstige Rechte“ im Sinne des § 823 BGB dadurch aus, dass sie eine den ausdrücklich normierten Rechten entsprechende Ausschlussfunktion besitzen (36).¹⁶ Es stellt sich deshalb die Frage, wer berechtigt ist, Dritte von der Nutzung eines Datenbestandes auszuschließen. Das Recht am eigenen Datenbestand ist daher eine Hilfskonstruktion, die zunächst mehr Probleme aufwirft als löst.

4.1.6 Datenschutzrecht

Möglicherweise bietet das Datenschutzrecht eine Lösung für die Zuordnungsproblematik. Es schützt aber „Daten“ nur dem Namen nach. Tatsächlich geht es um den dargestellten Inhalt, die Information, die sich mit einer Person in Beziehung bringen lässt. Das Datenschutzrecht betrifft nicht die Beziehungen einer Person zu einer Sache, sondern direkt die Beziehungen zwischen Personen und Daten. So können beispielsweise nach § 35 BDSG Löschungs- oder Berichtigungsansprüche gegen eine verarbeitende Stelle unabhängig davon geltend gemacht werden, in welcher Form die Daten vorhanden sind. Auch Auskunftsansprüche (z. B. aus § 34 BDSG) gewähren dem Betroffenen immer nur das Recht, Daten vorgelegt zu bekommen, die mit den gespeicherten Daten inhaltsgleich sind. Das umfasst zwar beispielsweise auch die Auskunft über die Bezeichnung

¹² Siehe dazu *BGH*, NJW 1996, 2924 (2925); ähnlich auch *Meyer/Wehlau*, NJW 1998, 1585 (1588).

¹³ *OLG Karlsruhe*, CR 1987, 19.

¹⁴ *OLG Oldenburg*, MDR 2012, 403; *Meyer/Wehlau*, NJW 1998, 1585 (1588).

¹⁵ *Meyer/Wehlau*, NJW 1998, 1585 (1588).

¹⁶ *Wagner*, in: *MüKo*, § 823 Rn. 143.

der Datei, in der die personenbezogenen Daten gespeichert sind (19).¹⁷ Der Betroffene erhält aber nicht das Recht auf Zugang zum Hauptspeichermedium. Das Datenschutzrecht schafft damit eine rechtliche Verantwortlichkeit für Daten. Es darf aber nicht dahingehend verstanden werden, dass der Betroffene Ausschließlichkeitsrechte an den einzelnen Datensätzen im Sinne eines „Eigen-tums“ besitzt.

4.1.7 Datenbank sui-generis

Daneben findet sich im Urheberrecht ein Hinweis auf eine mögliche Zuordnung von Rechten an Daten: Die §§ 87a ff. UrhG mit dem darin enthaltenen Leistungsschutzrecht für Datenbankhersteller. Das Urheberrecht schützt primär Inhalte. Mit den §§ 87a ff. UrhG wurde aber in Umsetzung der Richtlinie 96/9/EG ein viel diskutierter Schutz sui-generis geschaffen. Dass Datenbanken nicht gleichbedeutend mit den darin angeordneten Daten sind, versteht sich durch ihre Definition (48).¹⁸ Es stellt sich aber die Frage, ob der Gesetzgeber mit dieser Regelung zum Ausdruck gebracht hat, dass er Daten nur im Umfang dieses sui-generis-Schutzes für schützenswert hält. Der Schutz des Datenbankherstellers, der als Ausschließlichkeitsrecht und subjektiv-rechtliche Rechtsposition ausgestaltet ist,¹⁹ begründet aber nur den Schutz vor Vervielfältigung, Verbreitung und öffentlicher Wiedergabe, nicht jedoch ein eigentumsähnliches Vollrecht (23). Auch die Erschöpfungswirkung an körperlichen Vervielfältigungsstücken der Datenbank²⁰ zeigt, dass gerade nicht der Schutz von Daten bezweckt wird, sondern der des Dateninhalts (17, 13, 56). Dies zusammen genommen zeigt, dass mit der Regelung für Datenbanken ein spezieller Investitionsschutz geschaffen werden sollte, der vor wirtschaftlicher Ausbeutung fremder Leistung schützt, nicht aber eine rechtliche Zuordnung von Daten zu einer Person.

4.1.8 § 303a StGB: Daten als geschütztes Gut

Anders als im Zivilrecht, dem in Bezug auf die Frage nach dem Zuweisungsgehalt des Eigentums eine viel größere Rolle zukommt, gibt es im Strafrecht die Regelung des § 303a StGB, die sich explizit auf Daten als geschütztes Gut bezieht. Einiges spricht dafür, die Zuordnung der Daten im Rahmen des § 303a

¹⁷ HessVGH, RDV 1991, 187; Gola/Schomerus, § 34 Rn. 9.

¹⁸ Sendrowski, GRUR 2005, 369 (372).

¹⁹ Vgl. BGH, MMR 1999, 470 (472); Gaster, in: Hoeren/Sieber/Holznagel, Teil 7.6 Rn. 75 f.

²⁰ Gaster, S. 131 Rn. 522 ff.; Dreier, in: Schulze/Dreier, § 87b Rn. 18; Thum, in: Wandtke/Bullinger, § 87b Rn. 58.

StGB anhand des Skripturaktes, also anhand des technischen Herstellungsprozesses der Daten, vorzunehmen.

4.1.9 Fazit

Es erscheint geboten, im bestehenden Geflecht der zivilrechtlichen Normen Lösungen für die eigentumsrechtliche Zuordnung von Daten zu suchen. Zwar mag es auf den ersten Blick konstruiert anmuten, neben den ohnehin schon bestehenden Differenzierungen hinsichtlich Sacheigentum, Urheberrecht, Schutz des Datenbankherstellers, Datenschutz etc. zusätzlich eigenständiges Dateneigen-
tum anzunehmen. Dennoch sind Daten zu einem selbstverständlichen, fühlbaren Wert geworden. Die Entwicklung eines Schutzregimes anhand der aufgezeigten Ansätze ist deshalb notwendig. Hierzu wird auf die Langversion im Anhang der Studie verwiesen.

4.2 Urheberrecht

Der Umgang mit Big Data wirft eine Vielzahl urheberrechtlicher Fragen auf. Eine allgemeingültige und für jede Big Data-Lösung einheitliche Begutachtung urheberrechtlicher Fragen ist nicht möglich. Die konkreten Probleme ergeben sich stets aus der Ausgestaltung des Verfahrens im Einzelfall. Im folgenden Abschnitt werden die urheberrechtlichen Grundlagen sowie deren konkreter Bezug zum Umgang mit Big Data dargestellt. Hierbei wird auf ungeklärte Rechtsfragen und sich daraus ergebende Forschungsperspektiven hingewiesen.

4.2.1 Anwendbarkeit des UrhG

Auch bei Big Data-Lösungen stellt sich zunächst die Frage, ob das deutsche Urheberrechtsgesetz (UrhG) Anwendung findet. Der räumliche Anwendungsbe-
reich des UrhG ist durch das Territorialitätsprinzip auf das Gebiet der Bundes-
republik Deutschland beschränkt (13, 14, 56).²¹ Die weltweite Nutzung und
Verbreitung urheberrechtlich geschützter Werke schränkt die territoriale Wir-
kung des UrhG ein. Nach dem Schutzlandprinzip findet das Recht des Staates

²¹ BGH, Urt. v. 16.06.1994 - I ZR 24/92 = GRUR 1994, 798; Dreier, in: Dreier/Schulze,
Vor § 120 Rn. 1; Kotthoff, in: Dreyer/Kotthoff/Meckel, § 120 Rn. 4; v. Welser, in:
Wandtke/Bullinger, Vor § 120 ff. Rn. 5.

Anwendung, für dessen Gebiet Schutz beansprucht wird (14, 56).²² Der persönliche Anwendungsbereich des UrhG richtet sich nach den §§ 120–128 UrhG, die an die Staatsangehörigkeit des Betroffenen anknüpfen.

Diese Grundsätze sind bei der rechtlichen Begutachtung der verschiedenen Konstellationen beim Umgang mit Big Data zu beachten. Neben der Frage, ob sich das datenverarbeitende Unternehmen selbst auf das UrhG berufen kann, kann es auch auf die Frage ankommen, ob die verarbeiteten Daten fremder Personen dem nationalen urheberrechtlichen Werkschutz unterfallen. Grundsätzlich wird bei Big Data-Lösungen, die innerhalb Deutschlands genutzt werden, auch deutsches Urheberrecht zu beachten sein. Dennoch ergeben sich bei Internet-Sachverhalten, etwa im Bereich des Cloud-Computing, aufgrund der Ubiquität des Internets und der Tatsache, dass sich Handlungen nicht stets einem bestimmten Schutzterritorium zuordnen lassen, kollisionsrechtliche Fragestellungen, die noch nicht abschließend geklärt sind (37).²³

4.2.2 Urheberrecht an und Werkqualität von Daten

Das Urheberrecht schützt Werke der Literatur, Wissenschaft und Kunst, insbesondere in den in § 2 Abs. 1 UrhG genannten Formen. Es muss sich um persönliche geistige Schöpfungen handeln, die über eine ausreichende Gestaltungshöhe verfügen, § 2 Abs. 2 UrhG. Das Merkmal der Gestaltungshöhe erfordert dabei einen bestimmten Mindestgrad an Individualität des Werkes und dient folglich der Aussonderung einfacher Alltagserzeugnisse, die keinen Urheberrechtsschutz genießen (56).²⁴ Wann eine ausreichende Gestaltungshöhe erreicht ist, ist umstritten und differiert zwischen den einzelnen Werkarten (56).²⁵ Der Rechtsprechung folgend, darf ebenjener Grad an Gestaltungshöhe jedoch nicht zu gering ange setzt werden, so dass das Werk eine erhebliche individuelle Prägung besitzen muss (56).²⁶ Eine solche kommt einzelnen Daten regelmäßig nicht zu (13, 52).²⁷ Denn Daten und Fakten, wie sie z.B. in Datenbanken gespeichert werden, erfüllen nicht die Voraussetzungen des § 2 Abs. 2 UrhG, weil sie entweder nur eine grundsätzlich freizuhaltende Tatsache enthalten oder ihnen die erforderliche

²² BGH, Urt. v. 15. 2. 2007 - I ZR 114/04 = GRUR 2007, 871; *Kotthoff*, in: Dreyer/*Kotthoff*/Meckel, § 120 Rn. 8; v. *Welser*, in: Wandtke/Bullinger, Vor § § 120 ff. Rn. 4.

²³ *Nägele/Jacobs*, ZUM 2010, 281 (284 f.).

²⁴ Bullinger, in: Wandtke/Bullinger, § 2 Rn. 23.

²⁵ Bullinger, in: Wandtke/Bullinger, § 2 Rn. 24f.

²⁶ Bullinger, in: Wandtke/Bullinger, § 2 Rn. 24.

²⁷ Schulze, in: Dreier/Schulze, § 2 Rn. 130; Spindler/Hillegeist, in: Handbuch Forschungsdatenmanagement, S. 63 f.

Individualität fehlt (15, 13).²⁸ Entscheidend ist auch nicht, ob es sich um fremde oder eigene Daten handelt, mit denen im Rahmen des Big Data-Managements umgegangen wird. Urheberrechtsverstöße sind beim Umgang mit einzelnen Daten deshalb zunächst nicht zu befürchten.

Die Big Data-Lösung muss je nach ihrer konkreten Ausgestaltung auch fremde Urheberrechte beachten. Dies kann etwa bei der Auswertung nutzergenerierter Inhalte (UGC)²⁹ Dritter in sozialen Netzwerken, denen unter Umständen Schutz als Sprachwerk (§ 2 Abs. 1 Nr. 1 UrhG) zukommt, der Fall sein (50).³⁰ Gleichermaßen gilt für nach § 72 UrhG geschützte Lichtbilder oder Lichtbildwerke i. S. d. § 2 Abs. 1 Nr. 5 UrhG, die von Nutzern sozialer Netzwerke zugänglich gemacht werden. Die Auswertung solcher Werke wird dann regelmäßig das Vervielfältigungsrecht und das Recht der öffentlichen Zugänglichmachung berühren (50).³¹

Erfolgt im Rahmen der Big Data-Lösung ein Zugriff auf fremde Datenbestände, um deren Inhalte auszuwerten, hängt die Zulässigkeit des Vorgehens zunächst davon ab, ob die fremden Datenbestände urheberrechtlichem Schutz unterliegen. Datenbanken können gem. § 4 Abs. 2 UrhG als Datenbankwerke geschützt sein oder gem. §§ 87a ff. UrhG dem sui-generis-Schutz für Datensammlungen unterfallen.

Datenbankwerk

Ein Datenbankwerk nach § 4 Abs. 2 UrhG ist ein Sonderfall des Sammelwerks, dessen Elemente systematisch oder methodisch angeordnet und einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sind. Ein Schutz als Datenbankwerk kommt nur in Betracht, wenn die Auswahl oder Anordnung der enthaltenen Elemente auf einer schöpferischen Leistung beruht (13, 56, 15).³² In den meisten Fällen wird jedoch ein Zugriff auf Datenbestände erfolgen, deren Inhalt nach logischen Gesichtspunkten geordnet und auf Vollständigkeit ausgerichtet ist, so dass keine schöpferische Leistung vorliegt und der Schutz als Datenbankwerk nicht entgegensteht (52).³³

²⁸ Nordemann, in: Fromm/Nordemann, § 2 Rn. 76; Schulze, in: Dreier/Schulze § 20 Rn. 130.

²⁹ Siehe dazu auch Kapitel „II. L. 2.“.

³⁰ Solmecke/Wahlers, ZD 2012, 550 (553 f.).

³¹ Solmecke/Wahlers, ZD 2012, 550 (554).

³² Dreier, in: Dreier/Schulze, § 4 Rn. 1; Marquardt, in: Wandtke/Bullinger, § 4 Rn. 8; Czuchowski, in: Fromm/Nordemann, § 4 Rn. 13.

³³ Spindler/Hilleggeist, in: Handbuch Forschungsdatenmanagement, S. 64.

Sui-generis-Datenbankschutz

Wesentlich häufiger als der Schutz einer Datensammlung als Datenbankwerk wird der sui-generis-Schutz-für Datenbanken gem. den §§ 87a UrhG greifen, die Investitionsschutz für denjenigen, der zwar keine eigene schöpferische Leistung erbracht hat, jedoch das Investitionsrisiko für die Erstellung einer Datenbank trägt, gewähren (13, 56).³⁴ In Abgrenzung zum Datenbankwerk fordert der sui-generis-Datenbankschutz für die systematische und methodische Anordnung der Elemente also keine besondere Schöpfungshöhe, sondern eine wesentliche Investition (46, 13).³⁵ Die Frage, ob eine Investition als nach „Art oder Umfang wesentlich“ zu bezeichnen ist, muss durch eine Abwägung im Einzelfall und eine Gesamtbetrachtung aller Umstände ermittelt werden (15, 13).³⁶ Nach h. M. statuiert der unbestimmte Begriff der „Wesentlichkeit“ eine niedrige Schutzzschwelle, die nur ganz unbedeutende Investitionen ausschließen will (56, 13, 17).³⁷ Datensammlungen unterfallen daher regelmäßig dem sui-generis-Schutz.

Etwas anderes gilt jedoch dann, wenn die wesentlichen Kosten schon im Rahmen der Datenerhebung und nicht erst im Rahmen der Verwaltung der Datenbank anfallen (52).³⁸ Diese Investitionen sind nach der „British Horseracing Board“-Entscheidung³⁹ des EuGH nicht berücksichtigungsfähig. Wie die Abgrenzung zwischen Datenerhebungs- und -verwaltungsphase zu erfolgen hat, ist jedoch ungeklärt.

Der Schutzmfang des sui-generis-Schutzes bestimmt sich nach § 87b UrhG. Dem Rechteinhaber stehen das Vervielfältigungs- und Verbreitungsrecht sowie das Recht zur öffentlichen Wiedergabe zu. Satz 1 beschränkt die Rechte hierbei zunächst auf die Nutzung der gesamten Datenbank sowie eines wesentlichen Teils einer Datenbank. Ohne Zustimmung des Datenbankherstellers dürfen zur Datengewinnung also nur unwesentliche Teile übernommen werden, wobei die genaue Bestimmung der Wesentlichkeit wiederum vom Einzelfall abhängt.

³⁴ Dreier, in: Dreier/Schulze, Urheberrecht, § 4 Rn. 3; Thum, in: Wandtke/Bullinger, Vor § § 87a ff. Rn. 1.

³⁵ Vogel, in: Schricker/Loewenheim, § 87a Rn. 26; Dreier, in: Dreier/Schulze, Vor § § 87a ff. Rn. 1

³⁶ Czychowski, in: Fromm/Nordemann, § 87a Rn. 15; Dreier, in: Dreier/Schulze, § 87a Rn. 15.

³⁷ Thum, in: Wandtke/Bullinger, § 87a Rn. 55 ff; Dreier, in: Dreier/Schulze, § 87a Rn. 14; Gaster, Der Rechtsschutz von Datenbanken Rn. 476.

³⁸ Spindler/Hillegeist, in: Handbuch Forschungsdatenmanagement, S. 64.

³⁹ EuGH, Urt. v. 9. 11. 2004 - C-203/02 The British Horseracing Board Ltd u.a./William Hill Organization Ltd = GRUR 2005, 244.

Eigener Schutz der Big Data-Anwendung

Nicht nur die Quelle der genutzten Daten, sondern auch die Big Data-Anwendung selbst kann nach dem UrhG geschützt sein. Denkbar ist, dass die Anwendung als Datenbankwerk geschützt wird oder dem sui-generis-Schutz für Datenbanken unterfällt (27).⁴⁰ Naheliegend ist zudem der urheberrechtliche Schutz der Big Data-Lösung als Computerprogramm (§§ 69a ff. UrhG). Computerprogramme in diesem Sinne sind Programme in jeder Gestalt (§ 69a Abs. 1 UrhG). Sie werden geschützt, wenn sie das Ergebnis der eigenen geistigen Schöpfung ihres Urhebers sind, wobei es insbesondere nicht auf qualitative oder ästhetische Kriterien ankommt, § 69a Abs. 3 UrhG.

4.2.3 Betroffene Verwertungsrechte

Werden fremde Leistungen durch Urheber- oder Leistungsschutzrechte geschützt, stellt der Zugriff auf diese regelmäßig einen Eingriff in fremde Rechte dar. Welche Verwertungsrechte allerdings speziell betroffen sind, hängt maßgeblich von der konkreten Ausgestaltung der Big Data-Anwendung ab. Bei der Übernahme urheberrechtlich geschützter Texte kommt das Vervielfältigungsrecht des Rechteinhabers in Betracht. Daneben ist außerdem für das Big Data-Management das Recht der öffentlichen Zugänglichmachung (§ 19a UrhG) relevant.

Ein lang erwartetes Grundsatzurteil zum Verbreitungsrecht und dessen Er schöpfung beim Onlinevertrieb „gebrauchter“ Software fällte der EuGH in seiner „UsedSoft“-Entscheidung.⁴¹ Das Gericht urteilte zur umstrittenen Frage der Anwendbarkeit des urheberrechtlichen Erschöpfungsgrundsatzes auf den Onlinebezug von Computerprogrammen, dass sich das Recht auf ausschließliche Verbreitung der Programmkopien mit dem Erstverkauf erschöpfe. Diese Entscheidung eröffnet wichtige Fragen für das Big Data-Management. Die Forschung wird klären müssen, inwieweit sich das Urteil auf andere digitale Güter als Software erstrecken lässt (22, 25).⁴²

⁴⁰ Vgl. Beispiele bei *Kieun*, Database Technology for Large Scale Data, <http://www.cubrid.org/blog/web-2-0/database-technology-for-large-scale-data/> (zuletzt abgerufen am 5.7.2013).

⁴¹ EuGH, U. v. 3. 7. 2012, Akz. C-128/11 = MMR 2012, 586 m. Anm. *Heydn*.

⁴² Verneinend *LG Bielefeld*, GRUR-RR 2013, 281; vgl. auch *Hartmann*, GRUR-Int 2012, 980 (981 ff.); *Hoeren/Försterling*, MMR 2012, 642 (647).

4.2.4 Schranken des Urheberrechts

Der Eingriff in urheberrechtliche Verwertungs- oder Leistungsschutzrechte ist zulässig, wenn diese Maßnahmen durch entsprechende Schrankenregelungen ausdrücklich legitimiert sind. Regelmäßig wird das beim Umgang mit großen Datenmengen und dem damit verbundene Eingriff in fremde Verwertungsrechte nicht der Fall sein. Insbesondere erscheinen weder das Zitatrecht (§ 51 UrhG) noch das Recht Vervielfältigungen zum privaten und sonstigen eigenen Gebrauch anzufertigen (§ 53 UrhG) als einschlägige Schranken.

Es drängt sich damit eine weitere Forschungsperspektive auf. Muss das UrhG um eine weitere Schranke ergänzt werden und wie ist diese ggf. auszustalten, um den neuen Dimensionen des Datenverkehrs gerecht zu werden? Verschärft wurde diese Problematik durch ein aus dem Januar stammendes Urteil des *EGMR*⁴³ in Straßburg. Hiernach sei das geistige Eigentum durch die EMRK nicht uneingeschränkt geschützt, sondern stets im Einzelfall mit der Meinungs- und Informationsfreiheit der Medien und ihrer Nutzer nach Art. 10 EMRK abzuwegen. Dieses Urteil liefert wesentliche Neuerungen im Hinblick auf die Auslegung urheberrechtlicher Schranken, die Abwägung von Urheberrecht und Informationsinteresse und bringt daher die bisherige nationale Rechtsprechung ins Wanken.

4.2.5 Einräumung von Nutzungsrechten

Greifen keine legitimierenden Schranken, hängt die Zulässigkeit der urheberrechtlich relevanten Handlung von der Zustimmung des Rechtsinhabers ab. Dieser kann seine Zustimmung entweder durch Einwilligung oder vertraglich erteilen.

Vor allem die vertragliche Einräumung von Nutzungsrechten nach §§ 31 ff. UrhG eröffnet zahlreiche praktische Probleme: Die neuen Technologien des Internetzeitalters führen dazu, dass fast zwangsläufig in zahllose fremde Ausschließlichkeitsrechte eingegriffen wird,⁴⁴ so dass eine erhebliche Zahl vertraglicher Vereinbarungen getroffen werden müsste (28). Dies ist häufig bereits rein praktisch nicht möglich. Regelmäßig wird es mithin im Rahmen des Big Data-Managements Fälle geben, in denen weder legitimierende Schranken greifen noch eine vertragliche Einräumung von Nutzungsrechten stattgefunden hat.

⁴³ *EGMR*, Urt. v. 10.1.2013, Nr. 36769/08 - Ashby Donald/Frankreich.

⁴⁴ Klass, ZUM 2013, 1.

In vergleichbaren Konstellationen bediente sich der *BGH* der Konstruktion einer schlichten konkludenten Einwilligung, um zu interessensgerechten und praxisorientierten Lösungen zu gelangen. Hieraus ergibt sich unmittelbar die Frage, inwiefern sich die Grundsätze des *BGH* auch auf das Big Data-Management übertragen lassen. Von grundsätzlicherer Bedeutung ist allerdings das Problem, ob die Konstruktion der schlichten Einwilligung überhaupt haltbar ist. Große Teile des Schrifttums üben Kritik an der Rechtsprechung des *BGH* und äußern dogmatische Bedenken gegenüber der schlichten Einwilligung (11, 20, 26, 53).⁴⁵ Denkbare Lösungen für diesen (unbefriedigenden) Zustand könnten die Etablierung einer neuen oder die Erweiterungen einer bestehenden Schranke sein (28).⁴⁶ Als zielführend könnte sich die Einführung einer „Fair-Use-Klausel“ nach US-Vorbild erweisen, die in der Lage wäre, neuartige Entwicklungen flexibel zu erfassen und nicht der potenziellen Blockade des Gesetzgebers im Konflikt zwischen den verschiedenen Lobbygruppen unterliege (53).⁴⁷

4.3 Datenschutzrecht

4.3.1 Datenschutzrechtliche Grundsätze

Das Datenschutzrecht wird von verschiedenen Grundsätzen geprägt, deren Zweck der Schutz des Grundrechts auf informationelle Selbstbestimmung (Art. 2 Abs. 1, Art. 1 Abs. 1 GG) ist. Das Verständnis dieser Grundsätze ist daher auch Basis für die Lösung datenschutzrechtlicher Fragen im Zusammenhang mit Big Data. Neben den allgemeinen Rechtsgrundsätzen gelten im Datenschutzrecht spezielle Grundsätze wie *Zweckbindung*, *Transparenz*, *Direkterhebung* und *Datenvermeidung* bzw. *Datensparsamkeit*. Zudem steht die Datenverarbeitung nach dem BDSG unter einem *Verbot mit Erlaubnisvorbehalt*. Die speziellen datenschutzrechtlichen Grundsätze fanden in unterschiedlicher Form Eingang in das BDSG. Gem. § 3a BDSG hat jeder Datenumgang mit dem Ziel zu erfolgen, so wenige personenbezogene Daten wie möglich zu erheben und diese möglichst zu anonymisieren oder zu pseudonymisieren. Der Grundsatz der Direkterhebung findet sich in § 4 Abs. 2 S. 1 BDSG wieder, wonach personenbezogene Daten beim Betroffenen zu erheben sind. Der Zweckbindungsgrundsatz wiederum besagt, dass die Daten grundsätzlich nur zu dem Zweck verarbeitet und genutzt werden dürfen, zu dem sie ursprünglich erhoben wurden. Schließlich

⁴⁵ Vgl. etwa Conrad, ZUM 2010, 585; Götting, LMK 2010, 309481; Hüttner, WRP 2010, 1008 ff.; Spindler, GRUR 2010, 785.

⁴⁶ Klass, ZUM 2013, 1 (9f.) mit Verweis auf den Vorschlag von Schaefer, Urheberrechtliche Rahmenbedingungen für Bildersuchmaschinen de lege lata und de lege ferenda, S. 265 ff.

⁴⁷ Spindler, GRUR 2010, 785 (792).

gewährleistet der Transparenzgrundsatz jedem Betroffenen das Recht, Einsicht in gespeicherte personenbezogenen Daten zu nehmen, was einerseits durch Auskunftsrechte des Betroffenen und andererseits durch Informationspflichten der verarbeitenden Stelle sichergestellt wird. Nur in begrenzten Ausnahmefällen, etwa aufgrund eines überwiegenden Allgemeininteresses, kann die Zweckbindung entfallen (57).⁴⁸

4.3.2 Anwendbarkeit des BDSG - Kollisionsrechtliche Fragen bei grenzüberschreitendem Bezug

Besonders im Zusammenhang mit Big Data-Anwendungen auf ausländischen Servern ist noch völlig ungeklärt, wann ein ausländischer Anbieter überhaupt deutsches Datenschutzrecht zu beachten hat. Dabei ist fraglich, wie der Ort zu bestimmen ist, an dem die Erhebung der personenbezogenen Daten stattfindet. Aufgrund des Territorialitätsprinzips stellt sich im Zusammenhang mit Big Data-Anwendungen zudem grundsätzlich die Frage, inwieweit sich der Standort bzw. ein Standort der Datenmengen überhaupt verlässlich bestimmen lässt, da durch Cloud Computing Daten weltweit auf verschiedene Standorte verteilt sind,⁴⁹ die großen Datenmengen zudem flüchtig sind und sich der einmal ermittelte Standort binnen kurzer Zeit erneut ändern kann (37).

4.3.3 Datenverarbeitung – Verbot mit Erlaubnisvorbehalt

Das deutsche Datenschutzrecht ist geprägt durch das Volkszählungsurteil⁵⁰ des Bundesverfassungsgerichts, das dem Einzelnen das Recht zusichert, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen. Für den Umgang mit personenbezogenen Daten gilt ein sog. Verbot mit Erlaubnisvorbehalt, das in § 4 Abs. 1 BDSG kodifiziert ist und besagt, dass die Erhebung, Verarbeitung und Nutzung personenbezogener Daten nur zulässig ist, soweit dies gesetzlich gestattet ist oder der Betroffene eingewilligt hat. Für den Bereich der Telemedien gilt § 12 Abs. 1 TMG entsprechend.

Ein Datenumgang ist also legitimiert, wenn der Betroffene seine Einwilligung hierzu erteilt hat oder eine gesetzliche Ermächtigungsgrundlage vorliegt. Der Grundsatz des Verbots mit Erlaubnisvorbehalt wird zunehmend als untauglich

⁴⁸ Weichert, ZD 2013, 251 (256).

⁴⁹ Nägele/Jacobs, ZUM 2010, 281 (289 f.).

⁵⁰ BVerfG, NJW 1984, 419.

kritisiert, da heute potentiell jedermann zum Datenverarbeiter wird (44).⁵¹ Jede Äußerung und Lebensregung, die im Netz stattfindet, ist schon ein unmittelbar verarbeitetes Datum (43).⁵² Durch die „Ubiquität der Datenverarbeitung“⁵³ in Smartphones usw. findet Datenverarbeitung beinahe in jeder Lebensphase statt (43). Die Grundannahme des Gesetzes, nach der zunächst alle Verarbeitungen verboten sind, berücksichtigt nicht sämtliche Grundrechte der Beteiligten, sondern stellt den Schutz der informationellen Selbstbestimmung über alles weitere, nur um dann aber sehr viele und zum Teil sehr weite Erlaubnistatbestände vorzusehen (24,44).⁵⁴ Die Ausnahmen werden somit zur Regel (44).⁵⁵ Ein grundsätzlich generelles Verbot lässt unberücksichtigt, dass die moderne Informationsgesellschaft die Autonomie des Einzelnen erfordert, um „selbst seine Daten verarbeiten, verbreiten und ebenso auch die Daten anderer verarbeiten und verbreiten“⁵⁶ zu können (44). Im Rahmen weiterer Forschung zu Big Data-Anwendungen ist deshalb zu überprüfen, inwieweit sich bei einer effektsteigenden Modernisierung des Datenschutzes das Verbotsprinzip noch als tauglich erweist (44).⁵⁷

4.3.4 Datenqualität – Personenbezogene Daten

Sobald ein Umgang mit personenbezogenen Daten stattfindet, ist der Anwendungsbereich des BDSG eröffnet, § 1 Abs. 1 u. 2 BDSG (19, 49).⁵⁸ Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person, § 3 Abs. 1 BDSG. Sie entstehen insbesondere in sozialen Netzwerken, dem sog. Web 2.0 (57).⁵⁹ Das Bundesverfassungsgericht hat in seiner „Volkszählungsentscheidung“ festgestellt, dass „es unter den Bedingungen der automatisierten Datenverarbeitung kein ‚belangloses‘ Datum mehr“⁶⁰ gibt. Im Sinne eines effektiven

⁵¹ Schneider, AnwBI 2011, 233 (236).

⁵² Roßnagel/Pfitzmann/Garstka, Modernisierung des Datenschutzrechts, S. 22.

⁵³ Roßnagel/Pfitzmann/Garstka, Modernisierung des Datenschutzrechts, S. 22.

⁵⁴ Hoeren, ZD 2011, 145 (145); Schneider, AnwBI 2011, 233 (233).

⁵⁵ Schneider, AnwBI 2011, 233 (233).

⁵⁶ Schneider, AnwBI 2011, 233 (236).

⁵⁷ Schneider, AnwBI 2011, 233 (234).

⁵⁸ Gola/Schomerus, § 1 Rn. 22; Kühling/Seidel/Sivridis, S. 79; Sokol, in: Simitis, § 1 Rn. 57.

⁵⁹ Weichert, ZD 2013, 251 (252).

⁶⁰ BVerfG, NJW 1984, 419 (422).

Grundrechtsschutzes sind somit alle Daten, die eine Person betreffen, vom Begriff des personenbezogenen Datums erfasst (49, 29).⁶¹

Die Daten müssen den Bezug zu einer bestimmten oder bestimmbaren natürlichen Person zulassen. Vielfach erfolgt bei Big Data-Anwendungen ein Umgang mit anonymisierten Daten, so dass vermeintlich kein Personenbezug vorliegt. Dabei steigt mit der Menge an vorhandenen Daten auch die Wahrscheinlichkeit der Identifizierbarkeit einer bestimmten Person (8, 34, 57).⁶² Es wird daher weiterer Untersuchung bedürfen, ob Big Data als Ansammlung großer Datenmengen trotz der Anonymisierung von Daten die Re-Identifikation einzelner Personen ermöglicht (8, 34, 57).⁶³

4.3.5 Erlaubnisgründe

Die Tatsache, dass die Verarbeitung personenbezogener Daten unter einem Verbot mit Erlaubnisvorbehalt steht, führt dazu, dass den verschiedenen Erlaubnisgründen für die Zulässigkeit des Big Data-Managements eine erhebliche Bedeutung zukommt (57).⁶⁴

Einwilligung

Besondere Schwierigkeiten ergeben sich bei Big Data im Zusammenhang mit der Einwilligung gem. § 4a BDSG, die in der Praxis den am häufigsten gewählten Erlaubnisgrund darstellt (34).⁶⁵ Der Betroffene muss in die Lage versetzt werden, im Vorfeld „Anlass, Ziel und Folgen der Verarbeitung korrekt abzuschätzen“ (19, 39).⁶⁶ Wichtig ist, dass dem Einwilligenden die Folgen der Erteilung stets klar vor Augen geführt werden – ist dies nicht oder nicht ordnungsgemäß geschehen, folgt daraus die Unwirksamkeit der Einwilligung (39, 49).⁶⁷ Im Bereich der Telemmediendienste kann die Einwilligung auch elektronisch erklärt werden, § 13 Abs. 2 TMG. Sie ist aber zu protokollieren und dem Nutzer jederzeit zum Abruf bereitzustellen.

⁶¹ Dammann, in: Simitis, § 3 Rn. 7; Kühling/Seidel/Sivridis, S. 79; Sokol, in: Simitis, § 1 Rn. 57.

⁶² Baeriswyl, digma 2013, 14 (15); Mayer-Schönberger/Cukier (2013), p.155; Weichert, ZD 2013, 251 (257).

⁶³ Baeriswyl, digma 2013, 14 (15); Mayer-Schönberger/Cukier (2013), p. 154; Weichert, ZD 2013, 251 (258).

⁶⁴ Weichert, ZD 2013, 251 (255).

⁶⁵ Mayer-Schönberger/Cukier (2013), p. 153.

⁶⁶ Gola/Schomerus, § 4a Rn. 25; Plath, in: Plath, § 4a Rn. 31.

⁶⁷ Plath, in: Plath, § 4a Rn. 82; Sokol, in: Simitis, § 4 Rn. 7.

Im Rahmen von Big Data-Anwendungen ist genau zu prüfen, in welchen Fällen sich ein Rückgriff auf gesetzliche Erlaubnistratbestände anbietet und wann eine Einwilligung einzuholen ist. Kennzeichnend für Big Data ist, dass bereits vorhandene Daten auf neue Weise analysiert bzw. die vorhandenen Daten mit weiteren verfügbaren Daten zusammengeführt werden, um neue Erkenntnisse zu erhalten (57).⁶⁸ Dies ist derzeit immer erst nach Einholung einer Einwilligung möglich, was die Entwicklung von Big Data deutlich hemmt.

Datenschutzrechtliche Einwilligungen als eine Art „Handelsgut“ erlangen bei Big Data-Anwendungen große Bedeutung. Unentgeltliche Online- und Offline-Dienste sind keinesfalls kostenlos, sondern werden vielmehr im Rahmen eines Tauschgeschäfts (Einwilligung gegen Dienstleistung) abgewickelt (10, 42).⁶⁹ Die Einwilligung gilt mittlerweile als „Kommerzialisierungsinstrument“⁷⁰, das den spezifischen wirtschaftlichen Wert personenbezogener Daten einerseits widerspiegelt und andererseits auch abschöpft,⁷¹ denn in Zeiten von Big Data sind Daten an sich bereits wertvoll (10, 42, 34).⁷² Zukünftige Forschung wird sich der Frage annehmen müssen, inwieweit ein solches Tauschgeschäft mit der Idee der Freiwilligkeit der Einwilligung vereinbar ist und ob und ggf. wie die Freiwilligkeit im Zuge eines immer sorgloseren Umgangs der Nutzer mit Einwilligungen sichergestellt werden kann. Eine Forschungsperspektive bei Big Data-Anwendungen ergibt sich aus der Frage, welche gesetzgeberischen Konsequenzen sich aus der veränderten Natur der Einwilligung und wie sich die notwendige Transparenz für die Betroffenen herstellen lässt (44).⁷³

Es wird zu erforschen sein, welche Formen der Einwilligung für Big Data-Anwendungen zulässig sind. Die strenge Schriftform des BDSG war ein deutsches Kurirosum und einmalig in Europa (12, 42).⁷⁴ Vorgeschlagen wird auch die Begrenzung der Wirkungsdauer einer Einwilligung (42).⁷⁵ Klärungsbedürftig ist deshalb, inwieweit hierdurch einerseits eine verbesserte Praktikabilität im Umgang mit der Einwilligung erreicht und andererseits effektiver Grundrechtsschutz sichergestellt werden kann.

⁶⁸ Weichert, ZD 2013, 251 (256).

⁶⁹ Buchner, DuD 2010, 39 (39); Rogosch, S. 41.

⁷⁰ Buchner, DuD 2010, 39 (39).

⁷¹ Rogosch, S. 41.

⁷² Mayer-Schönberger/Cukier (2013), p. 100.

⁷³ Schneider, AnwBl 2011, 233 (239).

⁷⁴ Drewes/Siegert, RDV 2006, 139, (144); Rogosch, S. 196.

⁷⁵ Rogosch, S. 192.

Erlaubnisnormen für öffentliche und nicht-öffentliche Stellen (§§ 12 ff BDSG und §§ 28 ff BDSG)

Die im BDSG enthaltenen Rechtsgrundlagen der Datenverarbeitung differenzieren zwischen der Datenverarbeitung öffentlicher Stellen einerseits (§§ 12 ff. BDSG) und der Datenverarbeitung nicht-öffentlicher Stellen und öffentlich-rechtlicher Wettbewerbsunternehmen andererseits (§§ 27 ff. BDSG).

Öffentliche Stellen

Öffentliche Stellen verarbeiten Daten grundsätzlich aufgabenbezogen und unterliegen dabei einem strengen Zweckbindungsgrundsatz (57).⁷⁶ Insoweit kommt eine Big Data-Auswertung durch öffentliche Stelle nur aufgrund gesetzlicher Spezialregelungen in Betracht (57).⁷⁷

Nicht-öffentliche Stellen

Grundsätzlich ist gem. § 28 Abs. 1 BDSG das Erheben, Speichern, Verändern oder Übermitteln personenbezogener Daten oder ihre Nutzung in drei verschiedenen Alternativen zulässig. Als Mittel zur Erfüllung eigener Geschäftszwecke, zur Wahrung berechtigter Interessen der verantwortlichen Stelle ohne entgegenstehendes, überwiegendes schutzwürdiges Interesse des Betroffenen oder schließlich dann, wenn ein Umgang mit allgemein zugänglichen Daten erfolgt. Für Big Data-Anwendungen ist insbesondere die Kategorie der allgemein zugänglichen Daten gem. § 28 Abs. 1 Nr. 3 BDSG von erhöhtem Forschungsinteresse, denn die Art und Weise der Verwendung dieser Daten sowie deren Umfang und Zweck kann frei bestimmt werden (57).⁷⁸ Allgemein zugängliche Daten liegen vor, wenn sie „sich sowohl ihrer Zielsetzung als auch ihrer Publikationsform nach dazu eignen, einem individuell nicht bestimmbarer Personenkreis Informationen zu vermitteln (19).“⁷⁹ Grundsätzlich zählen hierzu auch Angaben auf Internetseiten (19, 39, 57).⁸⁰ Für die Forschung besonders relevant ist dabei die Frage, wann Daten als „öffentlicht“ einzuordnen sind.

Anonymisierung, Pseudonymisierung & De-Personalisierung

Das BDSG steht dem Big Data-Management auch dann nicht im Wege, wenn die verarbeiteten Daten ihren Personenbezug verlieren (57).⁸¹ Anonymisierung und Pseudonymisierung von Daten stellen Möglichkeiten dar, die Auswertung von Big Data zu legitimieren. Anonymisieren ist das Verändern

⁷⁶ Weichert, ZD 2013, 251 (254).

⁷⁷ Weichert, ZD 2013, 251 (254).

⁷⁸ Weichert, ZD 2013, 251 (255).

⁷⁹ Gola/Schomerus, § 28 Rn. 151.

⁸⁰ Gola/Schomerus, § 28 Rn. 151; Plath, in: Plath, § 28 Rn. 76; Weichert, ZD 2013, 251 (257).

⁸¹ Weichert, ZD 2013, 251 (258).

personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können, § 3 Abs. 6 BDSG. Die identifizierenden Daten müssen gelöscht oder zumindest derart verändert werden, dass eine (auch zufällige) Zuordnung nicht mehr möglich ist (57).⁸² Bei Telemedien ist gem. § 13 Abs. 6 TMG eine anonymisierte oder pseudonymisierte Nutzung zu ermöglichen, soweit dies technisch möglich und zumutbar ist.

Forschungsperspektiven eröffnen sich in diesem Bereich insbesondere im Zusammenspiel mit der Informatik im Hinblick auf die Entwicklung von Big Data-Verfahren, die eine Re-Identifikation sicher ausschließen. Für Big Data-Anwendungen werden insbesondere neue Möglichkeiten der De-Personalisierung durch Technik zu prüfen sein. In Betracht kommt hier die Verschlüsselung von Daten, die normalerweise dem Bereich der Datensicherheit zugeordnet wird. Hier würde es möglicherweise genügen, entsprechende Verschlüsselungstechniken (Encryption) zu nutzen. Entscheidend ist, dass die Verschlüsselung nach dem aktuellen Stand der Technik als sicher eingestuft werden kann und die Gefahr der Decodierung durch unbefugte Dritte nicht besteht (55).⁸³ Das BDSG ist somit grundsätzlich bei der Übermittlung von verschlüsselten Daten zu beachten, es sei denn der Datenexporteur weist nach, dass die Daten sicher verschlüsselt sind und dass nur der Betroffene oder ein nach dem BDSG Berechtigter die Daten mit dem Schlüssel einsehen kann (51).⁸⁴

4.3.6 Auftragsdatenverarbeitung und Cloud Computing

Die Auftragsdatenverarbeitung hat als „zentraler Anwendungsfall bei der Nutzung neuer Formen der Datenverarbeitung“⁸⁵ auch für das Big Data-Management eine immense Bedeutung (7). So stellt etwa das Cloud Computing einen typischen Fall der Auftragsdatenverarbeitung dar. Diese hat in § 11 BDSG eine detaillierte, allerdings als nicht mehr zeitgemäße Ausgestaltung erfahren (7, 39).⁸⁶ Erfolgt die Erhebung, Verarbeitung oder Nutzung personenbezogener Daten im Auftrag durch andere Stellen (Auftragnehmer), ist der Auftraggeber für

⁸² Weichert, ZD 2013, 251 (258).

⁸³ Schultze-Melling, in: Taeger/Gabel, § 9 Rn. 83.

⁸⁴ Spies, MMR-Aktuell 2011, 313727.

⁸⁵ AG Rechtsrahmen des Cloud Computing, *Trusted Cloud - Datenschutzrechtliche Lösungen für Cloud Computing*, S. 4.

⁸⁶ AG Rechtsrahmen des Cloud Computing, *Trusted Cloud - Datenschutzrechtliche Lösungen für Cloud Computing*, S. 4; Plath, in: Plath, § 11 Rn. 48.

die Einhaltung der Vorschriften des BDSG und anderer Vorschriften über den Datenschutz verantwortlich.

Im Zusammenhang mit Big Data stellen sich daher besondere Herausforderungen. Die Herrschaft des Kunden über die beim Auftragsdatenverarbeiter befindlichen Daten muss durch die technische und rechtliche Ausgestaltung der Datenverarbeitung gewährleistet werden (37).⁸⁷ Die Ausgestaltung klarer und den Vorgaben der §§ 9, 11 BDSG entsprechender vertraglicher Vereinbarungen gestaltet sich dabei schwierig (21, 27).⁸⁸ Darüber hinaus sind Art und Umfang der gesetzlichen Kontrollpflichten weitgehend ungeklärt.

Die angedeuteten Probleme, die sich im Rahmen des § 11 BDSG im Hinblick auf das Cloud Computing ergeben, wurden insbesondere von der „AG Rechtsrahmen des Cloud Computing“ detailliert herausgearbeitet. Darüber hinaus wurden Lösungsansätze für eine zeitgemäße Ausgestaltung des § 11 BDSG entwickelt, unter anderem zum Schriftlichkeitserfordernis des § 11 Abs. 2 S. 2 BDSG und dem Kontrollerfordernis nach § 11 Abs. 2 S. 4 BDSG. Hierauf wird in der Langversion des Gutachtens ausführlich eingegangen.

4.4 Vertragsrecht

Die Vertragsgestaltung im Big Data-Business ist aufgrund der vielen technischen Fragen sowie der Berührungspunkte zum Datenschutz- und Urheberrecht kompliziert. Die Vielzahl denkbarer Big Data-Anwendungen und ihrer vertraglichen Umsetzungsmöglichkeiten führt dazu, dass sich pauschalisierte Aussagen zur Vertragsgestaltung schwerlich treffen lassen. Die Verteilung der Rechte und Pflichten aus dem jeweiligen Vertragsverhältnis liegt nach dem Grundsatz der Vertragsfreiheit in den Händen der beteiligten Parteien. Die rechtliche Einordnung eines Vertrages und damit einhergehend vertragsrechtliche Fragen müssen stets im Einzelfall beurteilt werden.

Vertragliche Regelungen können etwa bei im Internet weitgehend frei zugänglichen Daten (z.B. in sozialen Netzwerken) eine Rolle spielen. Denn bei der Nutzung fremder Daten bzw. fremder Inhalte eines Webseitenbetreibers können vertragliche Nutzungsbeschränkungen die rechtliche Zulässigkeit oder den Umfang einer zulässigen Nutzung einschränken (60).⁸⁹

⁸⁷ Nägele/Jacobs, ZUM 2010, 281 (290).

⁸⁸ Grünwald/Döpkens, MMR 2011, 287 (287); Nägele/Jacobs, ZUM 2010, 281, (284).

⁸⁹ Ausführlich dazu Zieger/Smirra, MMR 2013, 418 (421).

Fragen der Berechtigung sowie weitere grundlegende Probleme der Vertragsgestaltung im Geschäftsbereich von Big Data-Unternehmen lassen sich beispielhaft auch anhand des Cloud-Computing veranschaulichen, bei dem der Cloud-Nutzer zur Verwaltung seiner Datenbestände auf Ressourcen des Cloud-Anbieters zurückgreift. Vertragsrechtliche Fragen im Zusammenhang mit dem Cloud-Computing, einem Hauptanwendungsfall des Big Data-Managements, werden in der Folge exemplarisch betrachtet.

4.4.1 Hauptleistungspflichten

Geht man von der mietrechtlichen Beurteilung des Cloud-Computing Vertrages aus,⁹⁰ so schuldet der Cloud-Provider Überlassung und Erhaltung der Miet-sache (59, 54).⁹¹ Konkrete Anhaltspunkte, was das für die Hauptleistungspflichten bedeutet, ergeben sich daraus jedoch noch nicht. Deshalb müssen Verträge technisch genau abgestimmt sein. Quantitative und qualitative Vereinbarungen erfolgen im Rahmen von Service Level Agreements (SLA)⁹², in denen unter anderem die Fragen des Sicherheitsmanagements und des Daten-Backup geregelt werden (47).

Rückgriff auf das gesetzliche Leitbild

Was aber passiert, wenn eine dieser technischen Feinheiten nicht vertraglich geregelt wurde, ist unklar. Ob die vertraglichen Lücken durch Rückgriff auf das gesetzliche Leitbild des Mietrechts sachgerecht geschlossen werden können, erscheint fraglich. Dies zeigt sich beispielsweise bei der Frage der Verfügbarkeit von Übertragungskapazitäten. Es bedarf der Klärung, ob die Überlassung und Erhaltung der Mietsache ausnahmslos „rund um die Uhr“ geschuldet wird. Daneben ist unklar, wen das Risiko trifft, dass die Vertragsdurchführung durch technische Fehler gestört wird.

AGB-rechtliche Regelung

Um gar nicht erst Gefahr zu laufen, möglicherweise existenzgefährdenden Regressansprüchen ausgesetzt zu sein, regeln in der Praxis viele Unternehmen Haftungsausschlüsse in allgemeinen Geschäftsbedingungen. Diese sind im Gegensatz zu den privatautonom ausgehandelten Leistungspflichten, in die der Richter nicht korrigierend eingreifen darf, der vollen Inhaltskontrolle nach § 307 BGB unterworfen (9).⁹³ Ein großes Problem besteht dabei hinsichtlich der Frage, wie

⁹⁰ Dazu Wicker, MMR 2012, 783 (788).

⁹¹ Emmerich, in: Staudinger, § 535 Rn. 21.

⁹² Schumacher, MMR 2006, 12 (13).

⁹³ Spindler, in: Benlian/Hess/Buxmann, S. 34.

eine wirksame Klausel konkret ausgestaltet sein muss, um den strengen Anforderungen der Rechtsprechung gerecht zu werden. Langwierige und unternehmensschädliche Rechtsstreitigkeiten könnten durch die Erarbeitung rechtswirksamer Haftungsklauseln reduziert werden, so dass sich in diesem Bereich für das Big Data-Management eine lohnende Forschungsperspektive eröffnet.

Entgelt

Der Cloud-Nutzer schuldet die mietvertraglich vorgesehene Vergütung, typischerweise monatlich. Die Miete erhöht sich, wenn der Nutzer unberechtigterweise zu einer Übernutzung kommt. Zu bedenken ist, dass der *BGH* für Vermietlösungen CPU-Erhöhungsklauseln als wirksam angesehen hat.⁹⁴ Insofern kann der Cloud-Provider bei Veränderungen der technischen Infrastruktur auf Nutzerseite ein erhöhtes Nutzungsentgelt verlangen. Im Gegensatz dazu besteht auch die Möglichkeit der Mietminderung bei eingeschränkter Nutzbarkeit. Denkbar ist jedoch die Vereinbarung von Wartungsklauseln mit klar definierten Reaktionszeiten. Diese dürfen jedoch nicht so definiert werden, dass daraus de facto ein Gewährleistungsausschluss resultiert oder die Hauptleistungspflicht unterlaufen wird (9).⁹⁵

4.4.2 Nachvertragliche Pflichten

Aus Sicht des Cloud-Nutzers können nachvertragliche Pflichten eine entscheidende Bedeutung haben. Wurden diesbezüglich keine vertraglichen Regelungen getroffen, stellt sich die Frage, wie der Nutzer sicherstellen kann, dass er einerseits alle Daten zurückhält und andererseits keine Daten beim bisherigen Vertragspartner zurückbleiben. Hilfe könnte in dieser Situation § 539 Abs. 2 BGB bieten. Kommt es zur Beendigung des Vertrages, gewährt diese Vorschrift dem Mieter ein Wegnahmerecht. Bei konsequenter Anwendung der mietrechtlichen Vorschriften müsste dem Cloud-Nutzer bezüglich seiner eigenen Daten ebenfalls ein solches Recht zustehen. Inwiefern hinsichtlich dieser Regelungen klauselmäßige Modifikationen möglich sind und ob im Falle eines wirtschaftlichen Schadens Deliktsrecht anwendbar ist, bedarf einer genaueren Untersuchung. Gerade im Bereich der Vertragsabwicklung bestehen somit relevante Forschungsperspektiven.

⁹⁴ *BGH*, MMR 2003, 393.

⁹⁵ Spindler, in: Benlian/Hess/Buxmann, S. 36.

4.4.3 Schadensermittlung bei Datenverlust

Kommt es zum schuldhaften Datenverlust zum Nachteil eines Big Data-Unternehmens, stehen diesem unterschiedliche Schadensersatzansprüche zu. Diese können sich nicht nur aus dem Vertrag wegen Verletzung einer Haupt- oder Nebenleistungspflicht ergeben, sondern unter Umständen auch aus dem Deliktsrecht oder Spezialgesetzen wie § 7 BDSG. Allerdings stellt sich die Frage, wie im konkreten Fall die Höhe des Schadens zu ermitteln ist. Das *OLG Oldenburg* hat zur Schadensberechnung die Kosten zugrunde gelegt, die für die Wiederherstellung des Datensatzes erforderlich waren.⁹⁶ Denn im Rahmen des Schadensersatzes ist in erster Linie Naturalrestitution geschuldet. Möglich ist dies aber nur, soweit die Dateien aufgrund einer in anderer Form noch vorhandenen Vorlage technisch reproduzierbar sind.⁹⁷ Andernfalls träfe den Schädiger die Werterstattpflicht aus § 251 Abs. 1 BGB. Für diesen Fall hat der *BGH* festgestellt, dass die Vermögenseinbuße vom Tärichter nach freiem Ermessen zu schätzen ist.⁹⁸ Dementsprechend droht im Big Data-Business eine hohe Haftung für schuldhaften Datenverlust. Vor diesem Hintergrund wird die Notwendigkeit einer Haftungsbegrenzung erneut deutlich.

4.5 Informationshaftung und Informationsqualität

Die aus den Datenmengen abgeleiteten Informationen sind Grundlage für Entscheidungen in verschiedensten Lebensbereichen. Fehlerhafte Informationen können gravierende Auswirkungen haben. So lösten beispielsweise fehlerhafte Daten von Geschwindigkeitssonden eine Fehlerkette aus, die letztlich zum Absturz eines Flugzeugs führte, bei dem 228 Insassen ihr Leben verloren.⁹⁹ Angesichts der immer stärkeren Berücksichtigung von Daten in Entscheidungsprozessen sowie in Anbetracht der daraus entstehenden Gefahren stellt sich die Frage nach einem Haftungsregime für fehlerhafte Informationen.

4.5.1 Fehlinformation und Informationsqualität

Die Haftung eines Big Data-Unternehmens kommt nur in Betracht, wenn auf dessen Seite ein Fehlverhalten vorliegt. Ein solches könnte durch die

⁹⁶ *OLG Oldenburg*, B. v. 24.11.2011 - 2 U 98/11; MMR-Aktuell 2011, 326091.

⁹⁷ *BGH*, MMR 2009, 250 (251).

⁹⁸ *BGH*, MMR 2009, 250.

⁹⁹ *Lehnartz*, Welt vom 05.07.2012, abrufbar unter <http://www.welt.de/vermischt/weltgeschehen/article107913815/Wie-das-Airbus-Drama-in-elf-Kilometern-Hoehe-ablief.html>.

Übermittlung einer fehlerhaften Information gegeben sein. Es stellt sich allerdings die Frage, wann eine Fehlinformation vorliegt. Hierbei ist es unklar, ob es qualitative Mindeststandards für Informationen gibt und welche das sind. Betrachtet man nationale und internationale Rechtsquellen, so ergibt sich, dass die Normgeber bei informationsrelevanten Tätigkeiten stets die Einhaltung eines qualitativen Mindestlevels vorschreiben oder zumindest von dessen Einhaltung ausgehen. Neben der Richtigkeit verweist etwa Art. 4 lit. d) der EG-Datenschutzverordnung¹⁰⁰ auf die Vollständigkeit von personenbezogenen Daten. Konsens besteht somit jedenfalls dahingehend, dass Daten bzw. die daraus abgeleiteten Informationen korrekt, d.h. nachweislich richtig sein müssen.

Besonders intensiv hat sich der Gesetzgeber in den USA mit dem Thema Informationsqualität befasst. Dort wurden ein „Federal Data Quality Act“ sowie diesen konkretisierende Richtlinien erlassen.¹⁰¹ Diese beziehen sich zwar ausschließlich auf die Informationsverbreitung durch Behörden. Nichtsdestotrotz können die darin aufgestellten Kriterien womöglich zum Teil auf die Frage der Bewertung der Qualität großer Datenmengen übertragen werden. Ausdrücklich werden in diesen Richtlinien neben der Nützlichkeit die Kriterien der Objektivität, Nachvollziehbarkeit und Klarheit genannt (16).¹⁰² Inwieweit diese Maßstäbe auch für die Bewertung der Datenqualität im Rahmen des Big Data-Managements fruchtbar gemacht werden können, bedarf der Klärung.

Festzuhalten ist, dass die Frage, wie die Qualität einer Information allgemeinverbindlich beurteilt werden kann, bislang offen ist. Ob eine Information oder ein Datum „gut“ sind und somit eine Haftung mangels Fehlinformation ausgeschlossen ist, lässt sich abstrakt allerdings wohl nur mit dem Maßstab der sachlichen Richtigkeit bestimmen.

4.5.2 Vertragliche Haftung

Lässt sich im Rahmen einer vertraglichen Beziehung nach den oben dargestellten Grundsätzen die Übermittlung einer fehlerhaften Information zwischen einem Big Data-Unternehmen und seinem Kunden feststellen, so kommt ein Schadensersatzanspruch wegen Pflichtverletzung des Schuldverhältnisses gem. § 280 Abs. 1 BGB in Betracht. Dabei kann zum einen die Verletzung einer Hauptleistungspflicht, zum anderen die Verletzung einer Nebenleistungspflicht aus § 241 Abs. 2 BGB vorliegen.

¹⁰⁰ VO (EG) 45/2001.

¹⁰¹ Vgl. <http://www.whitehouse.gov/sites/default/files/omb/fedreg/reproducible2.pdf>.

¹⁰² Gasser, S. 11f.

Während es im Big Data-Bereich bisher keine höchstrichterlichen Entscheidungen bezüglich einer vertraglichen Haftung für fehlerhafte Informationen gibt, zeigen sich in verwandten Geschäftsfeldern vergleichbare Probleme. Das bereits angesprochene Kriterium der Vollständigkeit war Hauptstreitpunkt zwischen einem Anlageberater und einem Kunden, der wegen fehlerhafter Aufklärung und somit Verletzung einer Hauptleistungspflicht Schadensersatz verlangte.¹⁰³ Ähnliche Schwierigkeiten könnten auch im Big Data-Geschäftsbereich drohen.

Auch im Rahmen von vertraglichen Nebenpflichten können fehlerhafte Informationen für eine Haftung relevant werden. So entschied der *BGH*, dass ein Makler, der im Exposé eine objektiv falsche Angabe macht, seine Sorgfaltspflichten gegenüber dem Kunden verletzt.¹⁰⁴ Gleiches gilt für einen Krankenhausmitarbeiter, der sich dazu bereit erklärt, einem Patienten bei der Erstellung seines Testaments zu helfen.¹⁰⁵ Eine fehlerhafte Information sorgte in diesem Fall für die Unwirksamkeit des Testaments. Das *OLG Hamm* entschied, dass der Krankenhausräger dem nicht wirksam zum Erben eingesetzten Kläger dessen Schaden ersetzen muss. Ob diese Ergebnisse auch auf das Big Data-Management übertragbar sind, muss in Zukunft diskutiert werden.

4.5.3 Deliktische Haftung

Führt eine fehlerhafte Information zur Verletzung eines Rechtsgutes im Sinne des § 823 Abs. 1 BGB, kommt für den Geschädigten unabhängig von einer vertraglichen Beziehung zum Schädiger ein deliktischer Schadensatzanspruch in Betracht. In diesem Zusammenhang stellte der *BGH* verschiedene Grundsätze auf, die für die Informationshaftung von Bedeutung sein könnten.

Generalisiert man die Maßstäbe, die der *BGH* im Fall „Kochsalzlösung“¹⁰⁶ angewendet hat, so ergibt sich, dass ein Anbieter, dessen Produkt aufgrund fehlerhafter Informationen Schäden verursacht, grundsätzlich nicht deliktisch haftet. Ausnahmsweise kommt jedoch eine Haftung wegen Verletzung der Verkehrssicherungspflichten in Betracht, wenn keine Maßnahmen getroffen wurden, die vor Fehlern schützen. Dies könnte insbesondere der Fall sein, wenn sich die Informationen möglicherweise auf hochrangige Rechtsgüter, z.B. Leib und Leben, auswirken. Angesichts des großen Zeitaufwandes und der drohenden Kosten

¹⁰³ *BGH*, WM 2013, 836.

¹⁰⁴ *BGH*, NJW 2000, 3642.

¹⁰⁵ *OLG Hamm*, Urteil vom 13.04.2010 – I-21 U 94/09.

¹⁰⁶ *BGH*, NJW 1970, 1963.

sowie der Tatsache, dass Verkehrssicherungspflichten nach Ansicht des *BGH* die Ausnahme sein sollen, könnte Vieles dafür sprechen, eine generelle Verkehrssicherungspflicht für Big Data-Unternehmen abzulehnen (41).¹⁰⁷

Daneben könnten allerdings auch die Grundsätze der Produzentenhaftung gem. § 823 Abs. 1 BGB zu einer Haftung des Schädigers führen. Hierbei handelt es sich um eine Verschuldenshaftung mit Beweiserleichterungen für den Kläger (31).¹⁰⁸ Der Grundgedanke dieser Rechtsprechung besteht ebenfalls in der Annahme einer Verkehrssicherungspflicht. Weil die Produzentenhaftung ursprünglich für Industrieprodukte entwickelt wurde, ist jedoch unklar, ob ihre Grundsätze auch im Big Data-Bereich Anwendung finden könne (33).¹⁰⁹

Zusammenfassung:

- Die Frage nach der rechtlichen Zuordnung von Daten bzw. dem Eigentum an Daten bedarf einer eindeutigen Klärung durch den Gesetzgeber. Daten haben sich zu einem wichtigen Handelsgut entwickelt, dessen Bedeutung im Alltag nicht länger hinter derjenigen beweglicher Sachen zurücksteht.
- Die Auswertung großer Datenmengen ist unter urheberrechtlichen Gesichtspunkten möglich. Dennoch herrscht hinsichtlich verschiedener Fragen Rechtsunsicherheit. Ein Beispiel hierfür ist das Schutzlandprinzip, das sich bei der Bestimmung des anwendbaren Urheberrechts in Bezug auf Internetsachverhalte als wenig praktikabel erweist.
- Eine Modernisierung des Datenschutzrechts ist im Hinblick auf einen effektiven Datenschutz erforderlich. Es ist unter anderem zu überprüfen, inwieweit sich bei einer Modernisierung des Datenschutzrechts das Verbotsprinzip (Verbot mit Erlaubnisvorbehalt) noch als tauglich erweist. Darüber hinaus bedarf die gesetzliche Regelung für die Auftragsdatenverarbeitung (§ 11 BDSG), deren zentraler Anwendungsfall das Cloud-Computing ist, einer Anpassung an den aktuellen Stand der Technik.
- Das Big Data-Management eröffnet eine Vielzahl ungelöster vertragsrechtlicher und haftungsrechtlicher Fragen. Es müssen verbindliche Maßstäbe zur Bestimmung der Qualität einer Information und zur Frage, wann eine vertragliche oder deliktische Haftung für Fehlinformationen greifen kann, festgelegt werden.

¹⁰⁷ Ähnlich: *Reese*, DStR 1994, 1121 (1122f.).

¹⁰⁸ *Lehmann*, NJW 1992, 1721 (1721).

¹⁰⁹ *Meyer*, ZUM 1997, 26 (30).

5 Der Blick in die Zukunft: 10 Kernthesen

Im Folgenden werden die bisher erarbeiteten Erkenntnisse mit Hilfe von 10 Thesen reflektiert. Diese Thesen wurden in Gesprächen mit Unternehmensvertretern von Technologieherstellern aber auch Anwendern als pointierte Aussagen zum Innovationspotenzial von Big Data entwickelt. Daraus ergibt sich ein Stimmungsbild, welches zur Ableitung von vier Zukunftsszenarios genutzt wird. Diese Zukunftsszenarios beziehen sich, im Gegensatz zu anderen Studien, spezifisch auf den deutschen Markt.

5.1 Die 10 Kernthesen

Im Folgenden werden die 10 Kernthesen vorgestellt. Diese wurden aus den vorgehenden Kapiteln abgeleitet. Dabei wurde insbesondere darauf geachtet, von technischen Details zu abstrahieren und auf die Implikationen und vermuteten Wirkungen von Big Data einzugehen.

5.1.1 Big Data allein macht keinen Unterschied.

Es kommt auf die Einbettung in die Geschäftsprozesse an. Dieser These liegt die Annahme zu Grunde, dass Big Data vor allem dann zu einem echten Innovationstreiber wird, wenn dadurch neue Geschäftsprozesse ermöglicht werden. Die Möglichkeit, riesige Datenmengen schnell zu verarbeiten macht alleine noch keinen Unterschied. Selbst wenn Analysen schnell vorliegen, müssen sie in Entscheidungen umgesetzt werden. So wäre es beispielsweise denkbar, täglich einen „Jahres“-abschluss durchzuführen. Wenn diese Informationen aber (noch) nicht für unternehmerische Entscheidungen genutzt werden (können), entsteht kein Mehrwert. Selbst wenn bisher menschliche Entscheidungen automatisiert durchgeführt werden, müssen die geeigneten Geschäftsprozesse zu entsprechenden Wertschöpfung etabliert werden. Es wird daher postuliert, dass nur in Kombination mit einem Umbau bzw. einer Neugestaltung der Geschäftsprozesse Big Data wertschöpfend eingesetzt werden kann. Wichtig ist, dass die Einbettung von Big Data in die Geschäftsprozesse auch mit einem

Verständnis der Analyseproblematik von Big Data einhergeht. Dafür sind Kompetenzen in Data Science, sowohl in skalierbaren Datenanalysesystemen als auch in Datenanalyseverfahren, nötig, um beispielsweise die korrekten Analysemodelle (chaotisch vs. gaussartig) effizient auf den dafür geeigneten Systemen im Hinblick auf Datenvolumen, Datenrate und Heterogenität zu realisieren und korrekt zu interpretieren.

5.1.2 Big Data wird menschliche Entscheidungsprozesse automatisieren.

Dieser These liegt zu Grunde, dass Big Data als das Versprechen aufzufassen sein könnte, bessere Entscheidungen generieren zu können. So werden bereits Risikokalkulationen für Konsumentenversicherungen bzw. -kredite weitgehend automatisch durchgeführt. Auf Basis des kalkulierten Risikos wird automatisch eine Vergabeentscheidung getroffen. Big Data verspricht nun in solche Entscheidungsmodelle zusätzlich die Analyse unstrukturierter Daten zu integrieren, um noch effektivere Entscheidungen treffen zu können. Es gibt eine Vielzahl von Entscheidungen in vielen Branchen, die ähnliche Entscheidungsprofile aufweisen (z.B. Ausfall von Zulieferern, Lieferantentreue, Qualitätsmaße). Dem steht entgegen, dass der Wettbewerbsvorteil durch die Automatisierung solcher Entscheidungen nur von kurzfristiger Dauer ist, da Konkurrenten hier schnell nachziehen können. Alternativ kann Big Data lediglich als Entscheidungsgrundlage dienen. Das Wertschöpfungspotenzial bezieht sich dann darauf, Entscheidungsträger bestmöglich zu unterstützen, d.h. die richtigen Daten in der erforderlichen Qualität in der richtigen Menge zum richtigen Zeitpunkt an den richtigen Ort zu liefern. Hierbei ist zu beachten, dass die Datenanalysten hinreichende Kenntnisse in statistischen Verfahren und anderen Methoden der Datenanalyse mitbringen.

5.1.3 Big Data wird das Bauchgefühl als wesentliche Entscheidungsgrundlage nicht ablösen.

Dieser These liegt zu Grunde, dass trotz Analysen auf Basis von Big Data die menschliche Intuition und Erfahrungen wesentliche Ausgangspunkte für Entscheidungen bleiben. Dies trifft vor allem auf strategische Entscheidungen zu, bei denen hohe Unsicherheit bzw. sogar Unkenntnis über Handlungsalternativen bzw. deren Implikationen herrscht. Dem gegenüber stehen die technischen Möglichkeiten durch eine schnelle Analyse großer Datenmengen entsprechende Simulationen und Sensitivitätsanalysen zu ermöglichen. Hierdurch können Entscheidungsträger Handlungsalternativen explorieren und deren Implikationen besser abschätzen. Ebenso kann es hier zu entsprechenden Anforderungen seitens der Unternehmensaufsicht, den Kunden, bzw. staatlichen Organen

kommen, die für Entscheidungen von großer Tragweite entsprechen Analysen aller zur Verfügung stehenden Daten notwendig machen.

5.1.4 Big Data wird zu Fehlentscheidungen von erheblicher Tragweite führen.

Dieser These liegt die Annahme zu Grunde, dass automatisierte Entscheidungsprozesse in unvorhergesehenen Situationen bzw. auf Basis fehlerhafter Daten zu Entscheidungen führen können, die weitreichende wirtschaftliche und gesellschaftliche Folgen haben könnten. So führten beispielsweise bereits inadäquate Annahmen im automatisierten Wertpapierhandel zu erheblichen Verlusten. Darüber hinaus zeigen psychologische Studien, dass Menschen einer Vielzahl von unterbewussten Entscheidungsverzerrungen unterliegen. So weisen Menschen beispielsweise kleinen Mustern in großen Datenmengen tendenziell mehr Bedeutung zu, als es objektiv zulässig wäre. Demgegenüber steht, dass die effektive Kombination aus Datenanalyse, Visualisierungstechniken und Entscheidungssituation helfen kann, Entscheidungsträgern solche Entscheidungsverzerrungen bewusst zu machen.

5.1.5 Big Data schafft völlig neue Geschäftsmodelle und Dienstleistungen.

Dieser These liegt die Annahme zu Grunde, dass durch die gezielte Kombination unterschiedlicher Daten und leistungsfähiger Analysewerkzeuge neue Leistungsangebote entwickelt werden können. Dies betrifft neben dem Endkundenmarkt vor allem die zwischenbetriebliche Wertschöpfung. So ermöglicht die Integration unterschiedlichster Sensordaten die Vorhersage von Maschinenausfällen und eine rechtzeitige Wartung. Für die Kunden solcher „predictive maintenance“-Dienstleistungen ist besonders interessant, dass entsprechende Datenanalysen dem Anbieter ermöglichen, seine Erkenntnisse zu Ausfallgründen bzw. Best Practices mit dem gesamten Kundensegment zu teilen. Darüber hinaus ermöglicht Big Data die Entwicklung spezialisierter Datendienstleistungen. Forschungsprojekte und erste Unternehmen bieten maßgeschneiderte Einblicke und Analysen in ihre Daten und können so insbesondere kleinen und mittelständischen Unternehmen Analysen ermöglichen, die eigenständig nicht durchzuführen wären. Dieser Sicht steht entgegen, dass Big Data aktuell vor allem dazu genutzt wird, bestehende Geschäftsmodelle und Dienstleistungen kostengünstiger bzw. mit höheren Qualitätszusagen zu betreiben. Hier sind zum Beispiel die Einsatzgebiete in der Logistik zu nennen, bei denen Big Data „lediglich“ den handelnden Akteuren geringere Puffer bzw. genauere Liefervoraussagen ermöglicht. Dabei könnten durch Informationsmarktplätze völlig neue Anwendungen und Geschäftsmodelle entstehen.

5.1.6 Big Data wird in naher Zukunft eine kritische Infrastruktur darstellen.

Dieser These liegt die Annahme zu Grunde, dass die schnelle Analyse großer Datens Mengen in vielen Branchen zu einer wesentlichen Grundlage des unternehmerischen Handelns wird. Nachdem der Einsatz von Informationssystemen und die damit verbundene Automatisierung und Digitalisierung bereits in nahezu allen Branchen unverzichtbar ist, werden in Zukunft auch die Qualität, Menge, Aktualität und Zusammensetzung der darin verarbeiteten Daten über den Markterfolg entscheiden. Darüber hinaus kann Big Data aber auch für staatliche Organe ein unverzichtbarer Baustein des Handelns werden. Würden beispielsweise wichtige Wetterinformationen ausfallen bzw. fehlerhaft sein, so würden auf unternehmerischer und gesellschaftlicher Ebene systematisch Fehlentscheidungen getroffen werden. Darüber hinaus besteht die Möglichkeit, dass sich monopolistische Strukturen etablieren, innerhalb derer bestimmte Unternehmen eine marktbeherrschende Stellung durch die Sammlung und Aufbereitung bestimmter Daten erreichen. Dem gegenüber steht einerseits der schnelle technologische Fortschritt, der bestehende Wettbewerbsvorteile schnell erodieren lässt.

5.1.7 Big Data wird zu neuem Verständnis & Ausdifferenzierung hinzu Datenschutz, Urheberrecht und Haftungsrecht führen.

Diese These beruht auf der Beobachtung, dass die rechtliche Einschätzung zu Big Data vor allem im Kontext von Skandalen im Umgang mit personenbezogenen Daten, beispielsweise in sozialen Netzwerken, diskutiert wird. Allerdings zeigt die eigene Analyse, dass hier die aktuelle Rechtsprechung bereits eine differenzierte Analyse von rechtlichen Fragestellungen ermöglicht. Die These hat daher zum Ziel die oftmals diffuse Unsicherheit von Anwendern und Entscheidern hinsichtlich der rechtlichen Situation darzustellen. Diese Unsicherheit wird durch die globale Ausrichtung bestehender Big Data-Angebote verschärft. Allerdings kann eine effektive Adressierung dieser diffusen Unsicherheit ein erheblicher Wettbewerbsvorteil für einen Standort sein.

5.1.8 Big Data verschärft die Suche nach der Nadel im Heuhaufen.

Diese These beruht auf der Annahme, dass der Einsatz von Big Data das bereits bestehende Überangebot an Informationsquellen lediglich verschärft. Insbesondere die Möglichkeit, explorative Datenanalyse durchführen zu können, führt dazu, dass ein erheblicher Aufwand in die Aufbereitung und Analyse von Datenbeständen investiert wird, ohne dass damit nennenswerte wirtschaftliche Effekte erreicht werden können. Es fehlen häufig Kriterien, mit denen der Wert der durchgeführten Analysen beurteilt werden kann. So ermöglicht Big Data

beispielsweise die Abkehr von stichprobenartigen Analysen und erlaubt stattdessen die vollständige Analyse von Unternehmensdaten für die Identifikation von dolosen Handlungen. Ob die gefundenen Auffälligkeiten dann aber tatsächlich dolose Handlungen darstellen, ist wiederum Aufgabe einer weiteren Beurteilung. Erzeugt eine Analyse zu viele Auffälligkeiten, die sich letztendlich nicht als Betrugsfall (engl. Fraud) herausstellen, wird die zusätzliche Analyse eher zu Belastung.

5.1.9 Big Data verschärft die Suche von Unternehmen nach geeigneten Mitarbeitern.

Diese These basiert auf der Annahme, dass die effektive Nutzung der Innovationspotenziale von Big Data eine spezifische Kombination von informationstechnisch-mathematischen und betriebswirtschaftlichen Kompetenzen benötigt. Die betriebswirtschaftlichen Kompetenzen sind notwendig, um jene Bereiche im Unternehmen zu identifizieren, in denen die verfügbaren Datensätze den größten Nutzen stiften können. Ebenso sind detaillierte Kenntnisse zu Unternehmensstrategien, Geschäftsmodellen und Geschäftsprozessen notwendig, um die richtigen Fragen zu stellen und die Ergebnisse in Handlungsempfehlungen umsetzen zu können. Informationstechnisch-mathematische Kompetenzen sind notwendig um verlässliche und nachvollziehbare Analysen entwickeln und durchführen zu können. Insbesondere der Umgang mit unstrukturierten Daten und Daten, die keiner Normalverteilung unterliegen, erfordert spezifische Statistikkenntnisse um die Belastbarkeit aber auch die Grenzen der Analysen beurteilen zu können.

5.1.10 Big Data ist insbesondere eine Chance für den Mittelstand.

Diese These beruht auf der Vermutung, dass die umfangreichen technischen Voraussetzungen für Big Data-Analysen sowie die notwendige Vielfalt an Kompetenzen zur Ableitung verlässlicher Analysen die Potenziale von Big Data eher für große Unternehmen erreichbar machen. Dem gegenüber steht, dass gerade der Mittelstand eine interessante Zielgruppe für Big Data-basierte Dienstleistungen, beispielsweise in Form von Informationsmarktplätzen, ist. Insbesondere für Mit telständler, die international tätig sind, könnten solche Dienstleistungen einen erheblichen Wettbewerbsvorteil darstellen.

5.2 Ein Stimmungsbild zu den 10 Thesen

Die zehn Thesen wurden von 185 Experten und Entscheidern im Umfeld von Big Data vorgelegt, mit der Bitte Stellung zu beziehen. Die zehn Kernthesen wurden

mit drei zusätzlichen Fragen ergänzt, um das Potential von Big Data besser abschätzen zu können gerade im Hinblick zu KMU. Es handelt sich um die Aussagen: „Big Data wird für Unternehmen messbar Wert schaffen“, „Big Data wird in den nächsten 2-3 Jahren zu einem entscheidenden Faktor im globalen Wettbewerb“ und „Big Data ist lediglich ein Modewort, das zu Marketingzwecken eingesetzt wird“. Experten aller befragten Gruppen stimmen darin überein, dass Big Data als Paradigma allein keinen Unterschied macht (vgl. Abbildung 63). Vielmehr stimmen zwischen 41 Prozent der befragten Experten voll und ganz zu, dass Big Data in existierende Geschäftsprozesse integriert werden muss. An dieser Stelle stimmen aber zwischen 29 Prozent der Experten voll und ganz zu, dass Big Data neue Geschäftsmodelle und Dienstleistungen ermöglichen kann. Am wenigsten glauben Experten über alle Gruppen, dass Big Data menschliche Entscheidungsprozesse automatisieren, zu Fehlentscheidungen von erheblicher Reichweite führen und lediglich ein Modewort sein wird. Vielmehr glauben zwischen 47 Prozent der Experten, dass Big Data nicht bzw. auf keinen Fall die Suche nach der Nadel im Heuhaufen verstärken wird bzw. dass Big Data umgekehrt mit überzeugenden Konzepten daherkommt.

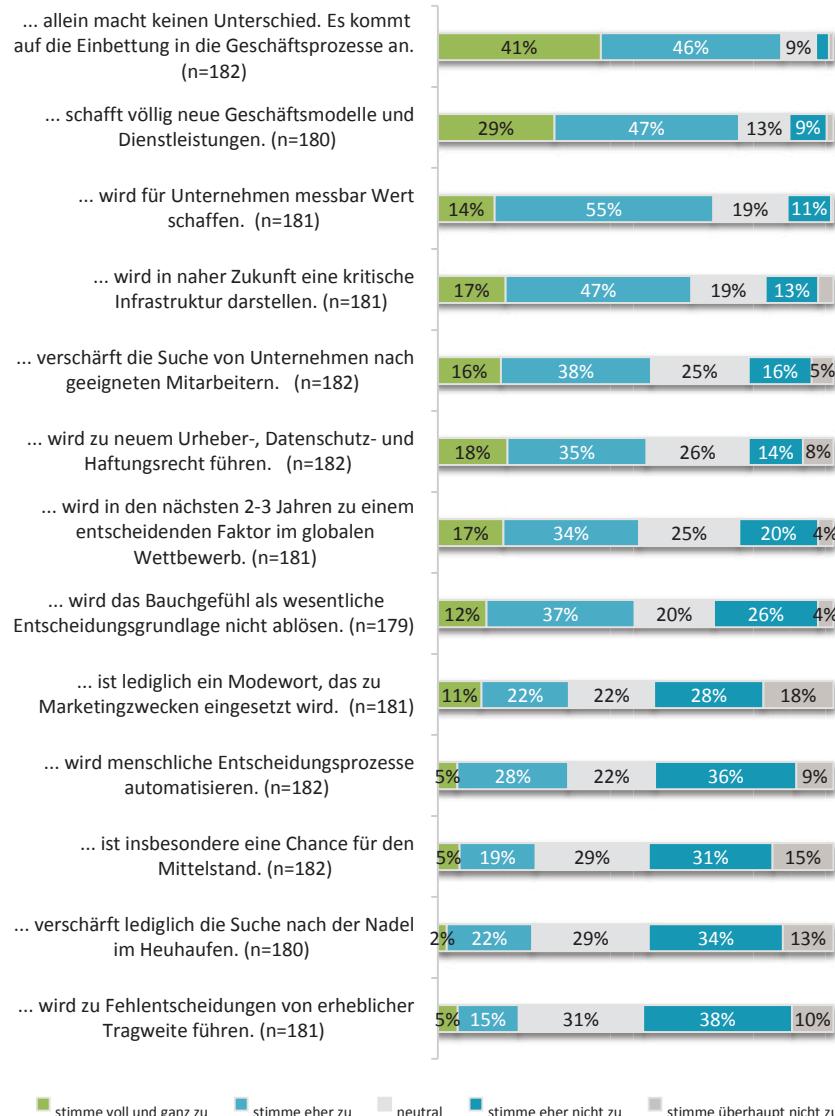


Abbildung 63: Allgemeine Einschätzungen zu Big Data (Quelle: Eigene Darstellung)

Wie zu erkennen ist, gibt es eine klare Überzeugung der Teilnehmer, dass Big Data zu neuen Geschäftsmodellen und Dienstleistungen führen kann (These 7). Damit verbunden ist die Überzeugung, dass Big Data nur dann wirtschaftlich sinnvoll genutzt werden kann, wenn entsprechende Analysen in die Geschäftsprozesse eingebettet werden können (These 1). Ebenso wird deutlich, dass Big Data als besonders kritisch für die zukünftige Wettbewerbsfähigkeit angesehen wird (These 6). Bemerkenswert ist zudem, dass die Entscheidungsautomatisierung (These 2) sowie der bestehende Wert von Intuition bei der Entscheidungsfindung (These 3) gleichgewichtig behandelt werden. So sind einige der Teilnehmer der Meinung, dass Entscheidungen künftig vermehrt automatisiert werden. Gleichzeitig messen sie dem „Bauchgefühl“ weiterhin eine Bedeutung

zu, welche bei allen Gruppen höher als die von These 2 ist. Dies lässt vermuten, dass hier zwischen strategischen und operativen Entscheidungen unterschieden wird. Während für letztere durchaus das Potenzial zur Automatisierung gesehen wird, werden dem Mensch die strategischen Entscheidungen zugesprochen. Dies deckt sich mit dem Stimmungsbild zu These 4. Die Teilnehmer gehen mehrheitlich nicht davon aus, dass aufgrund von Big Data Fehlentscheidungen von erheblicher Tragweite getroffen werden. Hinsichtlich der Einschätzung zu den Implikationen von Big Data für Urheberrecht, Datenschutz und Haftungsrecht (These 7) erwarten die Teilnehmer mehrheitlich Änderungen in der Gesetzeslage. Nachfragen lassen darauf schließen, dass sich die Teilnehmer dabei insbesondere auf Fragen des Datenschutzes beziehen.

Die Teilnehmer stimmen überwiegend nicht zu, dass es hinsichtlich der Gefahr der Datenflut bzw. der berüchtigten Suche nach der Nadel im Heuhaufen (These 8) schwierig wird. Die Experten der Teilnehmer geht davon aus, dass Big Data diese Suche noch verschärfen wird, während die andere Hälfte dem nicht zustimmt. Hier wird unter anderem deutlich, dass offensichtlich viele Erwartungen und Vorstellungen unter dem Begriff Big Data zusammengefasst werden. Eine Differenzierung anhand von Werttreibern von Big Data könnte hier zu einer Schärfung der Erwartungen beitragen. Wiederum mehrheitlich einig sind sich die Teilnehmer in Bezug auf die Verschärfung der Suche nach entsprechenden Kompetenzen bzw. Mitarbeitern (These 9). Einige Teilnehmer stimmen der These 10 zu, dass Big Data insbesondere eine Chance für den Mittelstand darstellt. Hier wird aber vor allem die neutrale Position gewählt. Es werden dem Mittelstand also durchaus Chancen eingeräumt, allerdings werden die nicht so klar gesehen. Insgesamt zeigt dieses Stimmungsbild, dass die wesentlichen Innovationspotenziale von Big Data in der Entwicklung neuer Geschäftsmodelle und Dienstleistungen liegen und dafür eine enge Verzahnung von Big Data mit den Geschäftsprozessen die Voraussetzung ist. Ebenso antizipieren die befragten Experten schon jetzt den Bedarf an entsprechenden Ausbildungs- und Fortbildungskonzepten.

5.3 Zukunftsszenarios

Um einen Kontext für die Empfehlungen der Studie zu entwickeln, werden im Folgenden vier Zukunftsszenarios beschrieben. In jedem Szenario wird dabei eine Voraussetzung beschrieben, die für die volle Ausnutzung des Innovationspotenzials von Big Data in Deutschland als entscheidend angesehen wird.

5.3.1 Datenschutz und Datensicherheit als Wettbewerbsvorteil

Wir schreiben das Jahr 2018. Big Data ist zu einem unverzichtbaren Bestandteil des gesellschaftlichen und wirtschaftlichen Handelns geworden. Im Weltmarkt sind dabei insbesondere deutsche Unternehmen in der Entwicklung und Nutzung von Big Data führend. Sie haben sich bereits früh mit den rechtlichen Rahmenbedingungen auseinandergesetzt und Methoden und Werkzeuge entwickelt, die es erlauben die Vorteile von Big Data zu nutzen ohne das Vertrauen der Kunden bzw. die rechtlichen Rahmenbedingungen zu verletzen. Hilfreich waren insbesondere Entwicklungen, die einen zweckbestimmten Umgang mit den Nutzerdaten ermöglichten. Darauf basierende Lösungen und Dienstleistungen werden aufgrund ihrer Transparenz und Vertrauenswürdigkeit weltweit nachgefragt. Darüber hinaus haben deutsche Unternehmen die Voraussetzungen entwickelt auch sensitive, nicht personenbezogene Daten im Sinne ihrer Kunden zu verarbeiten. Entsprechende innovative Dienstleistungen sind insbesondere im industriellen Umfeld anzutreffen. Deutschland hat hier seine Spitzenvorstellung im internationalen Wettbewerb behaupten und ausbauen können.

Die hohen Anforderungen an den Datenschutz in Deutschland wurden bisher teilweise als Hemmnis für die volle Ausnutzung des Innovationspotenzials von Big Data angesehen. Aktuelle Ereignisse zeigen, dass klare Regeln zu Datenschutz und Datensicherheit für einen transparenten und nachvollziehbaren Umgang mit Daten sorgen. Können die oben beschriebenen Hindernisse im deutschen und europäischen Rechtsraum behoben werden, kann sich der hohe Datenschutz als Wettbewerbsvorteil für deutsche Unternehmen etablieren. Da nicht nur deutsche Nutzer und Endkunden einen transparenten und rechtssicheren Umgang mit Ihren Daten bevorzugen wird durch diesen Wettbewerbsvorteil ein Grundstein für die internationale Markterschließung gelegt. Wenn in Deutschland die technischen Grundlagen für effektive Analysen etabliert sind und diese mit den Grundsätzen der Zweckbindung, der Transparenz, der Direkterhebung und der Datensparsamkeit vereinbar sind, könnte sich die damit verbundene Verlässlichkeit und Vertrauenswürdigkeit deutscher Big Data-Unternehmen positiv auf eine internationale Marktstellung auswirken.

Kann dieses Vertrauen und die Verlässlichkeit zudem mit einer hohen Datensicherheit kombiniert werden, entstehen weitere nennenswerte Vorteile für deutsche Unternehmen. Daten können in Cloud-basierten Lösungen gehalten und verarbeitet werden, was die Integration und schnelle Verarbeitung der Daten fördert, ohne das Vertrauen der Kunden und Nutzer zu gefährden. Dies ermöglicht einerseits die Ausnutzung erheblicher Kostenpotenziale, andererseits aber auch die Entwicklung von Unternehmensnetzwerken, die ihre Daten miteinander teilen um damit neue Geschäftsmodelle und Dienstleistungen zu erschaffen.

Datenschutz und Datensicherheit sind demnach Elemente, die durchgängig bei der Entwicklung von Big Data-basierten Systemen Anwendung und Berücksichtigung finden müssen. Beispielsweise müssen die Grundsätze des „Privacy by Design“ für Big Data-Systeme angepasst und deren Anwendbarkeit sichergestellt werden. Ebenso sind verschiedene technische Komponenten zu entwickeln, die einen flexiblen und agilen Einsatz von Big Data ermöglichen.

5.3.2 Informationsmarktplätze als Wettbewerbsvorteil des Mittelstands

Wir schreiben das Jahr 2018. Der deutsche Mittelstand floriert. Ein wesentlicher Treiber dafür ist die Tatsache, dass sich in Deutschland eine Vielzahl von Informationsmarktplätzen etabliert hat, die insbesondere kleinen und mittelständischen Unternehmen Zugang zu den Potenzialen von Big Data verschaffen. Diese Informationsmarktplätze übernehmen die technischen und organisatorischen Rahmenbedingungen für das Sammeln und Analysieren von Daten. Viele kleine und mittelständische Unternehmen, die entlang der Wertschöpfungskette arbeiten, teilen sich den Aufwand des Sammelns und der Analyse und können so gemeinsam von den Marktplätzen profitieren. Der Marktplatz stellt die Qualität, Nachvollziehbarkeit und Effizienz der Analysen sicher, während die beteiligten Unternehmen ihre Daten zur Verfügung stellen. Dies hat deutschen Unternehmen ermöglicht, schnell auf die wirtschaftlichen und politischen Umbrüche der vergangenen Jahre zu reagieren. Big Data wurde genutzt, um durch effizientere Produktionstechniken aber vor allem durch innovative Produkte und Dienstleistungen in vielen Branchen international eine Vorreiterstellung einzunehmen bzw. zu behaupten. Dies resultierte darin, dass auch konkurrierende Unternehmen gemeinsam Informationsmarktplätze nutzen um zumindest wettbewerbsunkritische Analysen miteinander zu teilen. Als Folge wurde die Reaktionsschnelligkeit und Innovationskraft des deutschen Mittelstandes erheblich ausgebaut.

Die hohen technischen, organisatorischen und personellen Aufwände für die Durchführung und Bereitstellung von verlässlichen und nachvollziehbaren Analysen von Big Data sind aktuell noch ein Hemmnis für den Einsatz in kleinen und mittelständischen Unternehmen.

5.3.3 Die antizipierende Fabrik

Wir schreiben das Jahr 2020. Die deutsche Industrie erlebt seit einigen Jahren einen deutlichen Aufschwung. Dies basiert insbesondere auf einer effektiven Nutzung der riesigen Datenmengen, die aufgrund von Sensoren in allen Teilen der Fertigung und des Betriebs eines Unternehmens anfallen. Insbesondere konnten deutsche Industrieunternehmen die Potenziale von Big Data dazu nutzen, ergänzende

Dienstleistungen zu ihren Produkten zu entwickeln und zu betreiben. Dies hat ihnen einen erheblichen Funktionalitäts- und Qualitätsvorsprung eingebracht und so die Aufrechterhaltung ihrer Wettbewerbsvorteile ermöglicht. Big Data wird dabei vor allem zur Antizipation von Engpässen und Schwachstellen im Produktionsbetrieb genutzt, aber auch um ungeplante Wartungen weitgehend zu vermeiden. So wurden Vorhersagemodelle mit sehr hoher Güte erarbeitet, die den Ausfall von Produktionsressourcen stark reduziert haben. Darüber hinaus bildet die Analyse von Kundenmeinungen in Online-Medien mittlerweile eine feste Säule der Produktentwicklung und -wartung. Das schnelle Erkennen von kritischen Kundenmeinungen weltweit hat deutschen Unternehmen geholfen, teure Rückrufaktionen zu vermeiden bzw. die damit verbundenen Reputationsschäden durch eine proaktive Kommunikation mit den Kunden stark zu reduzieren. Darüber hinaus sind entsprechende Analysen fester Bestandteil der Produktentwicklung wodurch sich der Markterfolg deutscher Produkte deutlich gesteigert hat.

Im industriellen Umfeld findet Big Data ideale Anwendungsfelder. Hier anfallende Daten sind in der Regel nicht personenbezogen sondern beschreiben den Zustand von Maschinen, Produktionsabläufen und Warenströmen. Werden personenbezogene Daten integriert so können leicht Methoden der Anonymisierung verwendet werden, da im Allgemeinen kein Rückschluss auf einzelne Personen oder sogar Personengruppen notwendig ist. Big Data kann so einen höchstmöglichen Überblick über den aktuellen Produktionsstand sowie präzise Vorhersagen zu möglichen Zuständen in der Zukunft ermöglichen. Darüber hinaus kann ein vollkommen neuer Grad der Automatisierung erreicht werden. Es gilt jedoch die dafür notwendigen Werkzeuge und Methoden zu entwickeln. Einerseits müssen die Datenanalysen so aufbereitet werden, dass sie von Produktionsleitern, Vorarbeitern und Maschinenführern unkompliziert zur schnellen Optimierung von Betriebsabläufen genutzt werden können. Andererseits müssen diese Informationen auch zur planerischen Optimierung von Geschäftsprozessen genutzt werden. Eine weitere Voraussetzung bilden geeignete Methoden und Werkzeuge zur Entwicklung und zum Betrieb von innovativen und industrienahen Dienstleistungen auf der Basis von Big Data. Die Kombination von Produkt und Datendienstleistungen verspricht insbesondere kundenindividuellen Mehrwert, der nur schwer von Konkurrenten imitiert werden kann.

5.3.4 Entscheiden heißt Daten analysieren

Wir schreiben das Jahr 2020. Unternehmerische Entscheidungen werden nur noch in Ausnahmefällen ohne entsprechende Datenanalysen getroffen. Big Data ist somit zu einem zentralen Baustein von Managementinformationssystemen geworden. Dies ermöglicht deutschen Entscheidern die präzise Einschätzung von Risiken und Chancen sowie eine detaillierte Auseinandersetzung mit möglichen Handlungsalternativen. So können Entscheidungsträger unkompliziert und leicht eigene

Analysen erstellen bzw. selbst Hypothesen für entsprechende Analysen formulieren. Die Zeitspanne zwischen der Entstehung der Daten und der Entscheidungsunterstützung konnte somit stark verkürzt werden. Ebenso konnten typische Fehlentscheidungen vermieden werden, da Entscheidungsverzerrungen erkannt und weitgehend vermieden werden können. Dies ermöglichte es Entscheidungsträgern belastbare und effektivere Entscheidungen zu treffen. Die entsprechend Zuständigen können flexibler und agiler auf Entwicklungen im Markt zu reagieren.

Big Data hat das Potenzial völlig neue Formen der Entscheidungsfindung bzw. Vorbereitung zu ermöglichen. Dafür müssen jedoch Visualisierungs- und Interaktionstechniken entwickelt werden, die es Entscheidern erlaubt, weitgehend selbstständig nachvollziehbare Datenanalysen durchzuführen. Eine zentrale Voraussetzung dafür ist ein fundiertes Verständnis von Entscheidungsverzerrungen wie sie im Umgang mit großen Datenmengen auftreten können. Die theoretischen Grundlagen hierfür sind bereits in der kognitiven Psychologie gelegt, es fehlt jedoch die Umsetzung dieses Wissens in nutzbare Visualisierungs- und Interaktionsmetaphern. Ebenso können hier bereits vorhandene Entwicklungen im Bereich der semantischen Technologien integriert werden.

Darüber hinaus sind entsprechende Fort- und Weiterbildungsangebote zu konzeptionieren, da insbesondere Entscheidungsträger die Rahmenbedingungen und Grenzen von Big Data-Analysen verstehen müssen. Gleichzeitig sind aber auch entsprechende Ausbildungskonzepte und Studiengänge zu entwickeln, die effektiv eine Brücke zwischen der Betriebswirtschaft, der Mathematik und der Informatik schlägt. Entsprechende innovative Ansätze zur Ausbildung von Data Scientists sind bereits in Planung oder Umsetzung. Um die zu erwartende Nachfrage nach entsprechenden Kompetenzen decken zu können, sind jedoch verstärkte Maßnahmen notwendig.

5.4 Zwischenfazit: Wegbereitung einer disruptiven Technologie

Die bisherigen Analysen lassen darauf schließen, dass Big Data bereits in den Unternehmen vorhanden ist. Somit stellt Big Data eine potentiell disruptive Technologie dar. Typisch für solche Technologien ist es, dass sie zunächst Herausforderungen aufwerfen, die nicht unmittelbar gelöst werden können (im Fall von Big Data z.B. die Datenschutzproblematik oder die noch schwierige Nachvollziehbarkeit von Analyseergebnissen für Entscheider) aber erhebliche zukünftige Potenziale erwarten lassen. Tabelle 19: Big Data als disruptive Technologie? bewertet Big Data hinsichtlich seiner Eigenschaft als potentiell disruptive Technologie. Hierzu wurden die Kriterien von Christensen (289) herangezogen.

Kriterium der Innovation	Beschreibung
Innovation führt zunächst zu Lösungen, die im Vergleich mit am Markt dominanten Lösungen schlechter abschneiden.	Big Data ermöglicht zwar die schnelle Verarbeitung von großen Datenmengen, erreicht aber noch nicht die Maßstäbe an Zuverlässigkeit, Konsistenz und Nachvollziehbarkeit wie bestehende Datenverarbeitungstechnologien.
Innovation schneidet in neuen Qualitätsdimensionen besser ab als am Markt dominante Lösungen.	Big Data ermöglicht die um ein Vielfaches schnellere Auswertung von Daten. Dies gilt insbesondere für Auswertungen auf Basis heterogener Datenformate.
Innovation verspricht geringere Kosten für Kunden und geringere Gewinnspannen für Anbieter.	Die Potenziale von Big Data versprechen eine Reihe von Möglichkeiten zur Senkung von Kosten. Hierfür sind allerdings hohe Anschaffungskosten für die Einrichtung der entsprechenden Infrastruktur und Aufbereitung der Daten notwendig. Gleichzeitig sind aber viele Anbieter von Big Data in der Phase der Gewinnung von strategischen Referenzkunden, die typischerweise geringere Gewinnmargen zulassen.
Innovation führt zu einfacheren und komfortableren Lösungen.	Insbesondere für Endkunden führt Big Data bereits heute schon zu komfortableren Lösungen, z.B. im Bereich der sozialen Netzwerke. Auch in Branchen wie der Werbewirtschaft sind einfachere Lösungen bereits vorhanden. Beispiele hierfür finden sich im Management von Werbekampagnen.
Innovation wird von den etablierten Anwendern bestehender Technologien zurückhaltend betrachtet.	Viele große Unternehmen verwenden bereits Pilotanwendungen für Big Data in der Entwicklung, die Kernprozesse werden jedoch nicht berücksichtigt. Insbesondere die diffuse Lage im Datenschutz sowie der Mangel an Fachkräften lässt viele etablierte Unternehmen abwarten.
Die ersten Anwender der Innovation stammen aus Nischenmärkten bzw. einzelnen Branchen.	Big Data wurde zunächst vor allem im wissenschaftlichen Bereich entwickelt und genutzt. Anschließend nutzten vor allem ausgewählte Branchen (Werbewirtschaft) die Vorteile von Big Data.
Vor allem kleine Unternehmen kommerzialisieren die Innovation als erste.	Zentrale technologische Grundlagen von Big Data wurden im Rahmen von Open Source-Projekten entwickelt. Ebenso gibt es eine große Anzahl von Big Data Start-Ups. Allerdings finden sich auch etablierte Unternehmen wie beispielsweise SAP oder Oracle unter den Anbietern von Big Data-Lösungen.
Innovation verändert etablierte Formen der Wertschöpfung.	Das Konzept der Informationsmarktplätze hat das Potenzial die Kollaboration und Vernetzung von Unternehmen, insbesondere im Mittelstand, nachhaltig zu verändern.
Die Innovation führt zu Umwürfen im Markt.	Sollten sich die Potenziale von Big Data erfüllen, sind einige momentan profitable Geschäftsmodelle gefährdet, die bspw. auf der Kalkulation von Risiken oder der Vorhaltung von Arbeitskräften zur Notfallwartung beruhen.

Tabelle 19: Big Data als disruptive Technologie?

Die oben stehende Analyse deutet darauf hin, dass Big Data ein großes Potential für eine disruptive Technologie aufweist. Daher sollte eine effektive Förderung darauf abzielen, die Voraussetzungen für den breiten Einsatz dieser Technologie voranzutreiben. Tabelle 20: Voraussetzungen für die Ausnutzung der Innovationspotenziale von Big Data.

Ebene	Voraussetzungen
Gesetzgebung und Rechtsprechung	Rasche Weiterentwicklung der Gesetzgebung Anpassung an die technischen Möglichkeiten und Erfordernisse sowie an die gesellschaftliche Akzeptanz
Märkte	Big Data als Teil einer konvergenten IT-Landschaft weiterentwickeln Rahmenbedingungen für die Nutzung von Big Data im Mittelstand schaffen Förderung von Big Data-Ökosystemen und -Netzwerken
Unternehmen	Methoden und Techniken zur Bewertung der ökonomischen Sinnhaftigkeit entsprechender Analysen Methoden und Techniken der Entwicklung von datenbasierten Dienstleistungen Methoden und Techniken zur Einbettung von Big Data in Geschäftsprozesse
Individuum	Methoden und Techniken zur Visualisierung und Interaktion mit Big Data- Ausbildung und Weiterbildung im Bereich Data Science

Tabelle 20: Voraussetzungen für die Ausnutzung der Innovationspotenziale von Big Data

Insgesamt kann festgehalten werden, dass die beschriebenen Potenziale die Verfügbarkeit zuverlässiger und nachvollziehbarer Technologie, belastbare Aussagen zur ökonomischen Sinnhaftigkeit und eine breite Akzeptanz samt rechtlicher Zulässigkeit voraussetzen. Die notwendigen technischen Grundlagen für die betriebswirtschaftlich sinnvolle Nutzung von Big Data sind bereits weitgehend vorhanden bzw. werden in den kommenden Jahren verfügbar gemacht. Hier besteht eine besondere Chance für die Förderung von innovativen Unternehmensgründungen in Deutschland, da einerseits wesentliche Entwicklungen von Big Data Technologie auch in Deutschland stattfinden und andererseits eine Vielzahl von Spezialanwendungen für den breiten Einsatz von Big Data notwendig sind. Ebenso kann dort das notwendige Anforderungsprofil an einen Data Scientist erarbeitet werden, bzw. entsteht so die Marktnachfrage nach entsprechenden Kompetenzen.

6 Empfehlungen für Entscheider

Big Data ist ein Sammelbegriff für Technologien zur Sammlung, Verarbeitung und Präsentation von großen, heterogenen Datenmengen die in sehr kurzen Zeiträumen anfallen und für sehr zeitnahe Entscheidungen genutzt werden können. Big Data kann damit disruptive Änderungen in Märkten und Unternehmen fördern. Aus diesen disruptiven Änderungen können erhebliche Chancen und Wettbewerbsvorteile für den deutschen Wirtschaftsstandort erwachsen. Gleichzeitig birgt Big Data aber auch Risiken, die aus Marktverwerfungen oder falschen Erwartungen und Konnotationen entstehen können. In einer globalisierten und eng verzahnten Wirtschaft ist es notwendig, die Rahmenbedingungen zu setzen, die es deutschen Unternehmen ermöglichen die Chancen von Big Data zu nutzen während gleichzeitig die damit verbundenen Risiken effektiv gesteuert werden.

Im Folgenden werden drei Prämissen sowie 6 Handlungsempfehlungen zur Entwicklung effektiver Rahmenbedingungen für Big Data vorgestellt. Die Prämissen betonen zentrale Aspekte einer Förderung von Big Data in Deutschland, die für alle Handlungsempfehlungen als Grundlage dienen. Die Handlungsempfehlungen stellen wichtige Entwicklungsrichtungen dar, die es ermöglichen, die Chancen von Big Data zu nutzen und die damit verbundenen Risiken zu kontrollieren.

6.1 Prämisse

6.1.1 Prämisse 1: Aufklärung und Management der Erwartungshaltung

Wettbewerbsvorteile durch Big Data brauchen einen sachlichen Diskurs zu Chancen und Risiken von Big Data-Technologien und deren Anwendungen mit breiter Beteiligung aus Wirtschaft, Politik, Gesellschaft und Wissenschaft.

Big Data verspricht neuartige Möglichkeiten der Erhebung, Analyse und Präsentation großer Datenmengen unterschiedlichster Herkunft in sehr kurzen Zeiträumen. Einerseits können damit bisherige Entscheidungsprobleme fundierter, schneller und effizienter in Unternehmen gelöst werden. Andererseits können aber auch vollständig neue Formen der unternehmerischen Entscheidungsunterstützung ermöglicht werden, die als Grundlage für neue Produkte bzw. Dienstleistungen, neue Unternehmen oder sogar neue Märkte dienen können. Unternehmen, Behörden, Haushalte und die Gesellschaft können so Daten für zielgenauere Entscheidungen nutzen. Big Data kann also eine wichtige Voraussetzung für Deutschland sein, um in einer globalisierten Wissensgesellschaft die Spitzenstellung unter den Industrienationen zu behaupten und auszubauen.

Aufgrund der Neuheit vieler Big Data-Technologien und der Vielfalt potenzieller Anwendungsbereiche gibt es derzeit aber noch viele Unklarheiten hinsichtlich der Chancen und Risiken von Big Data. Diese Unklarheit wird durch vollmundige Ankündigungen und Versprechen seitens der Technologieanbieter und durch eine, eher emotional, geführte Debatte über die Rahmenbedingungen und Grenzen, insbesondere mit großem Fokus auf Verwendung von personenbezogenen oder personenbeziehbaren Daten verstärkt. Dadurch entstehen sehr viele Zerrbilder bei Anwendern und Entscheidern in Unternehmen, Politik und Gesellschaft. Insbesondere besteht dadurch die Gefahr, dass insbesondere in der öffentlichen Wahrnehmung die Risiken von Big Data entscheidend für die Adoption sind und die Chancen dementsprechend nicht genutzt werden.

Daher ist eine realistische Kommunikation, Demonstration und Diskussion der Chancen und Risiken eine grundlegende Prämisse. Big Data als mögliche disruptive Technologie kann neue Dimensionen für die Qualität von Entscheidungen, aber auch neue Produkte und Dienstleistungen schaffen. Ein gesellschaftlicher Diskurs ist notwendig, um die möglichen Risiken dieser Entwicklungen nicht losgelöst sondern im Kontext der damit verbundenen Chancen für Wirtschaft und Gesellschaft zu sehen. Die Aufklärung der Bevölkerung ist eine wichtige Aufgabe der Anbieter und Anwender von Big Data-Technologien. Ein

wesentliches Element dieses Diskurses sollte die Demonstration der Chancen von Big Data Technologie in Industrien und Wirtschaftszweigen sein, in denen personenbezogene Daten eine untergeordnete Rolle spielen.

6.1.2 Prämisse 2: Verantwortungsvoller Umgang mit Daten

Wettbewerbsvorteile durch Big Data brauchen klare Anforderungen, Voraussetzungen und Grenzen eines verantwortungsvollen Umgangs mit personenbezogenen Daten.

Personenbezogene Daten spielen eine große Rolle bei Anwendung und Nutzung von Big Data-Technologien. Gleichzeitig stehen aber auch die damit verbundenen Risiken im Kern des gesellschaftlichen Diskurses zu Chancen und Risiken von Big Data. Die Analyse und Nutzung personenbezogener Daten ermöglicht es Unternehmen, kundenindividualisierte Produkte und Dienstleistungen anzubieten. Dies kann zu neuen Qualitätsdimensionen für Komfort, Sicherheit und Service bei zunehmend digitalisierten Produkten und Dienstleistungen führen.

Demgegenüber stehen aber Risiken des möglichen Datenmissbrauchs sowie der möglichen Einschränkung der Privatsphäre von Menschen. Insbesondere im Kontext einer globalisierten Wirtschaft und Gesellschaft stehen sich hier auch kulturell unterschiedliche Überzeugungen und Grundwerte gegenüber. Werden die damit verbundenen gesellschaftlichen Konflikte nicht gelöst, werden die Risiken eines Einsatzes von Big Data-Technologien überbewertet und die wirtschaftlichen wie gesellschaftlichen Chancen können nicht genutzt werden. Deutschlands Position im internationalen Wettbewerb wäre damit in diesem Bereich gefährdet.

Dem kann nur mit einem verantwortungsvollen Umgang mit personenbezogenen Daten in Wirtschaft und Gesellschaft begegnet werden. Neben der Entwicklung entsprechender rechtlicher Rahmenbedingungen sollte dabei insbesondere die transparente und proaktive Kommunikation von Unternehmen mit Konsumenten im Vordergrund stehen. Kern dieser Kommunikation sollte die Erläuterung und Aufklärung hinsichtlich der erhobenen, verarbeiteten und genutzten Daten sein. Interessanterweise eignen sich für diese Kommunikation insbesondere Big Data-Technologien. Gemeinsam mit Politik und Gesellschaft sollten Rahmenbedingungen entwickelt werden, die einen verantwortungsvollen Umgang mit personenbezogenen Daten für alle Beteiligten transparent im Kontext einer globalisierten Wirtschaft und Gesellschaft regeln. Solche

Rahmenbedingungen für den verantwortungsvollen Umgang mit personenbezogenen Daten können sich dabei zu einem entscheidenden Wettbewerbsvorteil auf internationaler Ebene entwickeln.

6.1.3 Prämisse 3: Kleine und mittelständische Unternehmen als wichtige Zielgruppe

Wettbewerbsvorteile durch Big Data brauchen die gezielte Unterstützung für kleine und mittelständische Anbieter und Anwender von Big Data Technologie.

Big Data Technologie ist ein Sammelbegriff für eine Vielzahl von generellen und spezialisierten Technologien zur Sammlung, Verarbeitung und Präsentation von Daten. Als potenziell disruptive Entwicklung kann Big Data somit neue Rahmenbedingungen und Wettbewerbsvorteile auf internationaler Ebene schaffen. Davon können insbesondere kleine und agilere Technologieanbieter profitieren. Ebenso entstehen an Universitäten und Hochschulen eine Vielzahl von Technologie- und Anwendungsinnovationen, die einen wichtigen Beitrag zur Adoption von Big Data leisten können. Ebenso kann Big Data die wirtschaftlichen Rahmenbedingungen der Sammlung, Verarbeitung und Präsentation von Daten drastisch verändern. So kann Big Data eine Entscheidungsunterstützung für kleine und mittelständische Unternehmen ermöglichen, die bisher nur großen Unternehmen zugänglich war.

Demgegenüber stehen die speziellen Anforderungen von kleinen und mittelständischen Unternehmen hinsichtlich Risikoappetit, Ressourcen und Reichweite. Zudem ist zwischen kleinen und mittelständischen Anbietern von Big Data Technologie und kleinen und mittelständischen Anwendern von Big Data Technologie zu unterscheiden. Anbieter stehen vor der Herausforderung mit ihren innovativen Lösungen in einem globalen Wettbewerb mit oftmals wesentlich größeren Konkurrenten zu bestehen. Kleine und mittelständische Anwender von Big Data stehen dagegen vor der Herausforderung mit oftmals begrenzten Personal-, Finanz- und Wissensressourcen neuartige Technologien einzuführen, die auch erhebliche organisatorischen Veränderungen nach sich ziehen. Werden jedoch geeignete Rahmenbedingungen durch die Unternehmen, Verbände und die Politik gesetzt kann sich Big Data zu einem Wettbewerbsvorteil für deutsche kleine und mittelständische Unternehmen entwickeln. Gerade für die oftmals international aktiven „Hidden Champions“ im Mittelstand können genauere Entscheidungsgrundlagen dabei helfen, bestehende Spitzenpositionen auszubauen und zu sichern. Für kleine und mittelständische Anbieter ist es

insbesondere wichtig, Ökosysteme mit potenziellen Kunden zu etablieren und mögliche Nischenpositionen frühzeitig identifizieren zu können. Ferner ist eine Strategie für das Bestehen im globalen Wettbewerb unerlässlich. Für kleine und mittelständische Anwender ist es wichtig, Netzwerke zu etablieren, die es ermöglichen, Ressourcen zu teilen bzw. zu bündeln.

6.2 Handlungsempfehlungen

Im Folgenden werden auf Basis der Prämissen spezifische Handlungsempfehlungen beschrieben mit dem Ziel, deutsche Unternehmen darin zu unterstützen, die Chancen von Big Data gezielt zu nutzen und die entsprechenden Risiken in einem akzeptablen Rahmen zu halten. Die Handlungsempfehlungen sind in drei Handlungsfelder gruppiert:

1. Daten nutzbar machen und wertschöpfend nutzen
2. Spitzentechnologien entwickeln und in innovative Produkte und Dienstleistungen umsetzen
3. Nachhaltige Adoption von Big Data sicherstellen

Diese Handlungsfelder sind dabei hinsichtlich des zeitlichen Horizonts der zu erwartenden Effekte geordnet. Kurzfristige Effekte lassen sich dadurch erreichen, in dem Unternehmen dabei unterstützt werden, bestehende Daten aufzubereiten und zur besseren Entscheidungsunterstützung heranzuziehen. Mittelfristige Effekte lassen sich erreichen, indem die Entwicklung innovativer Big Data-Technologien in Deutschland gefördert wird, durch Unterstützung bei der Umsetzung der Technologien in marktreife und innovative Produkte und Dienstleistungen. Langfristig kann eine nachhaltige Adoption von Big Data nur durch die Ausbildung geeigneter Fachkräfte und durch die Etablierung rechtssicherer Gestaltungsspielräume ermöglicht werden.

6.2.1 Daten nutzbar machen und wertschöpfend nutzen

Die folgenden Handlungsempfehlungen zielen darauf ab, Unternehmen dabei zu unterstützen ihre bestehenden Daten nutzbar zu machen und effektiv zur Optimierung von Geschäftsprozessen einzusetzen. Darüber hinaus sollen Daten als Grundlage für neue Produkte und Dienstleistungen genutzt werden können.

Handlungsempfehlung 1: Bisher ungenutzte Daten zur Optimierung operativer Geschäftsprozesse anwenden

Die Zusammenstellung und Aufbereitung von Daten ist ressourcen- und kostenintensiv. Unternehmen scheuen daher Investitionen in Big Data. Es sollten darum Pilotprojekte unterstützt werden, die helfen, den Aufwand und Nutzen von Big Data besser abschätzen zu können. Besonders geeignet dafür sind Datenarchive aus der Fertigung, der Entwicklung oder dem Betrieb. Diese umfassen in der Regel kaum personenbezogene Daten, so dass die rechtlichen Hürden geringer sind. Darüber hinaus können sie spezifischen operativen Prozessoptimierungen zugeordnet werden, so dass eine Nachvollziehbarkeit des Nutzens leichter gewährleistet werden kann. Können die Potenziale von Big Data wirtschaftlich sinnvoll realisiert werden, können belastbare Argumente für den Einsatz von Big Data abgeleitet werden.

Deutsche Unternehmen haben Petabytes von strukturierten Daten auf Band archiviert. Deutsche Unternehmen haben Petabytes von Daten in Archiven, können aber häufig aus Kostengründen auf diese Archive nicht mit analytischen Anfragen zugreifen: Die Daten befinden sich auf Band und müssten zunächst in ein Datenanalysesystem überführt werden, z.B. in Datenbanksysteme wie HANA, ParStream, Exasol, oder verteilte Datensysteme wie HDFS zur Analyse mittels Hadoop, Stratosphere, HADAPT/HAIL oder andere Big Data Systemen.

Unstrukturierte Datenarchive werden aufgrund des Aufwandes nicht genutzt. Archive von Banken, Versicherungen und anderen Unternehmen, die aufgrund von Compliance-Regelungen mehrere Petabyte Daten vorhalten, sind oftmals komplex strukturiert (Text, Audio, Video) und können nicht direkt von klassischer Data Warehouse-Technologie analysiert werden. Viele Big Data-Technologien, insbesondere die Datenspeicherung in verteilten Dateisystemen wie HDFS, erlauben mit Systemen wie z.B. Hadoop oder Stratosphere eine in-situ Analyse der Daten, ohne einen für Data-Warehouses üblichen, aufwändigen, kostspieligen ETL-Prozess. Dies kann die Erstellung einer Datenanalyse beschleunigen und deren Kosten erheblich senken. Durch Big Data-Technologien wird es erstmalig überhaupt erst möglich, existierende Daten aus Archiven auf Festplatten zu überspielen und die vollständige Grundgesamtheit zu analysieren.

Verknüpfung von Prozessoptimierung und Daten. Können entsprechende Datenarchive mit operativen Prozessoptimierungen verknüpft werden, ist es möglich belastbare Aussagen zum wirtschaftlich sinnvollen Umgang mit Big Data abzuleiten. Der Fokus auf unternehmensinterne Daten aus operativen Prozessen ermöglicht niedrige Hürden aus rechtlicher Perspektive. Werden beispielsweise Sensordaten aus der Fertigung genutzt, um die Auslastung von Maschinen in Echtzeit zu überwachen und zu steuern, fallen keine personenbezogenen Daten an. Da es sich um unternehmensinterne Prozesse handelt, stellen sich auch keine Fragen des Urheberrechts. In solchen Pilotprojekten können daher konkrete Aussagen zu Aufwand und Nutzen von Big Data abgeleitet werden.

Wirtschaftlicher Mehrwert als Ausgangspunkt. Der Ausgangspunkt solcher Projekte muss dabei der Geschäftsprozess bzw. eine konkrete Optimierungsfragestellung sein. Daraufhin werden die entsprechenden Datenarchive und Big Data-Technologien ausgewählt. Dieses Vorgehen stellt sicher, dass in diesen Pilotprojekten keine Technologieerprobung im Vordergrund steht, sondern die Optimierung eines konkreten wirtschaftlichen Problems.

Breite Basis an wertschöpfenden Big Data-Anwendungsbeispielen. Das anvisierte Ergebnis dieser Handlungsempfehlung ist eine breite Palette an Anwendungsbeispielen, die nicht nur den innovativen Charakter von Big Data hervorheben, sondern auch konkrete Aussagen zur Wirtschaftlichkeit zulassen.

Handlungsempfehlung 2: Aufbau und Stärkung von Ökosystemen für Datendienstleistungen

Big Data schafft die technologischen Rahmenbedingungen für Datendienstleistungen, d.h., Daten und Datenanalysen werden Wirtschaftsgüter. Es sollten daher Maßnahmen unterstützt werden, die große Datenmengen zur Analyse auch durch Dritte bereitstellen. Insbesondere sollten Unternehmen dadurch befähigt werden, Daten und Datenanalysen kommerziell zu handeln, zu tauschen oder offenzulegen. Diese Maßnahmen sollten es ermöglichen, belastbare Aussagen zu Struktur, Entwicklung und Vertrieb solcher Datendienstleistungen zu treffen. Aufgrund der Neuartigkeit bzw. Volatilität eines solchen Marktes sollte die Entstehung von komplementären Datenanbietern für Deutschlands Kernbranchen wie Industrie, Gesundheit und Mobilität für wichtige öffentliche Daten unterstützt werden.

Datenanbieter sind in Deutschland noch nicht ausreichend in wichtigen Branchen vertreten. Derzeit bieten noch zu wenige Unternehmen Daten an, bzw. verfolgen datengetriebene Verwertungsmodelle. Zu den bereits aktiven Branchen gehören Datenanbieter für Konsumentenprofile, Haushalts-, Geo- und Gebäudedaten, Datenbanken mit Umzugsadressen, Scoringdaten, Inkassodaten, Daten von Auskunfteien und Meldedaten.

Öffentliche Anbieter benötigen Hilfe bei der Verwertung der Daten in Anwendungen. Weitere Anbieter sind beispielsweise Bundes- und Landesämter mit Veröffentlichungspflicht. Leider fehlen diesen öffentlichen Anbietern häufig die entsprechenden Ressourcen, die sowohl die Technologie beherrschen als auch über genügend analytische Expertise verfügen. Durch die Kooperation von Wirtschaftsunternehmen und öffentlichen Institutionen können auch diese Daten entsprechend als Dienstleistungen konzipiert werden.

Erste Projekte beginnen das Monopol US-amerikanischer Anbieter für Webdaten aufzubrechen. Auch kleine und mittelständische Unternehmen und Universitäten bieten mittlerweile Daten zu News und zum Web an, ein Beispiel ist das BMWi geförderte Projekt www.MIA-Marktplatz.de das beispielsweise 80 Mio. News der Jahre 2008 bis 2013 sowie einen Crawl von mehr als 700 Mio. Webseiten anbietet. Derzeit ist MIA eine der wenigen Quellen im deutschsprachigen Raum mit einer derartigen Datengrundlage, die nur derzeit noch von US-amerikanisch gehosteten Servern bzw. Anbieter in China und Russland übertroffen wird.

Branchenspezifische und branchenübergreifende Kooperation ist notwendig. Die wenigen Anbieter in diesem Oligopol haben einen klaren Wettbewerbsvorteil durch die Größe ihres Datenbestandes (z.B. Sprachmodelle) und durch die Größe der Community, die diese Daten produziert (z.B. durch Suchanfragen und Klickströme). Es gibt derzeit leider noch keine Aktivitäten zur Erstellung und Verfügbarmachung von Sprachdaten, also von Daten, die zum Training von Ansätzen für das maschinelle Verarbeiten von Texten benötigt werden, auf Web-Größenordnungen. Selbst das Projekt „Deutscher Wortschatz“ (corpora.uni-leipzig.de) erreicht bei weitem nicht die notwendigen Daten-Größenordnungen und stellt keinerlei sprachtechnologische Vorverarbeitung zur Verfügung. Auch die Datensätze von großen internationalen Unternehmen wie Google, Microsoft, BaiDu, oder Yandex sind nicht vollständig, und v.a. in englischer, chinesischer oder russischer Sprache verfügbar. Nur durch entsprechende Kooperationen können daher die notwendigen wirtschaftlichen Rahmenbedingungen für solche Datendienstleistungen etabliert werden.

Ausgewählte wirtschaftlich nachhaltige Datendienstleistungen für Deutschlands Schlüsselbranchen. Das anvisierte Ergebnis dieser

Handlungsempfehlung ist es, dass für ausgewählte Schlüsselbranchen entsprechende Datendienstleistungen wirtschaftlich nachhaltig angeboten werden können. Hier ist insbesondere der Mittelstand als wichtiger Kunde solcher Dienstleistungen zu beachten.

6.2.2 Spitzentechnologien entwickeln und in innovative Produkte und Dienstleistungen umsetzen

Deutschland hat momentan eine Spitzenposition im Bereich der Forschung und Entwicklung von Big Data-Technologien. Dieser technologische Vorsprung muss für die Entwicklung marktreifer Produkte und Dienstleistungen genutzt werden. Neben den Technologieanbietern sind dafür auch die Anwender zu unterstützen.

Handlungsempfehlung 3: Stärkung von deutschen Technologieanbietern für Big Data (Technology Push)

Im Bereich der Big Data Technologien ist Deutschland durch Forschung an Universitäten und Entwicklungen in Unternehmen sehr gut positioniert. Es sollten daher Maßnahmen unterstützt werden, welche die Kommerzialisierung dieser Technologien in Deutschland und international zum Ziel haben. Dies kann vor allem durch die enge Kooperation von Technologieanbietern und potenziellen Anwendern erfolgen. Die potenziellen Anwender haben so die Möglichkeit, in den Endphasen der Technologieentwicklung in ihrem Sinne einzutreten, der Anbieter kann sein Produkt oder Dienstleistung auf die konkreten Bedürfnisse der Anwender ausrichten. Das Ziel dabei ist es, die Erfolgsschancen der auf den Markt kommenden deutschen Big Data Technologieanbieter zu erhöhen (Technology Push).

Big Data als Chance für deutsche Technologieanbieter. Die neuen Herausforderungen durch Big Data werden durch etablierte Systeme nicht abgedeckt. Daher besteht derzeit die Chance, durch innovative Technologien das im klassischen Datenbankbereich vorhandene Quasimonopol US-amerikanischer Anbieter zu durchbrechen. Eine Vielzahl an deutschen Anbietern, Forschungsinstitutionen und Universitäten sind hierzu sehr gut positioniert. Neben SAP, das mit HANA eine In-Memory-Technologie zur Datenanalyse entwickelt, sowie Software AGs Terracotta-System, sind hierbei insbesondere kleine und

mittelständische Unternehmen zu nennen, wie beispielsweise ParStream, welches aus einem BMWi ZIM Projekt entstanden ist oder Exasol, sowie innovative Technologien aus dem universitären Umfeld wie Stratosphere, Hyper oder Hadoop++/HAIL, welche kommerzialisiert werden könnten. Diese durch Universitätskooperationen entstandenen Unternehmen haben bereits mehrere internationale Preise gewonnen bzw. schlagen die US-amerikanischen Wettbewerber in wichtigen Benchmarks, z.B. im TPC Benchmark.

Ausnutzung der Forschungsexzellenz zu Big Data in Deutschland. An deutschen Universitäten sind mit den Systemen Stratosphere (TU Berlin, HU Berlin, Hasso Plattner Institut), Hyper (TU München), Hadapt/HAIL (Universität des Saarlandes) weitere sehr innovative Systeme und Prototypen entstanden, mit neuen disruptiven Technologien im Bereich der effizienten Spezifikation und skalierbaren Verarbeitung von Verfahren des maschinellen Lernens bzw. der Verarbeitung von Mixed Workloads. Einige dieser Systeme haben bereits ein Ökosystem und eine Nutzerbasis geschaffen und bieten großes Potential für eine erfolgreiche Kommerzialisierung.

Big Data „made in Germany“. Das anvisierte Ergebnis dieser Handlungsempfehlung ist eine enge Kooperation von Anbietern und Anwendern. Dabei sollte das Ziel verfolgt werden, die an deutschen Universitäten entwickelten neuen Technologien schnell zur Produktreife zu bringen. Eine gezielte Förderung von Kooperationen von Unternehmensgründern aus Universitäten und potenziellen Anwendern ermöglicht hier eine schnelle Anwendbarkeit der entsprechenden Technologien. Besonders vielversprechend ist die Verknüpfung von deutschen Anwendern und Anbietern über den kompletten Anwendungsfall und Technologiystack hinweg. Auf diese Weise werden deutsche Anwender und Anbieter gestärkt und Beispieldaten und Referenzkunden etabliert, was für den Markterfolg insbesondere für Startups und kleine und mittelständische Unternehmen ein kritischer Erfolgsfaktor ist.

Handlungsempfehlung 4: Aufbau und Stärkung von branchenspezifischen und branchenübergreifenden Innovationsnetzwerken für Big Data (Market Pull)

Eine wesentliche Herausforderung für die breite Nutzung der Potenziale von Big Data ist es, Bedarf bei potenziellen Anwendern zu identifizieren bzw. zu etablieren (Market Pull). Es sollten dafür Maßnahmen unterstützt werden, die es potenziellen Anwendern und Technologieanbietern von Big Data ermöglicht, sich in Innovationsnetzwerken zusammenzuschließen und datengetriebene Innovationen zu entwickeln. Die Form der Innovationsnetzwerke sichert dabei die Nachhaltigkeit der Innovation über einzelne Unternehmen hinweg und ermöglicht neue Formen der Kooperation in der Nutzung von Daten. Dabei stehen vor allem die betriebswirtschaftlichen Aspekte im Vordergrund sowie die Umsetzung von Big Data Potenzialen in neue Produkte, Dienstleistungen und Geschäftsmodelle. Können hier konkrete und belastbare Anforderungen an Big Data Technologie etabliert werden, entstehen Chancen für Unternehmen entsprechende Angebote zu entwickeln.

Big Data erlaubt die Analyse des gesamten Datenarchivs, aber wann ist das wirtschaftlich sinnvoll? Bisher waren Analysten häufig noch zu einer Verarbeitung von kleinen Stichproben gezwungen. Das Abschätzen geeigneter Stichproben ist fehleranfällig und Interpretationen sind nur eingeschränkt möglich. Big Data erlaubt die Analyse kompletter Datenarchive bei oftmals geringeren Aufwänden. Ein mögliches Anwendungsgebiet ist beispielsweise der Kundendienst, welcher heute hauptsächlich in Web-Foren oder Callcentern durchgeführt wird. Analog ergänzen Außendienstmitarbeiter diese Daten durch weitere unstrukturierte Daten. Big Data Technologie erlaubt nun eine vollständige Analyse dieses heterogenen Datenarchivs hinsichtlich der Zufriedenheit der Kunden mit den Produkten eines Unternehmens. Daraus können neben einer besseren Kundenbetreuung auch wichtige Informationen für die Weiterentwicklung der Produkte abgeleitet werden.

Wann schafft eine interaktive, explorative und visuelle Datenerkundung wirtschaftlichen Mehrwert? Big Data ermöglicht es Endanwendern, sich explorativ mit einem Datenarchiv auseinanderzusetzen. Ein Anwendungsbeispiel hierfür ist die ad-hoc Analyse im Rahmen der strategischen Unternehmenssteuerung. Die Berechnung von Ergebnissen in Sekunden, die durch neue System- und Hardwareentwicklungen ermöglicht werden kann, erlaubt iteratives, interaktives Herantasten an die beste Lösung. Dadurch können auch Domänenexperten ohne technologisches Verständnis, aber mit einem über die Jahre

geprägten, intuitiven Gefühl schnell eine annäherungsweise gute bzw. ausreichende Lösung entwickeln.

Welchen Wert bringt Datenintegration? Netzwerke können als Ausgangspunkt für die Integration von Daten der unterschiedlichen Netzwerkteilnehmer dienen. Insbesondere die branchenübergreifende Kooperation ermöglicht die Integration und nachfolgende holistische Exploration von Daten, welche die Unternehmen einzeln nicht sammeln konnten. Durch die gemeinsame Analyse können die beteiligten Unternehmen ihre individuellen Prozesse optimieren oder gemeinsam neue Produkte oder Dienstleistungen entwickeln.

Innovationsnetzwerke als Experimentierplattformen für Big Data. Es stellt sich allerdings die Frage, wann, für wen und wofür diese und weitere Potenziale von Big Data wirtschaftlich sinnvoll sind. Innovationsnetzwerke ermöglichen es Anwendern, Ressourcen und Kompetenzen zu teilen mit dem Ziel der Entwicklung eines wirtschaftlich tragfähigen Konzepts für den Einsatz von Big Data. Dabei sind branchenspezifische Netzwerke wie auch branchenübergreifende Netzwerke denkbar. Im Kern des Netzwerkes steht die Zusammenarbeit mehrerer Anwender mit ausgewählten Anbietern um ein gemeinsames Thema zu bearbeiten. Ein mögliches Beispiel wäre hier der regionale Zusammenschluss mehrerer kleiner und mittelständischer Unternehmen mit dem Zweck Big Data zur optimierten Materialwirtschaft anzuwenden.

Big Data wird von Unternehmen nachgefragt. Das anvisierte Ergebnis dieser Handlungsempfehlung ist es, tragfähige Innovationen auf Basis von Big Data zu entwickeln. Durch den Zusammenschluss mehrerer Unternehmen können die damit verbundenen Risiken kontrolliert und der Fokus auf nachhaltige Innovation gelegt werden. Es entstehen Konzepte und Prototypen für Produkte und Dienstleistungen, die als Ausgangspunkt für Anbieter genutzt werden können.

6.2.3 Nachhaltige Akzeptanz und Adoption von Big Data sicherstellen

Langfristig können die Chancen von Big Data nur dann genutzt werden, wenn die wirtschaftlichen und gesellschaftlichen Risiken von Big Data kontrolliert werden können. Zentrales Element ist hier ein hohes Maß an Rechtssicherheit im Umgang mit Daten und Datenanalysen. Des Weiteren brauchen die Unternehmen hochqualifizierte Fachkräfte für die wertschöpfende Nutzung von Big Data.

Handlungsempfehlung 5: Erhöhung der Rechtssicherheit für den Umgang mit Big Data und Beseitigung bestehender Hemmnisse

Die Anwendung von Big Data ist bereits heute rechtskonform möglich. Dennoch hemmt die bestehende Rechtslage Unternehmen darin, das volle wirtschaftliche Potenzial von Big Data Anwendungen effektiv auszuschöpfen. Eine Anpassung des Rechtsrahmens an den aktuellen Stand der Technik, insbesondere auf dem Gebiet des Datenschutz- und Urheberrechts, könnte einen entscheidenden Beitrag dazu leisten, Hemmnisse zu beseitigen und die Rechtssicherheit zu erhöhen.

Der Erfolg und die gesellschaftliche Akzeptanz von Big Data hängen davon ab, ob das anwendbare Recht einen angemessenen Ausgleich zwischen dem berechtigten Interesse an der Sammlung und Auswertung der Daten einerseits und insbesondere dem Daten- und Urheberschutz andererseits schafft. Dabei zeigt sich die Notwendigkeit der Modernisierung des bestehenden Rechtsrahmens in nahezu jedem der betroffenen Rechtsgebiete: So ist die Eigentumsfähigkeit von Daten zurzeit noch ungeklärt, was aufgrund der immensen Bedeutung von Daten in der heutigen Informationsgesellschaft als äußerst unbefriedigend empfunden werden muss. Im Hinblick auf das Datenschutzrecht erscheint unter anderem die Regelung der Auftragsdatenverarbeitung (§ 11 BDSG) als nicht länger zeitgemäß. Aus urheberrechtlicher Sicht drängt sich zudem die Frage auf, ob das Urhebergesetz um einen weiteren Erlaubnisvorbehalt ergänzt werden sollte und wie dieser ggf. auszugestalten ist, um den neuen Dimensionen des Datenverkehrs gerecht zu werden. Die dadurch steigende Rechtssicherheit würde es auch Anwendern in Zukunft ermöglichen, die Zulässigkeit des Umgangs mit großen Datenmengen zuverlässiger zu beurteilen.

Handlungsempfehlung 6: Ausbau von Aus- und Weiterbildungsangeboten für Data Science als Schlüsselkompetenz

Es besteht dringender Bedarf an Aus- und Weiterbildungsangeboten zur quantitativen und qualitativen Analyse von großen heterogenen Datenmengen mit geringer Latenz. Hierbei sollten Angebote gefördert werden, welche die Systemsicht von Big Data mit der Analysesicht sowie mit einem verantwortungsvollen, rechtssicheren Umgang von Big Data integrieren. Ebenso müssen ökonomische Aspekte der Nutzung von Big Data berücksichtigt werden.

Data Science ist eine derzeit entstehende Disziplin, die den Umgang mit großen Datenmengen sowohl aus der Sicht der Datenanalyseverfahren als auch der Sicht der Datenanalysesysteme lehrt. Data Science stellt laut McKinsey eine Schlüsselkompetenz für die moderne Wirtschaft und Gesellschaft dar. Daher ist es wichtig, zum einen Studierende mit Techniken, Methoden und Wissen zu Data Science auszubilden. Zum anderen sind auch Fort- und Weiterbildungsmaßnahmen nötig, um Arbeitsnehmern und Entscheidern die Möglichkeiten von Big Data aufzuzeigen und das nötige Handwerkszeug zu vermitteln, um Informationen quantitativ und qualitativ zu analysieren und daraus fundierte Entscheidungen abzuleiten. Dabei sollte ein Data Science Curriculum wesentliche Methoden der Datenanalyse (z.B. Statistik, maschinelles Lernen, Informationsextraktion, Informationsintegration, Sprachverarbeitung, etc.) als auch wesentliche Methoden der skalierbaren Datenverarbeitung (z.B. Datenbanksysteme, Data Warehousing, NoSQL, skalierbare Datenanalyse und Data Mining, Interna von Datenbanken, verteilte Systeme, Rechnerarchitektur, etc.) sowie Kenntnisse zum rechtssicheren und verantwortungsvollen Umgang mit Informationen vermitteln. Insbesondere sollte auf die Gefahren hingewiesen werden, die durch falsche Schlussfolgerungen z.B. aufgrund von falscher Modellwahl (z.B. fälschliche Verwendungen von statischen Momenten bei chaotischen Datenverteilungen), numerischer Instabilität von Datenanalysealgorithmen, zu geringen Stichprobengrößen bzw. mangelnder Qualität der Datenquellen entstehen. Gleichzeitig ist es notwendig, den Studierenden die Methoden und Werkzeuge der ökonomischen Bewertung sowie der organisatorischen Einbettung (Prozessmanagement, Service Engineering) zu vermitteln. Mehrere amerikanische Universitäten haben die Ausbildung von Data Scientists als speziellen Masterstudiengang etabliert. Derzeit sind mehrere Data Science Studiengänge in Deutschland in Planung; an der TU Berlin besteht beispielsweise die Möglichkeit, im Rahmen des Masterstudienganges Informatik eine Spezialisierung als Data Scientist zu belegen. Es gilt, derartige Aktivitäten in Deutschland auszubauen und Konzepte von Data Science als einen festen Bestandteil der Aus- und Weiterbildung in die Curricula nicht nur in den Ingenieur- und Naturwissenschaften, sondern auch in Betriebs- und Volkswirtschaftslehre und in den Geisteswissenschaften zu integrieren.

7 Glossar

Amdahls Law

Gesetz aus der Informatik: Beschreibt den Zusammengang zwischen der Anzahl der Cores, dem Grad der Parallelisierbarkeit des Programms und dem maximal erzielbaren Speedup. Typischerweise ermöglicht nur ein fast zu 100 Prozent parallel ausführbares Programm einen zwei oder höherstelligen Speedup auf einem Rechnercluster (163).

Anfrageoptimierer

Übersetzt deklarative Datenverarbeitungsanfragen in Ausführungspläne. Da typischerweise mehrere äquivalente Ausführungspläne existieren, versucht ein Anfrageoptimierer den effizientesten Ausführungsplan auszuwählen.

Batch-basierte Verarbeitung

Datenverarbeitung bei der alle Daten initial verfügbar sind und verarbeitet werden. Während der Verarbeitung ist keine Interaktion mit dem Verarbeitungsprogramm möglich. Im Gegensatz dazu werden bei Datenstrombasierter Verarbeitung kontinuierlich neue Daten empfangen und verarbeitet. Auch stapel-basierte Verarbeitung genannt.

Big Data

Daten deren Größe, Erfassungsgeschwindigkeit oder Heterogenität die Fähigkeiten gängiger Datenbanksoftwareprodukte zur Verwaltung und zur Analyse übersteigt.

Business Intelligence (BI)	Technologie und Verfahren, um aus Unternehmensdaten relevante Informationen zu extrahieren. Informationen werden unter anderem für Statusreporte (Reporting) oder zur Entscheidungsfindung (Decision Support) verwendet. Häufig realisiert durch Data Warehousing Technologie.
Clustering	Gruppe von Data Mining Verfahren, die eine vorhandene Menge von Elementen anhand von bestimmten Eigenschaften in Gruppen einteilen. Im Gegensatz zur Klassifikation wird kein Trainingsdatensatz mit bekannten Gruppenzugehörigkeiten verwendet, sondern die Zugehörigkeiten aller Elements ist unbekannt.
Data Scientist	Dieses neuartige Berufsbild kombiniert Methodenkenntnisse in Datenanalyseverfahren (Statistik und maschinelles Lernen, Optimierung, linearer Algebra, Signalverarbeitung, Sprachverarbeitung, Data Mining, Text Mining, Video Mining, Bildverarbeitung) mit technischen Kompetenzen im Bereich des skalierbaren Datenmanagements (Datenbanksysteme, Data-Warehousing, verteilte Systeme, Rechnernetze, Rechnerarchitekturen) sowie praktischen Kompetenzen der Systemimplementierung. Dabei sollte eine derartige Ausbildung durch praktische Anwendungsprojekte zur Vermittlung von Fachkompetenzen in bestimmten Anwendungsdomänen flankiert werden.
Datenanalysesystem	Ein System, das komplexe Analyseprogramme auf Daten ausführt. Idealer Weise werden Analyseprogramme deklarativ spezifiziert und von einem Optimierer in einen effizienten Ausführungsplan übersetzt. Im Unterschied zu Datenbanksystemen sind Datenanalysesysteme leseorientiert und erlauben keine oder nur eingeschränkte Datenaktualisierungen.

Datenbanksystem	Ein System zum Verwalten von Daten. Datenbanksysteme erlauben das Schreiben, Aktualisieren und Abfragen von Daten. Datenbanksysteme stellen üblicherweise die Konsistenz von Daten auch bei gleichzeitigen Schreib- und Leseoperationen sicher.
Datenfluss	Modell zum Beschreiben von Datenverarbeitungsprogrammen. Daten „fließen“ von einer Datenquelle durch Operatoren, die die Daten verarbeiten, in eine Datensenke, die das finale Ergebnis enthält.
Datenmodell	Ein Datenmodell gibt vor welche Daten von einem System verarbeitet werden können und wie diese strukturiert werden müssen. Beispiele sind das Key-Value Modell, das Schlüssel-Wert Paare definiert, oder das relationale Modell, das Tupel mit festem Schema erfordert. Strikte Datenmodelle sind häufig effizienter zu verarbeiten.
Datenschutz	Regelungsbereich, der den Betroffenen vor Gefahren schützt, die sich aus dem Umgang mit ihm betreffenden personenbezogenen Daten ergeben. Verwirklicht wird dies durch verbindliche Normen, die sich an diejenigen Stellen richten, die personenbezogene Daten erheben, verarbeiten und nutzen.
Deklarative Programmiersprachen	Gruppe von Programmiersprachen. Im Gegensatz zu den imperativen Sprachen, beschreiben deklarative Sprachen das gewünschte Ergebnis und spezifizieren nicht den Weg, um dieses Ergebnis zu berechnen. Die tatsächliche Berechnungsvorschrift von deklarativen Programmen wird automatisch von einem Kompilierer oder Optimierer erstellt. Prominentes Beispiel einer deklarativen Sprache ist SQL.

Extract-Transform-Load (ETL)	Traditionelle Schritte zur Integration von Daten. Daten werden aus der ursprünglichen Quelle extrahiert (extract), in ein einheitliches Format konvertiert (transform) und schließlich in das Zielsystem importiert (load).
Fehlertoleranz	Wichtige Eigenschaft von parallelen Datenverarbeitungssystemen. Da es in Umgebungen mit vielen Rechnern häufig zu Ausfällen von Hardware und Software kommt, müssen parallele Systeme in der Lage sein, auch im Fehlerfall den Betrieb und die Ausführung von Datenanalyseprogrammen fortzusetzen.
Haftungsrecht	Sammelbegriff für Rechtsnormen, die regeln, ob und wie eine bestimmte Person mit ihrem Vermögen für eine eigene oder fremde Schuld einstehen muss.
Heterogene Daten	Beschreibt Daten mit unterschiedlichen Datenmodellen und Schemata, zum Beispiel relationale Daten, Fließtext und Bilddaten.
Informationsextraktion (IE)	Verfahren, um aus unstrukturierten Daten (z.B. Fließtext oder Tonaufnahmen) strukturierte Daten zu extrahieren. Wird benötigt, da viele Analyseprogramme und -verfahren strukturierte Daten erfordern.
Klassifikation	Teilgebiet des Maschinellen Lernens mit dem Ziel bestimmte Elemente (z.B. Filme) anhand von ihren Eigenschaften (z.B. Regisseur, Länge, Schauspieler) einer Gruppe (z.B. einem Genre) zu zuordnen. Dafür wird mit einem Trainingsdatensatz von bekannten Elementen ein Klassifikationsmodell trainiert, das ungesuchte Elemente einordnen kann.

Materialisierung	Datenverarbeitungstechnik bei der im Gegensatz zum Pipelining Daten zunächst gesammelt werden bevor sie an den nächsten Operator weitergegeben werden. Dadurch kann zum Teil der Datendurchsatz verbessert werden, jedoch auch die Antwortlatenz leiden.
Natural Language Processing (NLP)	Klasse von Informationsextraktionsverfahren aus natürlich-sprachlichen Texten.
Online Analytical Processing (OLAP)	Klasse von Datenbankanwendungen, die hauptsächlich lesend auf Daten zugreift. Dabei werden häufig große Datenmengen gelesen und aggregiert. Anfragen sind häufig sehr komplex und rechenintensiv.
Online Transactional Processing (OLTP)	Klasse von Datenbankanwendungen, die schreibend und lesend auf Daten zugreift. Schreib- und Leseoperationen sind dabei typischerweise jeweils auf wenige Datensätze beschränkt und vergleichsweise einfach. OLTP ist herausfordernd bei sehr vielen gleichzeitigen Lese- und Schreiboperationen.
Part-of-Speech (POS)	Häufiger (Vor-)Verarbeitungsschritt für die Informationsextraktion aus Fließtexten. Annotiert die einzelnen Wörter in einem Text mit ihrer lexikalischen Klasse (z.B. Verb, Adjektiv oder Nomen).
Pipelining	Datenverarbeitungstechnik bei der Daten von einem Operator direkt zum nächsten Operator weitergegeben werden. Dadurch kann die Latenz der Antwortzeit reduziert werden. Ist jedoch nicht immer möglich, da für einige Operatoren (z.B. Sortierung) alle Daten gesehen werden müssen.
Predictive Analysis	Teilgebiet des maschinellen Lernens mit dem Ziel aus historischen Daten Vorhersagemodelle für die Zukunft zu berechnen.

Programmiermodell	Schnittstelle um Programme zu definieren. Wichtige Eigenschaften von Programmiermodellen sind Abstraktion und Ausdruckstärke. Abstraktion beschreibt den Aufwand, um ein Programm zu schreiben. Ausdruckstärke die „Menge“ von Programmen, die mit einem Programmiermodell möglich sind.
Relationales Datenmodell und relationale Algebra	Grundlage von relationalen Datenbanksystemen und der Anfragesprache SQL. Definiert die Struktur von Daten, Operatoren und Äquivalenzregeln zwischen Ausdrücken.
Replikation	Technik um Ausfallsicherheit in verteilten Systemen zu gewährleisten. Daten werden auf mehreren Rechnern gespeichert, um im Fehlerfall immer noch mindestens eine Kopie der Daten verfügbar zu haben.
Rule Mining	Maschinelles Lernverfahren, das aus Daten bestimmte Zusammenhänge in Form von Regeln erlernt. Zum Beispiel können Warenkörbe ausgewertet werden, um herauszufinden welche Produkte häufig den Kauf von anderen Produkten zur Folge haben.
Schutzrecht sui generis	Begriff, der genutzt wird, um ein Schutzrecht zu beschreiben, das nicht die typischen Charakteristika anderer Rechte aufweist und deshalb auch nicht in ein vorhandenes Klassifikationssystem eingegliedert werden kann.
Territorialitätsprinzip	Grundsatz, der klärt, welches Recht auf welche Personen wann und an welchem Ort anwendbar ist. Generell gilt, dass alle Personen den Gesetzen des Staates unterworfen sind, auf dessen Territorium sie sich jeweils befinden.

Urheberrecht	Rechtsgebiet, das dem Schutz des Urhebers in seiner geistigen und persönlichen Beziehung zum Werk und in der Werknutzung dient. Neben Werken werden auch Leistungsschutzrechte im UrhG normiert, deren Schutzmfang im Vergleich zum Werkschutz jedoch beschränkt ist.
User-Defined Function (UDF)	Von einem Benutzer geschriebener Programmcode, der von einem Datenverarbeitungssystem ausgeführt wird. Datenverarbeitungssysteme stellen Schnittstellen für UDFs für erweiterbare Funktionalität bereit.
User-Generated Content	Inhalte, die nicht vom Anbieter sondern vom Nutzer eines Medienangebotes erstellt werden.
Variety	Eine der Eigenschaften von Big Data, die die große Heterogenität der Daten beschreibt.
Velocity	Eine der Eigenschaften von Big Data, die den hohen Datendurchsatz beschreibt.
Verkehrssicherungspflicht	Pflichten, die vom Verkehrssicherungspflichtigen fordern, die Allgemeinheit vor Gefahrenquellen zu schützen. Verkehrssicherungspflichtig sind Personen, die Gefahrenquellen schaffen oder unterhalten. Bei Nichtbeachtung drohen Schadensersatzansprüche.
Volume	Eine der Eigenschaften von Big Data, die das große Datenvolumen beschreibt.

8 Literatur

- (1) Christoph Boden, Alexander Löser, Christoph Nagel, Stephan Pieper: FactCrawl: A Fact Retrieval Framework for Full-Text Indices. WebDB 2011
- (2) Xin Luna Dong, Divesh Srivastava: Detecting Clones, Copying and Reuse on the Web (Tutorial). DASFAA (2) 2012: 324-325
- (3) Xin Dong, Laure Berti-Equille, Yifan Hu, Divesh Srivastava: Global Detection of Complex Copying Relationships Between Sources. PVLDB 3(1): 1358-1369 (2010)
- (4) Xin Luna Dong, Felix Naumann: Data fusion - Resolving Data Conflicts for Integration. PVLDB 2(2): 1654-1655 (2009)
- (5) Georg Rehm und Hans Uszkoreit (Hrsg.): Strategic Research Agenda for Multilingual Europe 2020. Springer (2013)
- (6) Georg Rehm und Hans Uszkoreit (Hrsg.): Die Deutsche Sprache im Digitalen Zeitalter. Springer (2012)
- (7) AG Rechtsrahmen des Cloud Computing, Trusted Cloud - Datenschutzrechtliche Lösungen für Cloud Computing, 2012
- (8) Baeriswyl, «Big Data» ohne Datenschutz-Leitplanken, Zeitschrift für Datenrecht und Informationssicherheit, 2013, 14
- (9) Benlian/Hess/Buxmann, Software as a Service - Anbieterstrategien, Kundenbedürfnisse & Wertschöpfungsstrukturen, Wiesbaden 2010
- (10) Buchner, Die Einwilligung im Datenschutzrecht – vom Rechtfertigungsgrund zum Kommerzialisierungsinstrument, DuD 2010, 39 ff.
- (11) Conrad, Anmerkung zu BGH, Urteil vom 29. April 2010, I ZR 69/08 – Vorschaubilder, ZUM 2010, 585
- (12) Drewes/Siegert, Die konkludente Einwilligung in das Telefonmarketing und das Ende des Dogmas von der datenschutzrechtlichen Schriftform, RDV 2006, 139
- (13) Dreier/Schulze, Urheberrechtsgesetz, 4. Auflage München 2013
- (14) Dreyer/Kotthoff/Meckel, Urheberrecht, 2. Auflage Heidelberg 2008
- (15) Fromm/Nordemann, Urheberrecht, 10. Auflage Stuttgart 2008

- (16) Gasser, Information Quality and the Law, or, How to Catch a Difficult Horse, Research Publication No. 2003-08 (11/2003), <http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/2003-08.pdf> (zuletzt abgerufen am 05.07.2013)
- (17) Gaster, Der Rechtsschutz von Datenbanken, Köln 1999
- (18) Gerstenberg, Löschen von Tonbändern als neuer strafrechtlicher Tatbestand, NJW 1956, 540
- (19) Gola/Schomerus, Kommentar zum Bundesdatenschutzgesetz, 11. Auflage München 2012
- (20) Götting, Anmerkung zu BGH, Urteil vom 29.04.2010 - I ZR 69/08, LMK 2010, 309481
- (21) Grünwald/Döpkens, Cloud Control? - Regulierung von Cloud Computing-Angeboten, MMR 2011, 287
- (22) Hartmann, Weiterverkauf und „Verleih“ online vertriebener Inhalte - Zugleich Anmerkung zu EuGH, Urteil vom 3. Juli 2012, Rs. C-128/11-UsedSoft ./ Oracle, GRUR-Int 2012, 980
- (23) Hoeren/Sieber, Multimedia-Recht, 32. Ergänzungslieferung München 2012
- (24) Hoeren, Wenn Sterne kollabieren, entsteht ein schwarzes Loch – Gedanken zum Ende des Datenschutzes, ZD 2011, 145
- (25) Hoeren/Försterling, Onlinevertrieb „gebrauchter“ Software - Hintergründe und Konsequenzen der EuGH-Entscheidung „UsedSoft“, MMR 2012, 642
- (26) Hüttner, 1, 2, 3, 4 Eckstein, keiner muss versteckt sein? - Wer sich im Internet präsentiert, muss mit Google rechnen!, WRP 2010, 1008
- (27) Kieun, Database Technology for Large Scale Data, <http://www.cubrid.org/blog/web-2-0/database-technology-for-large-scale-data/> (zuletzt aufgerufen am 2.5.2013)
- (28) Klass, Neue Internettechnologien und das Urheberrecht: Die schlichte Einwilligung als Rettungsanker?, ZUM 2013, 1
- (29) Kühling/Seidel/Sivridis, Datenschutzrecht, 2. Auflage Heidelberg 2011
- (30) Lackner/Kühl, Kommentar zum Strafgesetzbuch, 27. Auflage München 2011
- (31) Lehmann, Produkt- und Produzentenhaftung für Software, NJW 1992, 1721
- (32) Maunz/Dürig, Grundgesetz-Kommentar, 66. Ergänzungslieferung München 2012
- (33) Meyer, Die Haftung für fehlerhafte Aussagen in wissenschaftlichen Werken, ZUM 1997, 26
- (34) Mayer-Schönberger/Cukier (2013), Big Data – A revolution that will transform how we live, work, and think, Boston
- (35) Meyer/Wehlau, Die zivilrechtliche Haftung für Datenlöschung, Datenverlust und Datenzerstörung, NJW 1998, 1585

- (36) Münchener Kommentar, Bürgerliches Gesetzbuch, Bd. 1, 6. Auflage München 2012
- (37) Nägele/Jacobs, Rechtsfragen des Cloud Computing, ZUM 2010, 281
- (38) Palandt, Bürgerliches Gesetzbuch, 72. Auflage München 2013
- (39) Plath (Hrsg.), BDSG Kommentar, Köln 2013
- (40) Redeker, Software – ein besonderes Gut, NJW 2008, 2684
- (41) Reese, Produkthaftung und Produzentenhaftung für Hard- und Software, DStR 1994, 1121
- (42) Rogosch, Die Einwilligung im Datenschutzrecht, Frankfurt a.M. 2013
- (43) Roßnagel/Pfitzmann/Garstka, Modernisierung des Datenschutzrechts, Berlin 2001
- (44) Schneider, Hemmnis für einen modernen Datenschutz: Das Verbotsprinzip, Anwaltsblatt 2011, 233
- (45) Schönke/Schröder, Kommentar zum Strafgesetzbuch, 28. Auflage München 2010
- (46) Schricker/Loewenheim, Urheberrecht, 4. Auflage München 2010
- (47) Schumacher, Service Level Agreements: Schwerpunkt bei IT- und Telekommunikationsverträgen, MMR 2006, 12
- (48) Sendrowski, Zum Schutzrecht „sui generis“ an Datenbanken, GRUR 2005, 369
- (49) Simitis, Kommentar zum Bundesdatenschutzgesetz, 7. Auflage Baden-Baden 2011
- (50) Solmecke/Wahlers, Rechtliche Situation von Social Media Monitoring-Diensten - Rechtskonforme Lösungen nach dem Datenschutz- und dem Urheberrecht, ZD 2012, 550
- (51) Spies, Cloud Computing: Keine personenbezogenen Daten bei Verschlüsselung, MMR-Aktuell 2011, 313727
- (52) Spindler/Hillegeist, Rechtliche Probleme der elektronischen Langzeitarchivierung von Forschungsdaten, in: Büttner/Hobohm/Müller (Hrsg.), Handbuch Forschungsdatenmanagement, Bad Honnef 2011
- (53) Spindler, Bildersuchmaschinen, Schranken und konkludente Einwilligung im Urheberrecht - Besprechung der BGH-Entscheidung „Vorschaubilder“, GRUR 2010, 785
- (54) Staudinger, Kommentar zum Bürgerlichen Gesetzbuch mit Einführungsgesetz und Nebengesetzen, Buch 3 (Berlin 2002), Buch 2 (Berlin 2004, 2011)
- (55) Taeger/Gabel (Hrsg.), Kommentar zum BDSG und zu den Datenschutzzvorschriften des TKG und TMG, Frankfurt a.M. 2010
- (56) Wandtke/Bullinger, Praxiskommentar zum Urheberrecht, 3. Auflage München 2009.
- (57) Weichert, Big Data und Datenschutz - Chancen und Risiken einer neuen Form der Datenanalyse, ZD 2013, 251
- (58) Welp, Datenveränderung (§ 303a StGB) – Teil 1, IuR 1988, 443 ff.

- (59) Wicker, Vertragstypologische Einordnung von Cloud Computing-Verträgen - Rechtliche Lösungen bei auftretenden Mängeln, MMR 2012, 783
- (60) Zieger/Smirra, Fallstricke bei Big Data-Anwendungen - Rechtliche Gesichtspunkte bei der Analyse fremder Datenbestände, MMR 2013, 418
- (61) D. Agrawal, P. Bernstein, E. Bertino, et al: Challenges and Opportunities with Big Data - A community white paper developed by leading researchers across the United States. <http://www.purdue.edu/discoverypark/cyber/assets/pdfs/BigDataWhitePaper.pdf>, (letzter Zugriff: 12.6.2013).
- (62) C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2007.
- (63) S. Bernhard, K. Al Zoukra, Ch. Schütte (2010) From non-invasive hemodynamic measurements towards patient-specific cardiovascular diagnosis. In: Quality Assurance in Healthcare Service Delivery, Nursing and Personalized Medicine. Hershey, PA: Medical Information Science Reference.
- (64) L. Bottou, O. Bousquet,: The Tradeoffs of Large Scale Learning, 2008, Advances in Neural Information Processing Systems (NIPS), 20:161–168.
- (65) J. Betteridge, A. Carlson, S. A. Hong, E. R. Hruschka Jr., E. L. M, Law, T. M. Mitchell, S. H. Wang: Toward never ending language learning. In: Proceedings of the 2009 AAAI, Spring Symposium on Learning by Reading and Learning to Read. 2009.
- (66) M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, O. Etzioni: Open information, extraction from the Web. In: Proceedings of the 20th IJCAI. 2670–2676, 2007.
- (67) P. Baumann, A. Dehmel, P. Furtado, R. Ritsch, N. Widmann. 1998. The multidimensional database system RasDaMan. SIGMOD Rec. 27, 2 (June 1998), 575-577.
- (68) D. Battré, S. Ewen, F. Hueske, O. Kao, V. Markl, D. Warneke: "Nephele/PACTs: a programming model and execution framework for web-scale analytical processing", In: SoCC, pp. 119-130 (2010).
- (69) M. Berlin. Consistent and Partition-Tolerant File Replication in the Cloud File System XtreemFS. Praxis der Informationsverarbeitung und Kommunikation, Vol. 36, No. 1, p. 39, 2013.
- (70) Y. Bu, B. Howe, M. Balazinska, M. D. Ernst: HaLoop: Efficient iterative data processing on large clusters. Proceedings of the VLDB Endowment, 3(1-2), pp. 285-296.6. (2010).
- (71) D. Battré, M. Hovestadt, O. Kao, A. Keller, K. Voss: Virtual Execution Environments for ensuring SLA-compliant Job Migration in Grids, 2008 IEEE International Conference on Services Computing (SCC 2008), 571-572. 2008.

- (72) A. Binder, A., K. R. Müller, M. Kawanabe: On Taxonomies for Multi-class Image Categorization, 2012, International Journal of Computer Vision, 99(3):281-301.
- (73) P. v. Bünaeu,, F. C. Meinecke, F. Király, K. R. Müller: Finding Stationary Subspaces in Multivariate Time Series, 2009, Physical Review Letters, 103:214101.
- (74) F. Bießmann, F. C. Meinecke, A. Gretton, A. Rauch, G. Rainer, N. K. Logothetis, K. R. Müller, Temporal kernel CCA and its application in multimodal neuronal data analysis, Machine Learning 79 (1-2), 5-27, 2010.
- (75) Bosch-Nord-Amerika ist an der Entwicklung des Programms AIDA beteiligt, welches zur stark automatisierten Berechnung vom Materialeigenschaften eingesetzt wird. Es wurde auch damit begonnen, Big- Data-Analysemethoden zu entwickeln: <http://www.bosch.us/content/language1/html/rtc.htm>.
- (76) N. Cesa-Bianchi, G. Lugosi: Prediction, learning, and games, Cambridge University Press, 2006.
- (77) S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, Computational Materials Science 68, 314-319 (2013).
- (78) E. Chang, S-F. Chang, A. G. Hauptmann, T. S. Huang, M. Slaney: Web-Scale Multimedia Processing and Applications, Proc. IEEE Vol. 100, No. 9, September 2012.
- (79) D. Chamberlin, D. Florescu, J. Robie, J. Simeon, M. Stefanescu: XQuery: A query language for XML. In: International Conference on Management of Data: Proceedings of the ACM SIGMOD international conference on Management of data, Vol. 9, No. 12, pp. 682-682 (2003).
- (80) Chu, Ch.-T., Kim, S. K., Lin, Y.-A., Yu, Y. Y., Bradski, G., Ng, A. Y., Olukotun, K.: Map-Reduce for Machine Learning on Multicore, 2006, Advances in Neural Information Processing Systems (NIPS).
- (81) P. Cudre-Mauroux et al. A demonstration of SciDB: a science-oriented DBMS, in Proceedings of the VLDB Endowment, 2(2):1534-1537, August 2009.
- (82) G. Cormode, S. Muthukrishnan: An improved data stream summary: The count-min sketch and its applications. 2005, LATIN 2004, J. Algorithm 55(1): 58-75.
- (83) T. M. Deserno (2011): Biomedical Image Processing, Series on Biological and Medical Physics, Biomedical Engineering, Springer.
- (84) J. Dean, S. Ghemawat: "MapReduce: simplified data processing on large clusters". In Communications of the ACM 51(1), pp.107-113, ACM (2008).

- (85) J. Dean, S. Ghemawat: MapReduce: simplified data processing on large clusters, In OSDI, pp. 137-150 (2004).
- (86) R. Datta, D. Joshi, J. Li, J. Z. Wang: Image retrieval: Ideas, influences, and trends of the new age, ACM Computing Surveys (CSUR), Volume 40, Issue 2, April 2008.
- (87) S. Edlich et al., „NoSQL - Einführung in nichtrelationale Web 2.0 Datenbanken“, 2. Auflage, Hanser Verlag, ISBN-13: 978-3446427532, Sept 2011.
- (88) M. Erdt, S. Steger, G. Sakas (2012): "Segmentation: A new view of image segmentation and registration.", Journal of Radiation Oncology Informatics 4(1), doi :10.5166/jroi-4-1-19 , pp. 1-23.
- (89) S. Ewen, K. Tzoumas, M. Kaufmann, V. Markl: Spinning Fast Iterative Data Flows. PVLDB 5(11): pp. 1268-1279 (2012).
- (90) G. M. Fiedler, A. Leichtle, J. Kase, S. Baumann, U. Ceglarek, K. Felix, T. O. F. Conrad, H. Witzigmann, A. Weimann, Ch. Schütte, J. Hauss, M. Büchler, J. Thiery (2009) Clinical Cancer Research, 15 (11). pp. 3812-3819.
- (91) B. Glocker, J. Feulner, A. Criminisi, D. R. Haynor E. Konukoglu (2012): Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012 (S. 590-598). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-33454-2_73.
- (92) A. Gupta, L. V. Kalé, D. Miložić, P. Faraboschi, S. M. Balle: HPC-Aware VM Placement in Infrastructure Clouds, IEEE Intl. Conf. on Cloud Engineering IC2E, 2013.
- (93) D. Gupta, K. V. Vishwanath, M. McNett, A. Vahdat, K. Yocom, A. Snoeren, G. M. Voelker: DieCast: Testing Distributed Systems with an Accurate Scale Model, ACM Transactions on Computer Systems, Volume 29 Issue 2, May 2011.
- (94) S. Harmeling, G. Dornhege, D. Tax, F. C. Meinecke,, K. R. Müller: From outliers to prototypes: ordering data, Neurocomputing, 69(13-15):1608-1618., 2006.
- (95) M. Hovestadt, O. Kao, A. Keller, A. Streit: Scheduling in HPC Resource Management Systems: Queuing vs. Planning, Proceedings of the 9th International Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP 2003), LNCS 2862, pp. 1-20, Springer 2003.
- (96) F. Hueske, M. Peters, A. Krettek, M. Ringwald, K. Tzoumas, V. Markl, J. Freytag: Peeking into the Optimization of Data Flow Programs with MapReduce-style UDFs. IEEE International Conference on Data Engineering (ICDE) (2013).
- (97) W. Hu, N. Xie, L. Li, X. Zeng: A survey on visual content-based video indexing and retrieval, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Volume 41, Issue 6, October 2011.

- (98) M. Isard, M. Budiu, Y. Yu, et al: "Dryad: distributed data-parallel programs from sequential building blocks". In Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems, ACM (2007).
- (99) O. Kao: On Parallel Image Retrieval with Dynamically Extracted Features, Journal for Parallel Computing. Elsevier, 700-709. 2008.
- (100) M. Kloft, U. Brefeld, S. Sonnenburg, P. Laskov, K. R. Müller, A. Zien: Efficient and accurate ℓ_p norm multiple kernel learning, Advances in neural information processing systems 22, 997-1005, 2012.
- (101) A. Khan, D. H. Winn, M. Bendixen: SQL Server Debugging in a Distributed Database Environment, Patent No.: US 7,500,225 B2, Mar 3., 2009.
- (102) H. Kache, W.-S. Han, V. Markl, V. Raman, S. Ewen: POP/FED: Progressive Query Optimization for Federated Queries in DB2. VLDB 2006: 1175-1178.
- (103) Y. Low, D. Bickson, J. Gonzalez, C. Guestrin, A. Kyrola, J. M. Hellerstein: DistributedGraphLab: A framework for machine learning and data mining in the cloud. Proceedings of the VLDB Endowment, 5(8), pp. 716-727 (2012).
- (104) D. Levin, M. Canini, S. Schmid, A. Feldmann: Toward Transitional SDN Deployment in Enterprise Networks, Open Networking Summit (ONS), 2013.
- (105) S. Lorenz, M. Scheffler, A. Gross, Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks. Chem. Phys. Lett. 395, 210-215 (2004), und Descriptions of surface chemical reactions using a neural network representation of the potential-energy surface (13 pages). Phys. Rev. B 73, 115431 (2006).
- (106) B. Lohrmann, O. Kao: Processing Smart Meter Data Streams in the Cloud, Innovative Smart Grid Technologies Conference Europe (ISGT Europe), 2011 IEEE PES, . Pp. 1 -- 8, Print ISBN: 978-1-4577-1422-1 2011, 2011.
- (107) N. Lao, T. Mitchell, W. W. Cohen: Random walk inference and learning in a large scale, knowledge base. In: Proceedings of EMNLP, Edinburgh, Scotland, UK., Association for Computational Linguistics (July 2011) 529–539.
- (108) Q. Li, M. Shao, V. Markl, K. S. Beyer, L. S. Colby, G. M. Lohman: Adaptively Reordering Joins during Query Execution. ICDE 2007: 26-35.
- (109) Lamecker, Pennec (2010): "Atlas to Image-with-Tumor Registration based on Demons and Deformation Inpainting." MIDAS Journal, Contribution to MICCAI 2010 Workshop on Computational Imaging Biomarkers for Tumors (CIBT).

- (110) B. Lohrmann, D. Warneke, O. Kao: Massively-parallel Stream Processing under QoS Constraints with Nephele, Proceedings of the 21st International Symposium on High-Performance Parallel and Distributed Computing. ACM, 271–282. 2012.
- (111) G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, G. Czajkowski: Pregel: a system for large-scale graph processing. In: Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, ACM, pp. 135-146 (2010).
- (112) "Materials Project" at MIT (accelerating materials discovery through advanced scientific computing and innovative design tools): <https://www.materialsproject.org/>.
- (113) M. Mintz, S. Bills, R. Snow, D. Jurafsky: Distant supervision for relation extraction without labeled data. In: Proceedings of ACL/AFNLP. 1003–1011, 2009.
- (114) McKinsey Global Institute: Big data: The next frontier for innovation, competition, and productivity. http://www.mckinsey.com/~media/McKinsey/dotcom/Insights_Prozent20and_Prozent20pubs/MGI/Research/Technology_Prozent20and_Prozent20Innovation/Big_Prozent20Data/MGI_big_data_full_report.ashx, (letzer Zugriff: 12.6.2013).
- (115) N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner. OpenFlow: enabling innovation in campus networks, ACM SIGCOMM Computer Communication Review, April 2008.
- (116) F. McSherry, R. Isaacs, M. Isard, D.G. Murray: Composable Incremental and Iterative Data-Parallel Computation with Naiad. Technical Report MSR-TR-2012-105, Microsoft Research, (2012).
- (117) C. du Mouza, W. Litwin, P. Rigaux: Large-scale indexing of spatial data in distributed repositories: the SD-Rtree. VLDB J. 18(4): 933-958 (2009).
- (118) V. Markl, G. Lohman, V. Raman: LEO: An autonomic query optimizer for DB2. IBM Systems Journal 42(1): 98-106 (2003).
- (119) K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, B. Schölkopf: An Introduction to Kernel-based Learning Algorithms, 2001, IEEE Neural Networks, 12(2):181-201.
- (120) G. Montavon, G. B. Orr, K. R. Müller, (Eds.): Neural Networks: Tricks of the trade, Reloaded, 2012, Springer, Lecture Notes in Computer Science (LNCS), 7700.
- (121) G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K. R. Müller, O. A. v. Lilienfeld, Machine Learning of Molecular Electronic Properties in Chemical Compound Space, New Journal of Physics, Focus Issue, Novel Materials Discovery, 2013.

- (122) V. Markl, V. Raman, D. E. Simmen, G. M. Lohman, H. Pirahesh: Robust Query Processing through Progressive Optimization. SIGMOD Conference 2004: 659-670.
- (123) Chang Sun Kong, Wei Luo, Sergiu Arapan, Pierre Villars, Shuichi Iwata, Rajeev Ahuja, and Krishna Rajan, Information-Theoretic Approach for the Discovery of Design Rules for Crystal Chemistry, *J. Chem. Inf. Model.* 52, 1812–1820 (2012)
- (124) C. D. Manning, H. Schütze: Foundations of Statistical Natural Language Processing. The MIT Press. ISBN 978-0-262-13360-9.1999.
- (125) M. Naaman: Social Multimedia: Highlighting Opportunities for Search and Mining of Multimedia Data in Social Media Applications. In *Multimedia Tools and Applications*, April 2010.
- (126) P. Ndjiki-Nya, D. Doshkov, H. Kaprykowsky, F. Zhang, D. Bull, T. Wiegand: Perception-oriented video coding based on image analysis and completion: A review, *Signal Processing: Image Communication*, vol. 27, no. 6, pp. 579-594, February 2012.
- (127) L. Neumeyer, B. Robbins, A. Nair, et al: "S4: distributed stream computing platform". In 2010 IEEE International Conference on Data Mining Workshops, pp. 170-177, IEEE (2010).
- (128) P. Ndjiki-Nya, G. Simo, T. Wiegand: Evaluation of Color Image Segmentation Algorithms Based on Histogram Thresholding, *VLBV - Lecture Notes in Computer Science*, pp. 214-222, 2006.
- (129) The "Materials Genome Initiative for Global Competitiveness" is a science and engineering initiative by the President of the United States of America, issued June 24 2011: http://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf.
- (130) A. Reinefeld, F. Schintke, T. Schütt, S. Haridi: A Scalable, Transactional Data Store for Future Internet Services. Towards the Future Internet - A European Research Perspective, G. Tselentis, J. Domingue, A. Galis, A. Gavras, D. Hausheer, S. Krco, V. Lotz, T. Zahariadis (Eds.), pp. 148-159, IOS Press, 2009.
- (131) G. Rätsch, S. Sonnenburg, J. Srinivasan, H. Witte, R. Sommer, K. R. Müller, B. Schölkopf: Improving the *C. elegans* genome annotation using machine learning, 2007, *PLoS Computational Biology*, 3:e20.
- (132) K. Reuter, C. Stampfl, M. Scheffler: Ab initio atomistic thermodynamics and statistical mechanics of surface properties and functions. In: *Handbook of Materials Modeling*, Vol. 1. (Ed.) Sidney Yip. Springer Berlin Heidelberg 2005, 149-194. ISBN 1-4020-3287-0.
- (133) M. Rupp, A. Tkatchenko, K. R. Müller, O. A. v. Lilienfeld: Fast and accurate modeling of molecular atomization energies with machine learning, *Physical Review Letters*, 108(5), 058301, 2012.
- (134) Storm: Distributed and fault-tolerant realtime computation: <http://storm-project.net/> (last accessed 2013-06-10).

- (135) D. Simmen, M. Altinel, V. Markl, S. Padmanabhan, A. Singh: DAMIA: Data Mashups for Intranet Applications, In Proc. ACM SIGMOD Conference, pp. 1171-1182 (2008).
- (136) J. Stender, M. Berlin, A. Reinefeld: XtreemFS – A File System for the Cloud. Data Intensive Storage Services for Cloud Environments, D. Kyriazis, A. Voulodimos, S.V. Gogouritis, T. Varvarigou (Eds.), IGI Global, 2013.
- (137) Newsfocus editorial, Materials scientists look to a data-intensive future. SCIENCE 335, 1434 (2012).
- (138) R. Sherwood, G. Gibb, KK Yap, G. Appenzeller, M. Casado, N. McKeown, G. Parulkar: Can the Production Network Be the Testbed? USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2010.
- (139) B. Sigelmann, L. A. Barroso, M. Burrows, P. Stephenson, M. Plakal, D. Beaver, S. Jaspan, C. Shanbhag: Dapper, a Large-Scale Distributed Systems Tracing Infrastructure, Google Technical Report, April 2010.
- (140) H. Seim, D. Kainmüller, H. Lamecker, Matthias Bindernagel, Jana Malinowski, Ch. Schütte, S. Zachow: Model-based Auto-Segmentation of Knee Bones and Cartilage in MRI Data Proc. Medical Image Analysis for the Clinic: A Grand Challenge (B. v. Ginneken et al., eds.), 2010.
- (141) F. Schintke, A. Reinefeld, S. Haridi, T. Schütt: Enhanced Paxos Commit for Transactions on DHTs. 10th IEEE/ACM Int. Conf. on Cluster, Cloud and Grid Computing (CCGrid2010), pp. 448-454, May 2010.
- (142) K. Sobe: Consistency and Fault Tolerance of Distributed Storage Systems. Doctoral thesis, HU-Berlin, 2013.
- (143) G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand: Overview of the High Efficiency Video Coding (HEVC) Standard, IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, December 2012.
- (144) Shi, Q., Petterson, J., Dror, G., Langford, J., Smola, A., S. V. N. Vishnawathan: Hash Kernels for Structured Data, Journal of Machine Learning Research, 10(Nov):2615–2637, 2009.
- (145) T. Schütt, F. Schintke, A. Reinefeld: A Structured Overlay for Multi-dimensional Range Queries. Euro-Par 2007, Parallel Processing, 13th International Euro-Par Conference, Rennes, France, August 2007.
- (146) Stratosphere Projekt: <https://www.stratosphere.eu/>. (letzter Zugriff: 12.6.2013).
- (147) A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain: "Content-based image retrieval at the end of the early years," IEEE Trans. on PAMI, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

- (148) J. Tan, S. Kavulya, R. Gandhi, P. Narasimhan: Visual, Log-Based Causal Tracing for Performance Debugging of MapReduce Systems, 2010 IEEE 30th International Conference on Distributed Computing Systems (ICDCS), p. 795-806.
- (149) TrekVid, <http://www-nplir.nist.gov/projects/tv2013/tv2013.html#sin>, letzter Zugriff, 7.7.2013
- (150) Y. Tian, S. Tatikonda, B. Reinwald: Scalable and Numerically Stable Descriptive Statistics in SystemML. ICDE 2012: 1351-1359
- (151) C. H. Teo: S.V.N. Vishwanathan, A. Smola, Q. V. Le: Bundle Methods for Regularized Risk Minimization, 2010, JMLR.
- (152) M. Vrhovnik, H. Schwarz, O. Suhre, B. Mitschang, V. Markl, A. Maier, T. Kraft: An Approach to Optimize Data Processing in Business Processes. VLDB 2007: 615-626
- (153) I. H. Witten, E. Frank, M. A. Hall: Data Mining: Practical Machine Learning Tools and Techniques (3 ed.). Elsevier. ISBN 978-0-12-374856-0. 2011.
- (154) T. White: Hadoop - The Definitive Guide: Storage and Analysis at Internet Scale (3. ed., revised and updated). O'Reilly, (2012).
- (155) D. Warneke, O. Kao: Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud, Parallel and Distributed Systems, IEEE Transactions on. IEEE Press, 985-997. 2011.
- (156) D. Warneke, O. Kao: Nephele: Efficient Parallel Data Processing in the Cloud, MTAGS '09: Proceedings of the 2nd Workshop on Many-Task Computing on Grids and Supercomputers. ACM, 1–10. 2009.
- (157) H.-E. Wichmann, R. Kaaks, W. Hoffmann, K.-H. Jöckel, K. H. Greiser, J. Linseisen. Die Nationale Kohorte Bundesgesundheitsblatt 55 (2012), pp. 781–789. DOI: 10.1007/s00103-012-1499-y.
- (158) A. Wundsam, D. Levin, S. Seetharaman, A. Feldmann. OFRewind: Enabling Record and Replay Troubleshooting for Networks, Usenix Annual Technical Conference (ATC), 2011.
- (159) T. Wiegand, G. J. Sullivan, G. Bjontegaard, A. Luthra: Overview of the H.264/AVC Video Coding Standard, IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, July 2003.
- (160) M. Zaharia, M. Chowdhury, T. Das, et al: "Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing". In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, USENIX (2012)
- (161) M. Zaharia, T. Das, H. Li, et al: Discretized Streams: "An Efficient and Fault-Tolerant Model for Stream Processing on Large Clusters". In Proceedings of the 4th USENIX conference on Hot Topics in Cloud Computing, USENIX (2012).

- (162) S. Zachow, M. Zilske, H.-C. Hege, (2007): "3D Reconstruction of Individual Anatomy from Medical Image Data: Segmentation and Geometry Processing." Proc. of 25th ANSYS Conference & CADFEM Users' Meeting 2007, Dresden.
- (163) Rodgers, David P. "Improvements in multiprocessor system design". ACM SIGARCH Computer Architecture News archive (New York, NY, USA: ACM) 13 (3): 225–231 (1985).
- (164) Louis Woods, Jens Teubner, Gustavo Alonso: Less watts, more performance: an intelligent storage engine for data appliances. SIGMOD Conference 2013: 1073-1076
- (165) Louis Woods, Jens Teubner, Gustavo Alonso: Complex Event Detection at Wire Speed with FPGAs. PVLDB 3(1): 660-669 (2010)
- (166) Peter M. Fischer, Jens Teubner: MXQuery with Hardware Acceleration. ICDE 2012: 1293-1296
- (167) Rodgers, David P. "Improvements in multiprocessor system design". ACM SIGARCH Computer Architecture News archive (New York, NY, USA: ACM) 13 (3): 225–231 (1985).
- (168) René Müller, Jens Teubner: FPGAs: a new point in the database design space. EDBT 2010: 721-723
- (169) Ian Buck, Tim Foley, Daniel Reiter Horn, Jeremy Sugerman, Kayvon Fatahalian, Mike Houston, Pat Hanrahan: Brook for GPUs: stream computing on graphics hardware. ACM Trans. Graph. 23(3): 777-786 (2004)
- (170) John Nickolls, Ian Buck, Michael Garland, Kevin Skadron: Scalable Parallel Programming with CUDA. ACM Queue 6(2): 40-53 (2008)
- (171) Juchang Lee, Michael Muehle, Norman May, Franz Faerber, Vishal Sikka, Hasso Plattner, Jens Krueger, Martin Grund: High-Performance Transaction Processing in SAP HANA. IEEE Data Eng. Bull. 36(2): 28-33 (2013)
- (172) Fabian Schomm, Florian Stahl, Gottfried Vossen: Marketplaces for data: an initial survey. ACM SIGMOD Record 42(1): 15-26 (2013)
- (173) A.Muschalle, F.Stahl, A. Löser, G. Vossen: Pricing Approaches for Data Markets. 6th. Business Intelligence for the Real Time Enterprise, Springer LNBIP 154, 2013, 129-144
- (174) Andreas Gadatsch: Big Data. WISU Das Wirtschaftsstudium 41(12): 1615-1621 (2012)
- (175) Christy Pettey, Rob van der Meulen: Gartner Reveals Top Predictions for IT Organizations and Users for 2012 and Beyond. <http://www.gartner.com/newsroom/id/1862714/> (zuletzt aufgerufen am 02.08.2013)
- (176) TATA Consultancy Service: The Emerging Big Returns on Big Data: A TCS 2013 global Trend Study 2013
- (177) Mathias Weber: Big Data im Praxiseinsatz - Szenarien, Beispiele, Effekte 2012

- (178) Ohne Verfasser: Big Data: What information, if you had it, would change the way you run your business? <http://www.gartner.com/technology/topics/big-data.jsp>. (zuletzt aufgerufen am 12.07.2013)
- (179) Helmut Krcmar: Informationsmanagement (5. Aufl.). Heidelberg, Dordrecht, London, New York: Springer 2009, ISBN 978-540-23015-1
- (180) Carsten Bange, Timm Grosser, Nikolai Janoschek: Big Data Survey Europe: Nutzung, Technologie und Budgets europäischer Best Practice Unternehmen. 2013.
- (181) Hsinchun Chen, Roger H. L. Chiang, Veda C. Storey: Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly 36(4): 1165-1188 (2012)
- (182) Maik Manig, Jens Giere: Quo vadis Big Data: Herausforderungen - Erfahrungen - Lösungsansätze. 2013.
- (183) Wolfgang Eggert, Steffen Minter: Stichwort: öffentlicher Sektor. Gabler Wirtschaftslexikon <http://wirtschaftslexikon.gabler.de/Archiv/14147/oefentlicher-sektor-v10.html> (zuletzt aufgerufen am 12.07.2013)
- (184) Michael Batty, Kay W. Axhausen, Fosca Giannotti, Alexei Bazzani, Armando Wachowicz, Georgios Ouzounis, Yuval Portugali: Smart cities of the future. The European Physical Journal Special Topics 214(1): 481-518 (2012)
- (185) Bundeswahlgesetz (BWG): Bundeswahlleiter, Editor 2013: BGBI. I 1288, 1594.
- (186) Bundeswahlordnung (BWO): Bundeswahlleiter, Editor 2013: BGBI. I 1255.
- (187) in 2 BvC 3/07, 2 BvC 4/07, Bundesverfassungsgericht, Editor 2009.
- (188) Randal E. Bryant, Randy H. Katz, Edward D. Lazowska: Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society. 2008
- (189) Eva Geisberger: agendaCPS: Integrierte Forschungsagenda Cyber-Physical Systems (acatech Studie) 2012
- (190) Henning Kagermann, Johannes Helbig, Wolfgang Wahlster: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0 - Abschlussbericht des Arbeitskreises Industrie 4.0. 2013, Berlin. 2013
- (191) Gerhard Kleinhenz, Martin Werding: Gesundheitswesen. 2013 <http://wirtschaftslexikon.gabler.de/Archiv/55801/gesundheitswesen-v8.html> (zuletzt aufgerufen am 18.07.2013)
- (192) Pavlo Baron: Big Data für IT-Entscheider: riesige Datenmengen und moderne Technologien gewinnbringend nutzen. München: Hanser.2013
- (193) TomTom Traffic: http://www.tomtom.com/de_de/services/traffic/ (zuletzt aufgerufen am 15.07.2013)
- (194) Alexander Strobel: Schneller ankommen, weniger verbrauchen - das versprechen zeitgemäße Navi-Systeme. Naviconnect 2: 12-17 (2009)

- (195) Bernhard Heißing, Metin. Ersoy, Stefan Gies: Zukunftsaspekte des Fahrwerks. Fahrwerkhandbuch: 639-676.
- (196) W Loose: Jahresbericht 2012/2013. 2013.
- (197) Gerd Wörn: Energie durch Big Data: sind Europas Energieversorger für das Datenzeitalter gerüstet?. T-Systems International GmbH.2013
- (198) Matthias Gottlieb, Wolfgang Wörndl: Analysis of Predictions and Clusters for a Sustainable Energy Monitoring Model. Proceeding of 3rd Colloquium of the Munich School of Engineering: "Research Towards Innovative Energy Systems and Material", München, 2013
- (199) T-Systems: Neue Energien aus der Wolke. 2013 <http://www.t-systems.de/news-media/big-data-bietet-energiebranche-grosse-chancen/913788/> (zuletzt aufgerufen am 12.07.2013)
- (200) Philip Russom: Big Data Analytics. tdwi Research. 2011
- (201) Hans R. Hansen, Gustaf Neumann: Wirtschaftsinformatik 1: Grundlagen und Anwendungen. Stuttgart. Lucius & lucius: 2009.
- (202) Brad Brown, Michael Chui, Jame Manyika: Are you ready for the era of 'big data'? McKinsey Quarterly 4: 24-35 (2011)
- (203) Christian Hofmeister, Michael Ehrmanntraut: Veränderungen im Geschäftsmodell von Versicherungen durch Big Data. IBM Information Management Forum. 2013
- (204) Günter Neumann: Informationsextraktion. Carstensen et al. (eds): Computerlinguistik und Sprachtechnologie - Eine Einführung, Heidelberg: Spektrum-Verlag, ISBN 3827414075, second edition., 2004.
- (205) K. Fukuda, T. Tsunoda, A. Tamura und T. Takagi: Toward Information Extraction: Identifying protein names from biological papers, Proceedings of the Pacific Symposium on Biocomputing, Hawaii, pp. 707–718, 1998.
- (206) Yunyao Li, Rajasekar Krishnamurthy, Sriram Raghavan, Shivakumar Vaithyanathan, and H. V. Jagadish: Regular expression learning for information extraction. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '08). Association for Computational Linguistics, Stroudsburg, PA, USA, 21-30, 2008.
- (207) Tomoko Ohta, Yuka Tateisi, and Jin-Dong Kim: The GENIA corpus: an annotated research abstract corpus in molecular biology domain. In Proceedings of the second international conference on Human Language Technology Research (HLT '02). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 82-86, 2002.

- (208) Hristo Tanev, Jakub Piskorski, and Martin Atkinson. Real-Time News Event Extraction for Global Crisis Monitoring. In Proceedings of the 13th international conference on Natural Language and Information Systems: Applications of Natural Language to Information Systems (NLDB '08), Epaminondas Kapetanios, Vijayan Sugumaran, and Myra Spiliopoulou (Eds.). Springer-Verlag, Berlin, Heidelberg, 207-218, 2008.
- (209) Anthony Fader, Stephen Soderland, and Oren Etzioni. Identifying relations for open information extraction. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11). Association for Computational Linguistics, Stroudsburg, PA, USA, 1535-1545, 2011.
- (210) Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S. Weld: Open information extraction from the web. Commun. ACM 51, 12 (December 2008), 68-74, 2008.
- (211) Oren Etzioni, Anthony Fader, Janara Christensen, Stephen Soderland, and Mausam Mausam: Open information extraction: the second generation. In Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume One (IJCAI'11), Toby Walsh (Ed.), Vol. Volume One. AAAI Press 3-10, 2011.
- (212) Wei Wang, Romaric Besançon, Olivier Ferret, and Brigitte Grau: Filtering and clustering relations for unsupervised information extraction in open domain. In Proceedings of the 20th ACM international conference on Information and knowledge management (CIKM '11), Bettina Berendt, Arjen de Vries, Wenfei Fan, Craig Macdonald, Iadh Ounis, and Ian Ruthven (Eds.). ACM, New York, NY, USA, 1405-1414, 2011.
- (213) Mausam, Michael Schmitz, Stephen Soderland, Robert Bart und Oren Etzioni: Open Language Learning for Information Extraction, EMNLP-CoNLL 2012, 523-534, 2012.
- (214) Jakub Piskorski und Roman Yangarber: Information Extraction – Past, Present and Future. In: T. Poibeau et al. (eds.) Multi-source, Multilingual Information Extraction and Summarization 11, Theory and Applications of Natural Language Processing, Springer-Verlag, Berlin, Heidelberg, 23-49, 2013.
- (215) Angel X. Chang und Christopher D. Manning: SUTIME: A Library for Recognizing and Normalizing Time Expressions. 8th International Conference on Language Resources and Evaluation (LREC), 2012
- (216) Angel X. Chang und Christopher D. Manning: SUTime -- Evaluation in TempEval-3. Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2, Proceedings of the Seventh International Workshop on Semantic Evaluation, ACL, 78—82, 2013.
- (217) Robert Spence: Information Visualization -- Design for Interaction (2nd Edition), Prentice Hall, 2006.

- (218) Matthew Ward, Georges G. Grinstein und Daniel Keim: Interactive Data Visualization: Foundations, Techniques, and Applications Taylor & Francis Ltd., 2010.
- (219) Ben Shneiderman: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, In: Proceedings of Visual Languages, 1996.
- (220) Harold Lim, Herodotos Herodotou, Shivnath Babu: Stubby: a transformation-based optimizer for MapReduce workflows, In Proceedings of the VLDB Endowment 5(11), 2012, pp. 1196-1207.
- (221) Fabian Hueske, Mathias Peters, Matthias Sax, Astrid Rheinländer, Rico Bergmann, Aljoscha Krettek, Kostas Tzoumas: Opening the black boxes in data flow optimization, In Proceedings of the VLDB Endowment, 5(11), 2012, pp. 1256-1267.
- (222) Micheal J. Cafarella, Christopher Ré: Manimal: relational optimization for data-intensive programs, In Proceedings of the 13th International Workshop on the Web and Databases. ACM, 2010.
- (223) Futuristic Intel Chip Could Reshape How Computers are Built, Consumers Interact with Their PCs and Personal Devices , http://www.intel.com/pressroom/archive/releases/2009/20091202comp_sm.htm#story (Letzter Aufruf 08.09.2013)
- (224) Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, Jonathan Turner: OpenFlow: enabling innovation in campus networks, In SIGCOMM, April 2008, Vol 38, No. 2, pp. 69—74
- (225) D. J. DeWitt, R. H. Gerber, G. Graefe, M. Heytens, K. Kumar, M. Muralikrishna: Gamma - A High Performance Dataflow Database Machine, TR #635, März 1986, Computer Science Department, University of Wisconsin.
- (226) Teradata, <http://www.teradata.de> (Letzter Aufruf 08.09.2013)
- (227) Sanjay Ghemawat, Howard Gobioff, Shun-Tak Leung: The Google file system, In ACM SIGOPS Operating Systems Review. Vol. 37. No. 5. ACM, 2003.
- (228) Apache Pig, <http://pig.apache.org/> (Letzter Aufruf 08.09.2013)
- (229) Apache Hive, <http://hive.apache.org/> (Letzter Aufruf 08.09.2013)
- (230) Apache Mahout, <http://mahout.apache.org/> (Letzter Aufruf 08.09.2013)
- (231) Apache Giraph, <http://giraph.apache.org/> (Letzter Aufruf 08.09.2013)
- (232) Apache HBase, <http://hbase.apache.org/> (Letzter Aufruf 08.09.2013)
- (233) Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Mike Burrows, Tushar Chandra, Andrew Fikes, Robert E. Gruber: Bigtable: A distributed storage system for structured data, In ACM Transactions on Computer Systems (TOCS) 26.2, 2008.
- (234) Apache Oozie, <http://oozie.apache.org/> (Letzter Aufruf 08.09.2013)

- (235) M. Stonebraker, D. Abadi, D. J. DeWitt, S. Madden, E. Paulson, A. Pavlo, A. Rasin: MapReduce and parallel DBMSs: friends or foes?, In Communications of the ACM, 53(1), 64-71, 2010.
- (236) R. Chaiken, B. Jenkins, P. Å. Larson, B. Ramsey, D. Shakib, S. Weaver, J. Zhou: SCOPE: easy and efficient parallel processing of massive data sets, In Proceedings of the VLDB Endowment, 1(2), 1265-1276., 2008.
- (237) Biswajesh Chattopadhyay, Liang Lin, Weiran Liu, Sagar Mittal, Prathyusha Aragonda, Vera Lychagina, Younghoo Kwon, Michael Wong: Tenzing A SQL Implementation On The MapReduce Framework. In PVLDB 4(12):1318-1327, 2011.
- (238) J. Dittrich, J. A. Quiané-Ruiz, A. Jindal, Y. Kargin, V. Setty, J. Schad: Hadoop++: Making a yellow elephant run like a cheetah (without it even noticing), In Proceedings of the VLDB Endowment, 3(1-2), 515-529, 2010.
- (239) S. Richter, J. A. Quiane-Ruiz, S. Schuh, J. Dittrich: Towards Zero-Overhead Static and Adaptive Indexing in Hadoop, In VLDB Journal, 2013.
- (240) M. Y. Eltabakh, Y. Tian, F. Özcan, R. Gemulla, A. Krettek, J. McPherson: CoHadoop: flexible data placement and its exploitation in Hadoop, In Proceedings of the VLDB Endowment, 4(9), 575-585, 2011.
- (241) Avrilia Floratou, Jignesh M. Patel, Eugene J. Shekita, Sandeep Tata: Column-Oriented Storage Techniques for MapReduce, In PVLDB 4(7):419-429, 2011.
- (242) Hadapt, <http://hadapt.com/> (Letzter Aufruf 08.09.2013)
- (243) A. Abouzeid, K. Bajda-Pawlikowski, D. Abadi, A. Silberschatz, A. Rasin: HadoopDB: an architectural hybrid of MapReduce and DBMS technologies for analytical workloads, In Proceedings of the VLDB Endowment, 2(1), 922-933, 2009.
- (244) David J. DeWitt, Alan Halverson, Rimma V. Nehme, Srinath Shankar, Josep Aguilar-Saborit, Artin Avanes, Miro Flasza, Jim Gramling: Split query processing in polybase, In SIGMOD 2013, pp. 1255-1266.
- (245) IBM InfoSphere BigInsights, <http://www-01.ibm.com/software/data/infosphere/biginsights/> (Letzter Aufruf 08.09.2013)
- (246) S. Melnik, A. Gubarev, J. J. Long, G. Romer, S. Shivakumar, M. Tolton, T. Vassilakis: Dremel: interactive analysis of web-scale datasets, In Proceedings of the VLDB Endowment, 3(1-2), 330-339, 2010.
- (247) Marcel Kornacker, Justin Erickson: Cloudera Impala: Real-Time Queries in Apache Hadoop, For Real, <http://blog.cloudera.com/blog/2012/10/cloudera-impala-real-time-queries-in-apache-hadoop-for-real/> (Letzter Aufruf 08.09.2013)
- (248) Apache Incubator, <http://incubator.apache.org/drill/> (Letzter Aufruf 08.09.2013)

- (249) MADlib: an open-source library for scalable in-database analytics, <http://madlib.net/> (Letzter Aufruf 08.09.2013)
- (250) PostgreSQL, <http://www.postgresql.org/> (Letzter Aufruf 08.09.2013)
- (251) Pivotal Greenplum Database, <http://www.gopivotal.com/pivotal-products/data/pivotal-analytic-database> (Letzter Aufruf 08.09.2013)
- (252) EXASolution: ein relationales In-Memory-Datenbankmanagementsystem, <http://wwwexasol.com/exasolution.html> (Letzter Aufruf 08.09.2013)
- (253) R Statistik-Software-Projekt, <http://www.r-project.org/> (Letzter Aufruf 08.09.2013)
- (254) Apache Spark, <http://spark.incubator.apache.org/> (Letzter Aufruf 08.09.2013)
- (255) Microsoft SQL Server, <https://www.microsoft.com/en-us/sqlserver/solutions-technologies/business-intelligence/streaming-data.aspx> (Letzter Aufruf 08.09.2013)
- (256) IBM Stream Computing, <http://www-01.ibm.com/software/data/infosphere/stream-computing/> (Letzter Aufruf 08.09.2013)
- (257) Storm, distributed and fault-tolerant realtime computation, <http://storm-project.net/> (Letzter Aufruf 08.09.2013)
- (258) Arvind Arasu, Brian Babcock, Shivnath Babu, Mayur Datar, Keith Ito, Itaru Nishizawa, Justin Rosenstein, Jennifer Widom: STREAM: the stanford stream data manager, In Proceedings of the 2003 ACM SIGMOD, pp. 665—665.
- (259) Arvind Arasu, Shivnath Babu, Jennifer Widom: The CQL continuous query language: semantic foundations and query execution. In The VLDB Journal 15.2 (2006): 121-142.
- (260) AsterixDB Big Data Management System, <http://asterix.ics.uci.edu/> (Letzter Aufruf 09.09.2013)
- (261) HyPer: A Hybrid OLTP&OLAP High-Performance Database System, <http://www.hyper-db.com/> (Letzter Aufruf 09.09.2013)
- (262) HP Vertica, <http://www.vertica.com/> (Letzter Aufruf 09.09.2013)
- (263) ParStream: Real-time Big Data Analytics Platform, <http://www.parstream.com/> (Letzter Aufruf 09.09.2013)
- (264) Datahub: the free, powerful data management platform, <http://datahub.io/about> (Letzter Aufruf 10.09.2013)
- (265) DBpedia: crowd-sourced community effort to extract structured information from Wikipedia <http://dbpedia.org/About> (Letzter Aufruf 10.09.2013)
- (266) Freebase: community-curated database of people, places, and things <http://www.freebase.com/> (Letzter Aufruf 10.09.2013)
- (267) Wikidata: a free knowledge base, <http://meta.wikimedia.org/wiki/Wikidata> (Letzter Aufruf 10.09.2013)

- (268) EMC: Cloud Computing erleichtert Umgang mit steigenden Informationsmengen, <http://germany.emc.com/about/news/press/2010/20100504-01.htm>
- (269) BioCreative II:- Critical Assessment for Information Extraction in Biology challenge (2006-2007), http://biocreative.sourceforge.net/biocreative_2.html (Letzter Aufruf 10.09.2013)
- (270) BioNLP Shared Tasks, <http://2013.bionlp-st.org/tasks> (Letzter Aufruf 10.09.2013)
- (271) BioCreative IV: Critical Assessment of Information Extraction in Biology, <http://www.biocreative.org/events/biocreative-iv/CFP/> (Letzter Aufruf 10.09.2013)
- (272) Andrew McCallum: Joint Inference for Natural Language Processing. Proceedings of the Thirteenth Conference on Computational Natural Language Learning (2012)
- (273) Hoifung Poon, Pedro Domingos: Joint Inference in Information Extraction. AAAI 2007: 913-918
- (274) Bishan Yang, Claire Cardie: Joint Inference for Fine-grained Opinion Extraction. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2013: 1640—1649
- (275) Raymond Yiu Keung Lau, Chun Lam Lai, Peter B. Bruza, Kam F. Wong: Leveraging web 2.0 data for scalable semi-supervised learning of domain-specific sentiment lexicons. In Proceedings of the 20th ACM international conference on Information and knowledge management (CIKM '11), ACM, New York, NY, USA, 2011: 2457-2460.
- (276) Sameer Singh, Michael Wick, Andrew McCallum: Monte Carlo MCMC: Efficient Inference by Approximate Sampling. In Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012: 1104--1113
- (277) <https://sites.google.com/site/scanlp2013/> (Letzter Aufruf 10.09.2013)
- (278) Kathrin Eichler, Holmer Hemsen, Günter Neumann: Unsupervised Relation Extraction From Web Documents. LREC 2008: 1674-1679
- (279) Jannik Strötgen, Michael Gertz: HeidelTime: High Quality Rule-based Extraction and Normalization of Temporal Expressions. In: SemEval'10, 321-324, 2010
- (280) Katrin Tomanek: Resource-aware annotation through active learning. Dissertation, TU-Dortmund 2010
- (281) Tim Polzehl, Alexander Schmitt, Florian Metze, Michael Wagner: Anger recognition in speech using acoustic and linguistic cues. Speech Communication 53(9-10): 1198-1209 (2011)

- (282) D. Abadi, D. Carney, U. Cetintemel, M. Cherniack, C. Convey, C. Erwin, E. Galvez, M. Hatoun, J. Hwang, A. Maskey, A. Rasin, A. Singer, M. Stonebraker, N. Tatbul, Y. Xing, R. Yan, S. Zdonik: Aurora: a data stream management system. In Proceedings of the 2003 ACM SIGMOD: 666-666
- (283) Daniel J. Abadi, Yanif Ahmad, Magdalena Balazinska, Ugur Cetintemel, Mitch Cherniack, Jeong-Hyon Hwang, Wolfgang Lindner, Anurag S. Maskey, Alexander Rasin, Esther Ryvkina, Nesime Tatbul, Ying Xing, Stan Zdonik: The Design of the Borealis Stream Processing Engine. In CIDR. Vol. 5. 2005.
- (284) BMBF, <http://www.bmbf.de/de/19955.php>
- (285) Challenges and Opportunities with Big Data: <http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>
- (286) OBAMA ADMINISTRATION UNVEILS “BIG DATA” INITIATIVE: http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf
- (287) UK Prime Minister joins Sir Ka-shing Li for launch of £90m initiative in big data and drug discovery at Oxford, http://www.ox.ac.uk/media/news_stories/2013/130503.html
- (288) Japan Looks to Big Data for Timely Economic Indicator, <http://blogs.wsj.com/japanrealtime/2013/09/24/japan-looks-to-big-data-for-timely-economic-indicator/>
- (289) Christensen, C. (1997): The Innovator’s Dilemma. HarperCollins, New York 1997.
- (290) Ying Xu, Mi-Young Kim, Kevin Quinn, Randy Goebel, and Denilson Barbosa: Open Information Extraction with Tree Kernels. In NAACL-HLT, 2013.
- (291) Paul McNamee abd Hoa Trang Dang: Overview of the TAC 2009 Knowledge Base Population Track. In TAC, 2009.
- (292) Matthew W. Bilotti, Paul Ogilvie, Jamie Callan, and Eric Nyberg. Structured retrieval for question answering. In SIGIR, 2007.
- (293) Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky: Distant Supervision for relation extraction without labeled data. In ACL-IJCNLP, 2009.
- (294) Katsumasa Yoshikawa, Sebastian Riedel, Masayuki Asahara, Yuji Matsumoto: Jointly Identifying Temporal Relations with Markov Logic. In ACL-IJCNLP, 2009.
- (295) Yang Song, Jing Jiang, Wayne, Xin Zhao, Sujian Li, and Houfeng Wang: Joint Learning for Coreference Resolution with Markov Logic. In EMNLP, 2012.
- (296) Matthew Richardson and Pedro Domingos: Markov Logic Networks. Machine Learning 62 (1-2): 107-136, 2006.

- (297) Benjamin Roth, Tassilo Barth, Michael Wiegand, and Dietrich Klakow: A Survey of Noise Reduction Methods for Distant Supervision. In CIKM-AKBC Workshop, 2013.
- (298) Robert Leaman, Laura Wojtulewicz, Ryan Sullivan, Annie Skariah, Jian Yang, and Graciela Gonzalez: Towards Internet-Age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts in Health-Related Social Networks. In BioNLP, 2010.
- (299) Sarvnas Karimi, SuNam Kim, Lawrence Cavedon: Drug Side Effects: What Do Patient Forums Reveal? In CIKM-MedEx-Workshop, 2011.
- (300) Hariprasad Sampathkumar, Bo Luo, and Xue-wen Chen: Mining Adverse Drug Side-Effects from Online Medical Forums. In HISB, 2012.
- (301) Steven Abney: Semisupervised Learning for Computational Linguistics. Chapman & Hall/CRC Computer Science & Data Analysis Series 8. CRC Press, 2007.
- (302) Stanford University, <https://bigdata.stanford.edu/>
- (303) United Nations, Report of the World Commission on Environment and Development: Our Common Future, 1987.
- (304) AnHai Doan, Alon Y. Halevy, Zachary G. Ives: Principles of Data Integration. Morgan Kaufmann 2012.
- (305) Felix Naumann, Melanie Herschel: An Introduction to Duplicate Detection. Synthesis Lectures on Data Management, Morgan & Claypool Publishers 2010.
- (306) Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze: Introduction to information retrieval. Cambridge University Press 2008.

9 Abbildungsverzeichnis

Abbildung 1: Zusammenspiel der Rahmenbedingungen	9
Abbildung 2: Bekanntheit des Begriffs Big Data in der deutschen Bevölkerung in Prozent (n=1008, BITKOM 2013)	20
Abbildung 3: Zustimmung zu verschiedenen Beschreibungen des Big Data Begriffs in Prozent (n=79, Experton Group 2012)	20
Abbildung 4: Einstellung befragter Unternehmen zu Big Data in Prozent (n=n.a., computing research 2013)	21
Abbildung 5: Ziele von Big Data Anwendungen in Anzahl Nennungen (n=100, Fraunhofer IAIS 2013).....	24
Abbildung 6: Wichtigste Vorteile einer Nutzung von Big Data-Technologien (n=167, BARC 2013)	25
Abbildung 7: Positive Auswirkungen von Big Data (n=100, Experton Group 2012)	26
Abbildung 8: Zustimmung zu verschiedenen Aussagen über die Bedeutung von Big Daten in Prozent (n=1008, BITKOM 2013)	27
Abbildung 9: Zeithorizont und Priorität von Big Data in Unternehmen in Prozent (n=750, interxion 2013).....	27
Abbildung 10: Priorität von Big Data innerhalb der nächsten drei Jahre in Prozent (n=750, interxion 2013)	28
Abbildung 11: Handlungsbedarf bei Big Data nach Branche auf einer Skala von 1 bis 5 (n=1010, TNS Infratest 2012).....	30
Abbildung 12: Durchschnittliche erwartete Rendite von Big Data Investition nach Branchen im Jahre 2012 in Prozent (n=643, TCS 2013).....	31

Abbildung 13: Unternehmensaktivitäten und deren Potenzial (n=85, Experton Group 2012).....	32
Abbildung 14: Durchschnittliche erwartete Rendite von Big Data Investition nach Funktionsbereichen im Jahre 2012 in Prozent (n=643, TCS 2013).....	34
Abbildung 15: Einsatz verschiedener Datenarten für Big Data Analysen (n=274, BARC 2013).....	35
Abbildung 16: Vorhandensein einer Strategie für Big Data (n=274, BARC 2013).....	35
Abbildung 17: Beschäftigung mit Big Data (n=100, Experton Group 2012)	36
Abbildung 18: Veränderung des Ausmaßes der Beschäftigung mit Big Data 2012 bis 2013 (n=n.a., computing research 2013)	37
Abbildung 19: Nutzung von Big Data-Werkzeugen nach Unternehmensgröße (n=132, BARC 2013)	37
Abbildung 20: Herausforderungen beim Einsatz von Big Data (n=206, BARC 2013)	39
Abbildung 21: Technische Hindernisse bei der Umsetzung einer Big Data Strategie in Prozent (n=n.a., Computing Research)	41
Abbildung 22: Entwicklung des globalen Big Data Marktes 2011-2016 in Mio. € (Experton Group 2012).....	42
Abbildung 23: Deutscher Big Data Markt 2011-2016 nach Marktsegmenten in Mio. € (Experton Group 2012)	42
Abbildung 24: Durchschnittliche Ausgaben eines Unternehmens für Big Data nach Ländern in den Jahren 2012 und 2015 (prognostiziert) in Mio. USD (n=643, TCS 2013)	43
Abbildung 25: Tag-Cloud zum Thema Big Data (Eigene Darstellung)	49
Abbildung 26: Wichtigkeit des Themas Big Data in der Praxis (Eigene Darstellung)	50
Abbildung 27: Wertschöpfungspotential von Big Data (Eigene Darstellung).....	51

Abbildung 28: Messbarer Wertbeitrag von Big Data (Eigene Darstellung).....	52
Abbildung 29: Geschäftsmodelle von Big Data-Anbietern, Mehrfachnennungen möglich, n=32 (Eigene Darstellung)	52
Abbildung 30: Charakteristika von Big Data 1 (Eigene Darstellung)	53
Abbildung 31: Charakteristika von Big Data 2 (Eigene Darstellung)	54
Abbildung 32: Geschäftspotentiale von Big Data (Eigene Darstellung)	55
Abbildung 33: Big Data und unternehmerische Entscheidungen 1 (Eigene Darstellung)	56
Abbildung 34: Big Data und unternehmerische Entscheidungen 2 (Eigene Darstellung)	57
Abbildung 35: Wertschöpfungspotential für betriebliche Funktionen (Eigene Darstellung)	58
Abbildung 36: Die Rolle der IT im Big Data-Kontext 1 (Eigene Darstellung).....	59
Abbildung 37: Die Rolle der IT im Big Data-Kontext 2 (Eigene Darstellung).....	60
Abbildung 38: Einschätzung des Wertes von Unternehmensdaten (Eigene Darstellung)	61
Abbildung 39: Status von Big Data-Projekten (Eigene Darstellung)	62
Abbildung 40: Wann wird Big Data wettbewerbsentscheidend (Eigene Darstellung)	62
Abbildung 41: Angebot an Big Data Werkzeugen anhand der verarbeiteten Datentypen (Anbieterperspektive)	63
Abbildung 42: Potenzial für Big Data Werkzeuge aus Sicht der Anwender	64
Abbildung 43: Hürden und Herausforderungen im Zusammenhang mit Big Data 1 (Eigene Darstellung)	66
Abbildung 44: Hürden und Herausforderungen im Zusammenhang mit Big Data 2 (Eigene Darstellung)	67

Abbildung 45: Branchenaufteilung nach der Studie des BARC-Instituts (In Anlehnung an (180, S. 9).	71
Abbildung 46: Aufteilung BITKOM Fallbeispiele auf ausgewählte Branchen (n=34)	74
Abbildung 47: Business Model Canvas zur BITKOM Studie (In Anlehnung an Osterwald, Pigneur (2010)).....	95
Abbildung 48: Handlungsdruck nach Branchen im Bereich Big Data (Quelle: TNS Infratest(182) (2012)).....	96
Abbildung 49: Vorher-Nachher-Vergleich der Dimensionen Datenvielfalt zu Verarbeitungsgeschwindigkeit zur Potentialbewertung (Eigene Darstellung, Quelle: BITKOM).....	98
Abbildung 50: Vorher-Nachher-Vergleich der Dimensionen Kontrolle zu Transparenz der Wertschöpfung zur Potentialbewertung (Eigene Darstellung, Quelle: BITKOM)	100
Abbildung 51: Vorher-Nachher-Vergleich der Dimensionen Gültigkeit der Empfehlung zu Zeitraum der Datenlage zur Potentialbewertung (Eigene Darstellung, Quelle: BITKOM)	101
Abbildung 52: Bewertung des Potentials für die Dimensionen Verfügbarkeit der Entscheidungsgrundlage und seiner Zeitrelevanz (Eigene Darstellung, Quelle: BITKOM)	102
Abbildung 53: Potential nach Branchen und prozentualer Anteil an untersuchten Fallbeispielen (Eigene Darstellung) 105	
Abbildung 54: Abgeleitete Schwerpunkte aus den verschiedenen Branchen (eigene Darstellung).....	107
Abbildung 55: Überblick der Untersuchung (Eigene Darstellung)	108
Abbildung 56: Der Datenanalyseprozess	113
Abbildung 57: Schichtenmodell ausgewählter „Big Data Systeme“	134
Abbildung 58: Landschaft ausgewählter Anbieter	135
Abbildung 59: Fallende Entwicklung der Preise pro Gigabyte in USDfür Arbeitsspeicher	139

Abbildung 60: Zugriffsgeschwindigkeiten auf Festplatte,
SSD und Hauptspeicher, (nach: [http://queue.acm.org/
detail.cfm?id=1563874](http://queue.acm.org/detail.cfm?id=1563874)) 140

Abbildung 61: Sieben Anbietergruppen für Daten
und datennahe Dienstleistungen teilen sich eine
gemeinsame Plattform. Diese Zusammenarbeit
ermöglicht die Berechnung des ausschlaggebenden
Informationsvorsprungs für Entscheider. Quelle:
Umfrage mit Experten und CEOs (172) (173). 153

Abbildung 62: Das Diagramm zeigt eine
Preismodellwahl, die sich an der Nachfrage des Kunden
orientiert. 158

Abbildung 63: Allgemeine Einschätzungen zu Big Data
(Quelle: Eigene Darstellung)..... 196

10 Tabellenverzeichnis

Tabelle 1: Überblick der untersuchten Studien (n > 4492).....	19
Tabelle 2: Aggregation verschiedener Studien zu den Beweggründen für Big Data (Eigene Darstellung).....	23
Tabelle 3: Aggregation verschiedener Studien zur Bedeutung von Big Data nach Branchen (Eigene Darstellung)	29
Tabelle 4: Top 10 Unternehmensaktivitäten und deren Potenzial (n=643, TCS 2013)	33
Tabelle 5: Aggregation verschiedener Studien zu den Herausforderungen für Big Data (Eigene Darstellung)	39
Tabelle 6: Herausforderungen bei der Umsetzung von Big Data Initiativen (TCS 2013/Capgemini 2012).....	40
Tabelle 7: Durchschnittliche Ausgaben und Budgetentwicklung für Big Data nach Unternehmensgröße (n=unterschiedliche Basis, BARC 2013)	44
Tabelle 8: Hintergrund der Umfrageteilnehmer und Charakteristika der Unternehmen (Eigene Darstellung).....	49
Tabelle 9: Vorteile von Big Data-Technologie für Unternehmen in Prozent (n=274).....	72
Tabelle 10: SWOT-Analyse zur Potentialbewertung von Big Data im Allgemeinen (Eigene Darstellung)	73
Tabelle 11: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Öffentlicher Sektor (Eigene Darstellung)	78
Tabelle 12: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Industrie 4.0 (Eigene Darstellung) ..	81
Tabelle 13: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Gesundheitssektor (Eigene Darstellung)	83
Tabelle 14: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Marktforschung, (Social-) Media und Entertainment (Eigene Darstellung)	86

Tabelle 15: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Mobilitätsdienstleistungen (Eigene Darstellung).....	88
Tabelle 16: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Energie (Eigene Darstellung)	90
Tabelle 17: SWOT-Analyse zur Potentialbewertung von Big Data in der Branche Risikomanagement und Versicherungswesen (Eigene Darstellung).....	93
Tabelle 18: Potentialgewichtung nach ausgewählten Branchen (Eigene Darstellung)	105
Tabelle 19: Big Data als disruptive Technologie?	202
Tabelle 20: Voraussetzungen für die Ausnutzung der Innovationspotenziale von Big Data	203

11 Anhang: Big Data und Recht

Rechtliche Rahmenbedingungen

A. Eigentum an Daten

1. Die Eigentumsfähigkeit von Daten
2. Erschöpfungsgrundsatz im Urheberrecht
3. Möglichkeit der Verallgemeinerung der rechtlichen Selbstständigkeit von Daten
4. „Eigentumsrechtliche“ Zuordnung von Daten
5. Datenschutzrecht
6. Der *sui generis*-Schutz des Datenbankherstellers
7. § 303a StGB
8. § 453 BGB
9. § 950 BGB
10. Ausgestaltung des zivilrechtlichen Schutzes
11. Fazit

B. Insolvenzrecht

C. Urheberrecht

1. Anwendbarkeit des UrhG
2. Urheberrecht an und Werkqualität von Daten
3. Betroffene Verwertungsrechte
4. Schranken des Urheberrechts
5. Einräumung von Nutzungsrechten

D. Datenschutzrecht

1. Datenschutzrechtliche Grundsätze
2. Anwendbarkeit des BDSG - Kollisionsrechtliche Fragen bei grenzüberschreitendem Bezug
3. Telemediengesetz (TMG)
4. Datenverarbeitung – Verbot mit Erlaubnisvorbehalt
5. Erlaubnisgründe

6. Auftragsdatenverarbeitung und Cloud Computing
7. US Patriot Act
8. Datentransfer/Datenaustausch
9. Rechte der Betroffenen im nicht-öffentlichen Bereich
10. Rechtsfolgen und Haftung
11. Produktbezogene informationelle Selbstbestimmung – Privacy by Design
12. Profiling – Persönlichkeitsprofile
13. Ausblick auf die Datenschutz-GrundVO

E. Vertragsrecht

1. Hauptleistungspflichten
2. Nebenleistungspflichten
3. Nachvertragliche Pflichten
4. Schadensermittlung bei Datenverlust

F. Informationshaftung – Informationsqualität

1. Informationsqualität
2. Vertragliche Haftung
3. Deliktische Haftung
4. Anscheinsbeweis
5. Haftungsausschluss

G. Aufbewahrungs- und Löschungspflichten**H. Steuerrecht****I. Strafrecht****J. Daten und Produkthaftungsrecht****K. Roboterrecht****L. Datenmarktplätze**

1. Datenschutzrecht
2. Urheberrecht

M. Medizindaten

1. Konkrete Anwendungsbeispiele
2. Rechtliche Rahmenbedingungen

8.	Datentransfer/Datenaustausch	62
9.	Rechte der Betroffenen im nicht-öffentlichen Bereich	63
10.	Rechtsfolgen und Haftung	64
11.	Produktbezogene informationelle Selbstbestimmung – Privacy by Design	65
12.	Profiling – Persönlichkeitsprofile	66
13.	Ausblick auf die Datenschutz-GrundVO.....	67
E.	Vertragsrecht.....	68
1.	Hauptleistungspflichten.....	68
2.	Nebenleistungspflichten	70
3.	Nachvertragliche Pflichten.....	71
4.	Schadensermittlung bei Datenverlust.....	71
F.	Informationshaftung – Informationsqualität	72
1.	Informationsqualität	72
2.	Vertragliche Haftung	74
3.	Deliktische Haftung	76
4.	Anscheinsbeweis.....	78
5.	Haftungsausschluss	80
G.	Aufbewahrungs- und Löschungspflichten.....	80
H.	Steuerrecht.....	82
I.	Strafrecht.....	85
J.	Daten und Produkthaftungsrecht	88
K.	Roboterrecht	90
L.	Datenmarktplätze.....	91
1.	Datenschutzrecht.....	92
2.	Urheberrecht.....	94
M.	Medizindaten	97
1.	Konkrete Anwendungsbeispiele	97
2.	Rechtliche Rahmenbedingungen	97

Literaturverzeichnis

Adolphsen/Mutz, Das Google Book Settlement, GRUR Int 2009, 789

AG Rechtsrahmen des Cloud Computing, Trusted Cloud – Datenschutzrechtliche Lösungen für Cloud Computing, 2012

ARTICLE 29 DATA PROTECTION WORKING PARTY, Opinion 03/2013 on purpose limitation,
http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf (zuletzt abgerufen am 6.9.2013)

Baeriswyl, «Big Data» ohne Datenschutz-Leitplanken, Zeitschrift für Datenrecht und Informationssicherheit, 2013, 14

Beck, Roboter, Cyborgs und das Recht – von der Fiktion zur Realität, Berlin 2010

Beck, Grundlegende Fragen zum rechtlichen Umgang mit der Robotik, JR 2009, 225

Becker/Nikolaeva, Das Dilemma der Cloud-Anbieter zwischen US Patriot Act und BDSG, CR 2012, 170

Beckmann/Müller, Online übermittelte Informationen: Produkte i.S.d. Produkthaftungsgesetzes?, MMR 1999, 14

Benlian/Hess/Buxmann, Software as a Service – Anbieterstrategien, Kundenbedürfnisse & Wertschöpfungsstrukturen, Wiesbaden 2010

Bergmann/Möhrle/Herb, Kommentar zum Bundesdatenschutzgesetz, den Datenschutzgesetzen der Länder und zum Bereichsspezifischen Datenschutz, Grundwerk mit 45. Ergänzungslieferung Stuttgart 2012

Braun, Insolvenzordnung, 5. Auflage München 2012

Breyer, Korrigierter Leitfaden zur Speicherung von Verkehrsdaten veröffentlicht, ZD-Aktuell 2012, 03218

Breinlinger/Scheuing, Der Vorschlag für eine EU-Datenschutzverordnung und die Folgen für Verarbeitung und Nutzung von Daten für werbliche Zwecke, RDV 2012, 64

Buchner, Die Einwilligung im Datenschutzrecht – vom Rechtfertigungsgrund zum Kommerzialisierungsinstrument, DuD 2010, 39

Cahn, Produkthaftung für verkörperte geistige Leistungen, NJW 1996, 2899

Conrad, Anmerkung zu BGH, Urteil vom 29. April 2010 - I ZR 69/08 – Vorschaubilder, ZUM 2010, 585

Czychowski/Bröcker, ASP – Ein Auslaufmodell für das Urheberrecht?, MMR 2002, 81

Däubler/Klebe/Wedde/Weichert, Kompaktkommentar zum Bundesdatenschutzgesetz, 3. Auflage Frankfurt a.M. 2010

Deutsch, Das neue System der Gefährdungshaftungen: Gefährdungshaftung, erweiterte Gefährdungshaftung und Kausal-Vermutungshaftung, NJW 1992, 73

Deutsch, Die Zulässigkeit des so genannten „Screen-Scraping“ im Bereich der Online-Flugvermittler, GRUR 2009, 1027

Dickmann/Rienhoff, Langzeitarchivierung von Forschungsdaten – Eine Bestandsaufnahme, Kapitel 12 Medizin, Boizenburg 2012

Doyle, National Security Letters in Foreign Intelligence Investigations: Legal Background and Recent Amendments, CRS Report for Congress, September 8, 2009,
<http://www.fas.org/sgp/crs/intel/RL33320.pdf> (zuletzt abgerufen am 6.9.2013)

Drewes/Siegert, Die konkludente Einwilligung in das Telefonmarketing und das Ende des Dogmas von der datenschutzrechtlichen Schriftform, RDV 2006, 139

Dreier/Schulze, Urheberrechtsgesetz, 4. Auflage München 2013

Dreyer/Kotthoff/Meckel, Urheberrecht, 2. Auflage Heidelberg 2008

Duisberg, Wem gehören die Daten und wer hat außerdem Rechte daran? in: Big Data wird neues Wissen, 36-54, München 2012

Eichler, Schuldrecht und IT, Anwalt 5/2002, 20

Ellger, Der Datenschutz im grenzüberschreitenden Datenverkehr. Eine rechtsvergleichende und kollisionsrechtliche Untersuchung, Baden-Baden 1990

Engel, Produzentenhaftung für Software, CR 1986, 702

Erman, Kommentar zum Bürgerlichen Gesetzbuch, 13. Auflage Köln 2011

Fischer, Kommentar zum Strafgesetzbuch, 60. Auflage München 2013

Fromm/Nordemann, Urheberrecht, 10. Auflage Stuttgart 2008

Gasser, Information Quality and the Law, or, How to Catch a Difficult Horse, Research Publication No. 2003-08 (11/2003), <http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/2003-08.pdf> (zuletzt abgerufen am 6.9.2013)

Gaster, Der Rechtsschutz von Datenbanken, Köln 1999

Gerstenberg, Löschen von Tonbändern als neuer strafrechtlicher Tatbestand, NJW 1956, 540

Giesen, Zum Begriff des Offenbarens nach § 203 StGB im Falle der Einschaltung privatärztlicher Verrechnungsstellen, NStZ 2012, 122

Gola, Beschäftigtendatenschutz und EU-Datenschutz-Grundverordnung, EuZW 2012, 332

Gola/Schomerus, Kommentar zum Bundesdatenschutzgesetz, 11. Auflage München 2012

Götting, Anmerkung zu BGH, Urteil vom 29.04.2010 – I ZR 69/08, LMK 2010, 309481

Graf/Jäger/Wittig, Wirtschafts- und Steuerstrafrecht, München 2011

Graf v. Westphalen, Das neue Produkthaftungsgesetz, NJW 1990, 83

Graf v. Westphalen, Produkthaftungshandbuch, 12. Auflage München 2012

Gramberg-Danielsen/Kern, Die Schweigepflicht des Arztes gegenüber privaten Verrechnungsstellen, NJW 1998, 2708

Grünwald/Döpkens, Cloud Control? - Regulierung von Cloud Computing-Angeboten, MMR 2011, 287

Hackenberg, Datenschutz setzt Big Data Grenzen, Computerwoche 24.1.2013,
<http://www.computerwoche.de/a/datenschutz-setzt-big-data-grenzen,2531242> (zuletzt abgerufen am 6.9.2013)

Haft, Das Zweite Gesetz zur Bekämpfung der Wirtschaftskriminalität (2. WiKG) - Teil 2: Computerdelikte, NStZ 1987, 6

Hallermann, Wann müssen Auftragsdatenverarbeitungen vor Ort kontrolliert werden?, RDV 2012, 226

Hansen, Vertraulichkeit und Integrität von Daten und IT-Systemen im Cloud-Zeitalter, DuD 2012, 407

Hartmann, Weiterverkauf und „Verleih“ online vertriebener Inhalte – Zugleich Anmerkung zu EuGH, Urteil vom 3. Juli 2012, Rs. C-128/11 – UsedSoft ./ Oracle, GRUR-Int 2012, 980

Heckmann, jurisPR IT-Recht, 23/2012, Anmerkung 2

Hegmanns/Niehaus, Outsourcing im Versicherungswesen und der Gehilfenbegriff des § 203 III 2 StGB, NStZ 2008, 57

Heintschel-Heinegg (Hrsg.), Beck'scher Online-Kommentar StGB, 22. Auflage 2013

Henning/Bazan/Doyle/Liu, Government Collection of Private Information: Background and Issues Related to the USA Patriot Act Reauthorization, CRS Report March 2, 2010,
<http://fpc.state.gov/documents/organization/139232.pdf> (zuletzt abgerufen am 6.9.2013)

Herrmann/Dehijelles, Das Schutzrecht sui generis an Datenbanken, K&R 2009, 23

Hilgendorf, Informationsstrafrecht und Rechtsinformatik, Berlin 2004

Hilgendorf, Grundfälle zum Computerstrafrecht, JuS 1996, 890

Hoenike/Hülsdunk, Outsourcing im Versicherungs- und Gesundheitswesen ohne Einwilligung?, MMR 2004, 788

Hoeren, Dateneigentum, Versuch einer Anwendung von § 303a StGB im Zivilrecht, MMR 2013, 486

Hoeren, Skriptum IT-Recht, Stand Oktober 2011, <http://www.uni-muenster.de/Jura.itm/hoeren/lehre/materialien> (zuletzt abgerufen am 6.9.2013)

Hoeren/Sieber, Multimedia-Recht, 32. Ergänzungslieferung München 2012

Hoeren, Wenn Sterne kollabieren, entsteht ein schwarzes Loch – Gedanken zum Ende des Datenschutzes, ZD 2011, 145

Hoeren/Försterling, Onlinevertrieb „gebrauchter“ Software – Hintergründe und Konsequenzen der EuGH-Entscheidung „UsedSoft“, MMR 2012, 642

Hoeren/Giurgiu, Datenschutz – Der Datenschutz in Europa nach der neuen Datenschutz-Grundverordnung, NWB 2012, 1599

Hoeren, Der Erschöpfungsgrundsatz bei Software – Körperliche Übertragung und Folgeprobleme, GRUR 2010, 665

Hoeren, Zur Beweislastumkehr bei der Produzentenhaftung, Anmerkung, CR 1992, 606

Hoeren, Anmerkung zu EuGH, Urteil vom 9.11.2004 – Rs. C-203/02, MMR 2005, 29

Hogan Lovells, A Global Reality: Governmental Access to Data in the Cloud. A comparative analysis of ten international jurisdictions, 23.5.2012,
[http://www.hldataprotection.com/uploads/file/Revised%20Government%20Access%20to%20Cloud%20Data%20Paper%20\(18%20July%2012\).pdf](http://www.hldataprotection.com/uploads/file/Revised%20Government%20Access%20to%20Cloud%20Data%20Paper%20(18%20July%2012).pdf) (zuletzt abgerufen am 6.9.2013)

Hohmann, Haftung der Softwarehersteller für das „Jahr 2000“-Problem, NJW 1999, 521

Holler, FAZ v. 16.04.2013, S. 29

Hollmann, Die EG-Produkthaftungsrichtlinie, DB 1985, 2389

Honsell, Produkthaftungsgesetz und allgemeine Deliktshaftung, JuS 1995, 211

Hüttner, 1, 2, 3, 4 Eckstein, keiner muss versteckt sein? – Wer sich im Internet präsentiert, muss mit Google rechnen!, WRP 2010, 1008

Institute of Information Law (University of Amsterdam), Cloud Computing in Higher Education and Research Institutions and the USA Patriot Act, 2012,
http://www.ivir.nl/publications/vanhoboken/Cloud_Computing_Patriot_Act_2012.pdf (zuletzt abgerufen am 6.9.2013)

Jauernig, Bürgerliches Gesetzbuch, 14. Auflage München 2011

Jotzo, Gilt deutsches Datenschutzrecht auch für Google, Facebook & Co. bei grenzüberschreitendem Datenverkehr?, MMR 2009, 232

Kieun, Database Technology for Large Scale Data, <http://www.cubrid.org/blog/web-2-0/database-technology-for-large-scale-data/> (zuletzt abgerufen am 6.9.2013)

Kilian/Heussen, Computerrecht, 31. Ergänzungslieferung München 2012

Klass, Neue Internettechnologien und das Urheberrecht: Die schlichte Einwilligung als Rettungsanker?, ZUM 2013, 1

Klein, Abgabenordnung, 10. Auflage München 2009

Kleinemenke, Anmerkung LG Hamburg, Urteil vom 26.09.2008 – 308 O 42/06, CR 2009, 55

Köpke, Die Bedeutung des § 203 Abs. 1 Nr. 6 StGB für Private Krankenversicherer, insbesondere bei der innerorganisatorischen Geheimnisweitergabe, Heidelberg 2003

Kort, Produkteigenschaft medizinischer Software, CR 1990, 171

Kort, Software - eine Sache?, DB 1994, 1505

Kühling/Seidel/Sivridis, Datenschutzrecht, 2. Auflage Heidelberg 2011

Kröger/Gimmy, Handbuch zum Internetrecht, Berlin 2012

Lackner/Kühl, Kommentar zum Strafgesetzbuch, 27. Auflage München 2011

Ladeur, Persönlichkeitsschutz und „Comedy“ - Das Beispiel der Fälle SAT 1/Stahnke und RTL/Schröder, NJW 2000, 1977

Lehmann, Produkt- und Produzentenhaftung für Software, NJW 1992, 1721

Leipziger Kommentar, Strafgesetzbuch, 12. Auflage Berlin 2008

Lensdorf/Mayer-Wegelin/Mantz, Outsourcing unter Wahrung von Privatgeheimnissen, CR 2009, 62

Leutheusser-Schnarrenberger, Zur Reform des europäischen Datenschutzrechts, MMR 2012, 709

Lindinger, FAZ v. 2.1.2013, S. 2

Loewenheim, Handbuch des Urheberrechts, 2. Auflage München 2010

Maunz/Dürig, Grundgesetz-Kommentar, 66. Ergänzungslieferung München 2012

Mayer-Schönberger/Cukier (2013), Big Data – A revolution that will transform how we live, work, and think, Boston

Meyer, Die Haftung für fehlerhafte Aussagen in wissenschaftlichen Werken, ZUM 1997, 26

Meyer, Instruktionshaftung, Bielefeld 1992

Meyer/Wehlau, Die zivilrechtliche Haftung für Datenlöschung, Datenverlust und Datenzerstörung, NJW 1998, 1585

Mincke, Die rechtliche Einordnung des Software-Erwerbs – Bedenken gegen die Sachkauf-Theorie, jurPC 1990, 406

Möller/Florax, Kreditwirtschaftliche Scoring-Verfahren – Verbot automatisierter Einzelentscheidungen gem. § 6a BDSG, MMR 2002, 806

Moos, Unzulässiger Handel mit Persönlichkeitsprofilen? – Erstellung und Vermarktung kommerzieller Datenbanken mit Personenbezug, MMR 2006, 718

Müller-Broich, Telemediengesetz, Frankfurt a.M. 2012

Müller-Hengstenberg, Computersoftware ist keine Sache, NJW 1994, 3128

Münchener Kommentar, Strafgesetzbuch, 2. Auflage, München 2012

Münchener Kommentar, Bürgerliches Gesetzbuch, Bd. 1, 6. Auflage München 2012

Münchener Kommentar, Bürgerliches Gesetzbuch, Bd. 6, 5. Auflage München 2009

Münchener Kommentar, Bürgerliches Gesetzbuch, Bd. 5, 5. Auflage München 2009

Münchener Kommentar, Insolvenzordnung, 2. Auflage München 2008

Nägele/Jacobs, Rechtsfragen des Cloud Computing, ZUM 2010, 281

Ostendorf/Frahm/Doege, Internetaufrufe zur Lynchjustiz und organisiertes Mobbing, NStZ 2012, 529

Ott, Bildersuchmaschinen und Urheberrecht, ZUM 2009, 345

Palandt, Bürgerliches Gesetzbuch, 72. Auflage München 2013

Pallasky, Schwerpunkt - TK-Überwachung - USA PATRIOT Act: Neues Recht der TK-Überwachung, DUD 2002, 221

Piper/Ohly/Sosnitza, UWG, 5. Auflage München 2010

Plath (Hrsg.), BDSG Kommentar, Köln 2013

Redeker, IT-Recht, 4. Auflage München 2007

Redeker, Software – ein besonderes Gut, NJW 2008, 2684

Reese, Produkthaftung und Produzentenhaftung für Hard- und Software, DStR 1994, 1121

Reinemann/Remmertz, Urheberrechte an User-generated Content, ZUM 2012, 216

Rengier, Strafrecht, Besonderer Teil - Bd.1 Vermögensdelikte, 14. Auflage München 2012

Rogosch, Die Einwilligung im Datenschutzrecht, Frankfurt a.M. 2013

Roßnagel/Pfitzmann/Garstka, Modernisierung des Datenschutzrechts, Berlin 2001

Sassenberg/Bamberg, Steuerberatung, EDV und Verschwiegenheit, DStR 2006, 2052

Schack, Urheberrechtliche Gestaltung von Webseiten unter Einsatz von Links und Frames, MMR 2001, 9

Schaefer, Urheberrechtliche Rahmenbedingungen für Bildersuchmaschinen de lege lata und de lege ferenda, Marburg 2009

Schmerer/Whittaker, Cloud-Computing: US-Behörden dürfen auf Daten europäischer Server zugreifen, ZDNet vom 01.07.2011, <http://www.zdnet.de/41554621/> (zuletzt abgerufen am 6.9.2013)

Schmoekel/Rückert/Zimmermann (Hrsg.), Historisch-kritischer Kommentar zum BGB, Bd. I Allgemeiner Teil §§ 1-240, Tübingen 2003

Schneider, 40 Jahre Datenschutz – Die Zukunft des Datenschutzes, MMR 2009, VII

Schneider, Hemmnis für einen modernen Datenschutz: Das Verbotsprinzip, Anwaltsblatt 2011, 233

Schneider/Spindler, Der Kampf um die gebrauchte Software – Revolution im Urheberrecht?, CR 2012, 489

Schönke/Schröder, Kommentar zum Strafgesetzbuch, 28. Auflage München 2010

Schricker/Loewenheim, Urheberrecht, 4. Auflage München 2010

Schulz, Privacy by Design, CR 2012, 204

Schumacher, Service Level Agreements: Schwerpunkt bei IT- und Telekommunikationsverträgen, MMR 2006, 12

Schwartz (2004), "Property, Privacy, and Personal Data", Harvard Law Review (Nr. 17), p. 2055

Sedlmaier/Kolk, ASP - Eine vertragstypologische Einordnung, MMR 2002, 75

Selk, Das Schicksal von ASP- und SaaS-Services in der Insolvenz des Anbieters, ITRB 2012, 201

Sendrowski, Zum Schutzrecht „sui generis“ an Datenbanken, GRUR 2005, 369

Simitis, Kommentar zum Bundesdatenschutzgesetz, 7. Auflage Baden-Baden 2011

Söbbing, Cloud und Grid Computing, MMR 5/2008, XII

Solmecke/Wahlers, Rechtliche Situation von Social Media Monitoring-Diensten – Rechtskonforme Lösungen nach dem Datenschutz- und dem Urheberrecht, ZD 2012, 550

Spies, USA: Cloud Computing – Schwarze Löcher im Datenschutzrecht, MMR 2009, XI

Spies, Cloud Computing: Keine personenbezogenen Daten bei Verschlüsselung, MMR-Aktuell 2011, 313727

Spindler/Schuster, Kommentar zum Recht der elektronischen Medien, 2. Auflage München 2011

Spindler/Hillegeist, Rechtliche Probleme der elektronischen Langzeitarchivierung von Forschungsdaten, in: Büttner/Hobohm/Müller (Hrsg.), Handbuch Forschungsdatenmanagement, Bad Honnef 2011

Spindler, Bildersuchmaschinen, Schranken und konkludente Einwilligung im Urheberrecht - Besprechung der BGH-Entscheidung „Vorschaubilder“, GRUR 2010, 785

Spindler, Verschuldensunabhängige Produkthaftung im Internet, MMR 1998, 119

Spindler, Das Jahr 2000 – Problem in der Produkthaftung: Pflichten der Hersteller und der Softwarenutzer, NJW 1999, 3737

Staudinger, Kommentar zum Bürgerlichen Gesetzbuch mit Einführungsgesetz und Nebengesetzen, Buch 2, 16. Auflage Berlin 2011

Staudinger, Kommentar zum Bürgerlichen Gesetzbuch mit Einführungsgesetz und Nebengesetzen, Buch 3, 14. Auflage Berlin 2002

Stieper, Anmerkung zu EuGH, Urteil vom 3. Juli 2012 – Rs. C-128/11 – UsedSoft, ZUM 2012, 668

Sweet & Maxwell (eds.) (2010), *Clerk & Lindsell on Torts*, 20th edition

Taeger, Produkt- und Produzentenhaftung bei Schäden durch fehlerhafte Computerprogramme, CR 1996, 257

Taeger/Gabel (Hrsg.), Kommentar zum BDSG und zu den Datenschutzvorschriften des TKG und TMG, Frankfurt a.M. 2010

Taschner/Frietsch, Produkthaftungsgesetz und EG-Produkthaftungsrichtlinie, 2. Auflage München 1990

Tettenborn, Europäische Union: Rechtsrahmen für die Informationsgesellschaft, MMR 1998, 18

Thum, Anmerkung zu BGH, Urteil vom 19.10.2011 – I ZR 140/10, GRUR-Prax 2012, 215

Tinnefeld, Die Novellierung des BDSG im Zeichen des Gemeinschaftsrechts, NJW 2001, 3078

ULD, Unterstützung für Microsofts Ansatz zu „Privacy by Default, ZD-Aktuell 2012, 03008

Wagner, Wie schützt man Daten – ohne Gesetze?, ZRP 2013, 33

Wandtke/Bullinger, Praxiskommentar zum Urheberrecht, 3. Auflage München 2009

Weichert, Big Data und Datenschutz - Chancen und Risiken einer neuen Form der Datenanalyse, ZD 2013, 251

Weiser, The Computer for the 21st Century,
http://dl.acm.org/ft_gateway.cfm?id=329126&ftid=109731&dwn=1&CFID=187817080&CFTOKEN=22196943 (zuletzt abgerufen am 6.9.2013)

Welp, Datenveränderung (§ 303a StGB) – Teil 1, IuR 1988, 443

v. *Westerholt/Berger*, Der Application Service Provider und das neue Schuldrecht, CR 2002, 81

Wicker, Vertragstypologische Einordnung von Cloud Computing-Verträgen – Rechtliche Lösungen bei auftretenden Mängeln, MMR 2012, 783

Zieger/Smirra, Fallstricke bei Big Data-Anwendungen - Rechtliche Gesichtspunkte bei der Analyse fremder Datenbestände, MMR 2013, 418

Rechtliche Rahmenbedingungen

A. Eigentum an Daten

Für das Big-Data-Management bestehen insbesondere im Bereich des Urheber- und Datenschutzrechts erhebliche rechtliche Probleme. Die Frage nach dem Eigentum oder der Inhaberschaft sonstiger Rechte an Daten stellt diese jedoch noch in den Schatten. Die rechtliche Zuordnung von Daten ist zurzeit vollkommen unklar, was aufgrund der immensen Bedeutung von Daten und Rechten an Daten in der heutigen Informationsgesellschaft als äußerst unbefriedigend empfunden werden muss und zu gravierender Rechtsunsicherheit für die Betroffenen führt.

Sowohl Privatpersonen als auch Unternehmen arbeiten tagtäglich mit großen Mengen von Daten. Für diese ist selbst mit juristischem Rat nicht ersichtlich, ob und wenn ja auf welche Weise ihre Daten rechtlich geschützt sind. Ebenso schwierig ist es, festzustellen, ob Dritten Rechte an den betroffenen Daten zustehen. Auch der Umfang solcher Rechte ist – soweit sie bestehen – momentan kaum zu bestimmen. Erschwerend hinzu kommt noch, dass die Gerichte auf deutscher und europäischer Ebene unterschiedliche Lösungsansätze vertreten. Aufgrund der enormen wirtschaftlichen Bedeutung von Daten insbesondere im Hinblick auf Big-Data-Anwendungen ist zu befürchten, dass diese Unsicherheiten große Investitionshemmnisse darstellen. Das Risiko, sich zivil- oder strafrechtlichen Ansprüchen auszusetzen oder selbst keine gesicherte Rechtsposition in Bezug auf Daten zu erlangen, ist unkalkulierbar. Deswegen sollte eine möglichst baldige Klärung der bestehenden Rechtsfragen angestrebt werden. Im Folgenden werden daher die wichtigsten Problemkreise und mögliche Lösungsansätze aufgezeigt.

1. Die Eigentumsfähigkeit von Daten

Die „Eigentumsfähigkeit“ von Daten setzt voraus, dass sie nach der Rechtsordnung einem Rechtssubjekt zugeordnet werden können. DIN 44300 Teil 2 Nr. 2.1.13 definiert Daten im technischen Sinne als ein „Gebilde aus Zeichen oder kontinuierlichen Funktionen, die aufgrund bekannter oder unterstellter Abmachungen Informationen darstellen, vorrangig zum Zwecke der Verarbeitung oder als deren Ergebnis“. Aus dieser Definition liest Welp eine Teilung des Begriffs in zwei Ebenen heraus: Den Inhalt und die Darstellung.¹ Damit lässt sich feststellen, dass ein Datum eine von diesen Ebenen abstrakte Bedeutung besitzen muss. Die Zeichen können (nicht notwendigerweise) körperlich manifestiert sein, wohingegen eine Information immer unkörperlich ist.

Rechtlich sind also drei Ebenen getrennt zu betrachten: Erstens das Medium, als physisch wahrnehmbare Sache oder Eigenschaft einer Sache (Darstellung), zweitens die Information, also der Inhalt, und drittens das davon abzugrenzende Datum als „Gebilde“. Software wird teilweise über ihre Steuerungsfunktion von Daten abgegrenzt.² Vor allem aber ist Software Inhalt und in dieser Eigenschaft nach den §§ 69a ff. UrhG urheberrechtlich geschützt.

¹ Welp, IuR 1988, 443 (444).

² von dem Bussche/Schelinski, in: Leupold/Glossner, Teil 1 Rn. 11; Dreier, in: Dreier/Schulze, § 69a Rn. 12.

Begrifflich scheinen Daten also von Sachen und Informationen trennbar zu sein. Die Frage nach der Eigentumsfähigkeit von Daten setzt daher voraus, dass ihnen auch rechtlich eine Eigenständigkeit neben den beiden darin vereinten Eigenschaften zukommt.

Um sich der Frage zu nähern, muss zunächst abgegrenzt werden, in welcher Form Medium und Information für die Rechtsordnung von Interesse sind. Eine rechtliche Einordnung der Daten in diesem Kontext ergibt sich möglicherweise aus der Beziehung von Datenträger und Information zueinander.

Ein Medium ist als Sache unabhängig von der darauf oder darin enthaltenen Information, fällt zivilrechtlich unter den Begriff der Sache nach § 90 BGB³ und damit auch unter den strafrechtlichen Sachbegriff der Vermögensdelikte.⁴ Es kann somit dem Eigentum einer Person nach § 903 BGB unterliegen.

Inhalte können wegen ihrer vielfältigen Ausgestaltung von unterschiedlichen Rechten erfasst sein. So unterliegen ihrem Schaffensprozess oder der Investition nach schutzwürdige Inhalte dem Urheberrecht oder gewerblichen Schutzrechten, personenbezogene Daten dagegen dem Datenschutzrecht. Teilweise können diese Rechte zwar unter den Eigentumsbegriff des Art. 14 Abs. 1 GG fallen.⁵ Ihr Institut wird jedoch bereits durch die entsprechenden Gesetze (UrhG, GeschmacksmusterG etc.) gewährleistet und unterliegt nicht dem zivilrechtlichen Schutzregime für Sacheigentum. Inhalte sind sowohl in ihrer Entstehung als auch in ihrer späteren Zuordnung zu Personen von der Beziehung der Person zu dem Medium zu trennen. Zwar müssen Werke, um Urheberrechtsschutz zu genießen, wahrnehmbar sein, eine körperliche Fixierung ist jedoch dafür nicht zwingend notwendig; das gilt auch für Computerprogramme.⁶

Der Gesetzgeber hat jedoch erkannt, dass Substanz und Inhalt in vielen Fällen eine Beziehung zueinander haben, die das Recht nicht ignorieren darf. Das zeigt sich beispielsweise deutlich in § 952 BGB. Hier begründet die auf einem Medium (Urkunde) verbrieftete Rechtsstellung einer Person deren Sacheigentum an dem Medium. § 98 UrhG gewährt dem Inhaber eines nach dem UrhG geschützten Rechts Ansprüche gegen den Eigentümer oder Besitzer rechtswidrig hergestellter Vervielfältigungsstücke auf Vernichtung dieser Vervielfältigungsstücke.

2. Erschöpfungsgrundsatz im Urheberrecht

Für die Untersuchung der Eigenständigkeit von Daten ist besonders der urheberrechtliche Erschöpfungsgrundsatz von Interesse. Im Rahmen der allgemeinen Ausführungen zum Urheberrecht werden weitere Aspekte dieses Grundsatzes beleuchtet.⁷

³ BGH, NJW 1988, 406 (408); Stremann, in: MüKo, § 90 Rn. 25; Jickeli/Stieper, in: Staudinger, § 90 Rn. 12.

⁴ Eser/Bosch, in: Schönke/Schröder, § 242 Rn. 9; Kühl, in: Lackner/Kühl, StGB § 142 Rn. 2.

⁵ Papier, in: Maunz/Dürig, Art. 14 GG Rn. 197-200.

⁶ KG, GRUR 1984, 507 (508); Loewenheim, in: Schricker/Loewenheim, § 2 Rn. 20; Schulze, in: Dreier/Schulze, § 2 Rn. 13 f.

⁷ Vgl. Kapitel „C. 3. a)“.

a) Der Erschöpfungsgrundsatz in § 17 Abs. 2 UrhG

In § 17 Abs. 2 UrhG ist festgelegt, dass ein einzelnes Vervielfältigungsstück eines Werkes, das mit Zustimmung des Rechteinhabers in der EU oder dem europäischen Wirtschaftsraum durch Veräußerung in den Verkehr gebracht wurde, weiterverbreitet werden darf. Beschränkungen des Nutzungsrechts gelten nur gegenüber dem unmittelbaren Vertragspartner und enden mit der Verbreitung. Im Übrigen haben solche Beschränkungen grundsätzlich ausschließlich schuldrechtliche und nur in Ausnahmefällen dingliche Wirkung.⁸ Die herrschende Meinung begründet die Erschöpfungswirkung mit dem Interesse der Allgemeinheit an der Verkehrsfähigkeit des Werkstücks.⁹ Im Grundsatz tritt Erschöpfung nur an körperlichen Werkstücken ein,¹⁰ was im Grunde auch für Vervielfältigungsstücke von Software gilt. Dort bestimmt sich die Erschöpfungswirkung nach § 69c Nr. 3 S. 2 UrhG, der sich an Art. 4 Abs. 2 der EU-RL zum Schutz von Computerprogrammen¹¹ orientiert.

Der Erschöpfungsgrundsatz durchbricht das Prinzip der Trennung von Sacheigentum und Recht am Inhalt partiell: Es wird ein Nutzungsrecht an eine Sache gekoppelt, um diese verkehrsfähig zu halten.¹² Ein Recht zur Vervielfältigung des Inhalts durch Schaffung neuer Sachen besteht hingegen grundsätzlich nicht. Diesen Grundsatz hat der *BGH* erstmals in der Parfumflakon-Entscheidung¹³ angetastet. Er stellte fest, dass auch Vervielfältigungshandlungen in Ausnahmefällen erlaubt sein können, wenn sie erforderlich sind, um die Verkehrsfähigkeit eines Werkstücks zu gewährleisten, indem beispielsweise Abbildungen des Werkes zu Werbezwecken für den Verkauf der Werkstücke dienen.

b) Erschöpfungsgrundsatz bei unkörperlichen Vervielfältigungsstücken – die Used-Soft-Entscheidung des EuGH

Weitestgehend wird die Prämisse aufrechterhalten, dass Gegenstand der Erschöpfungswirkung nur die durch Veräußerung in den Verkehr gebrachte Kopie ist.¹⁴ Bisher waren besonders zwei Fragen in diesem Zusammenhang streitig.

(1) Hintergrund

Es war lange nicht eindeutig geklärt, ob die Erschöpfungswirkung des § 69c Nr. 3 S. 2 UrhG direkt oder analog auch eine Softwarekopie erfasst, die der rechtmäßige Erwerber einer aus dem Internet heruntergeladenen Programmdatei mit Zustimmung des Rechteinhabers anfertigt. Ferner war umstritten, ob eine solche Kopie auch nach § 69c Nr. 3 S. 2 UrhG von einem Zweiterwerber, nach Übermittlung von dem Ersterwerber über ein Netzwerk, erstellt werden darf. Diese Fragen waren Gegenstand einer Vorlage des *BGH* an den *EuGH*, der

⁸ *Heerma*, in: *Wandtke/Bullinger*, UrhG § 17 Rn. 30ff.

⁹ *BGH*, GRUR 2001, 153 (154) – OEM-Version; *BGH*, GRUR 1986, 736 – Schallplattenvermietung; *Heerma*, in: *Wandtke/Bullinger*, § 17 Rn. 30; *Wiebe*, in: *Spindler/Schuster*, 1. Teil, § 17 Rn. 7.

¹⁰ *Schulze*, in: *Dreier/Schulze*, § 17 Rn. 28; *Stögmüller*, in: *Leupold/Glossner*, Teil 5 Rn. 187; *Wiebe*, in: *Spindler/Schuster*, 1. Teil, § 17 Rn. 6.

¹¹ RL 2009/24/EG, entspricht Art. 4 lit. c der RL 91/250/EWG.

¹² *Hoeren*, GRUR 2010, 665 (669).

¹³ *BGH*, GRUR 2001, 51.

¹⁴ *OLG Düsseldorf*, ZUM 2010, 60.

diesbezüglich über die Auslegung der Computerprogramme-RL zu urteilen hatte.¹⁵ Für eine ausführlichere Wiedergabe des Streitstandes wird an dieser Stelle auf den Teil zum Urheberrecht verwiesen. Zentraler Punkt der Entscheidung des *EuGH*, und somit auch für das hier diskutierte Problem von höchster Bedeutung, ist jedoch die Frage, wie weit der Erschöpfungsgrundsatz von der Sache gelöst und auf Daten übertragen werden kann. In seiner Vorlageentscheidung lehnte der *BGH* die Erstreckung des Erschöpfungsgrundsatzes auf unkörperliche Softwarekopien mit dem Argument ab, dass Zweck der Erschöpfung nur sein kann, die Verkehrsfähigkeit eines Vervielfältigungsstückes zu gewährleisten, nicht aber, eine vorher nicht vorhandene Verkehrsfähigkeit zu begründen.¹⁶

(2) Das Urteil des *EuGH*

Dieses Argument zeigt deutlich, dass aus urheberrechtlicher Sicht nur auf die Belegenheit einer Information auf einer Sache abgestellt und eine darüber hinausgehende Verkehrsfähigkeit von Daten nicht anerkannt wird. Dem schloss sich der *EuGH* allerdings nicht an und entschied, gestützt auf die Erwägungsgründe der RL 2009/24/EG, dass die Erschöpfungswirkung an jeder im EWR erstverkauften Programmkopie eintritt, unabhängig davon, ob sie körperlich oder unkörperlich in den Verkehr gebracht wurde.¹⁷

(3) Bedeutung für die Frage des Dateneigentums

Hier wird zum ersten Mal deutlich, dass das „Verschicken“ von Daten, das im Grunde aus einer Vervielfältigung und anschließenden Unbrauchbarmachung besteht, nicht als Vervielfältigungshandlung im Sinne des Art. 4 Abs. 1 lit. a RL 2009/24/EG gelten soll. Daraus lässt sich entnehmen, dass der *EuGH* den Daten, die ein Computerprogramm beinhaltet, neben dem Speichermedium anders als der *BGH* eine selbstständige, rechtlich relevante Eigenschaft zugesteht. Dem Erwerber von Daten stehen sodann die Rechte aus Art. 5 RL 2009/24/EG (respektive § 69d UrhG) zu, da durch die Erschöpfung die Rechtmäßigkeit der Benutzung nicht mehr von der Zustimmung des Rechteinhabers abhängt. Zweifelhaft ist aber, ob sich diese Rechtsprechung auch auf andere Werke als Software übertragen lässt.¹⁸ Dies wurde in Deutschland von der Rechtsprechung in einem ersten Urteil nach der *Used-Soft*-Entscheidung des *EuGH* in Bezug auf E-Books abgelehnt,¹⁹ ist aber umstritten.²⁰ Diese Diskussion verdeutlicht die Schwierigkeiten und Unklarheiten, die mit der Frage der rechtlichen Eigenständigkeit von Daten einhergehen.

3. Möglichkeit der Verallgemeinerung der rechtlichen Selbstständigkeit von Daten

Damit die rechtliche Selbstständigkeit von Daten sicher bejaht werden könnte, wäre allerdings eine Verallgemeinerung der rechtlichen Unabhängigkeit von Daten erforderlich.

¹⁵ *BGH*, GRUR 2011, 418.

¹⁶ *BGH*, GRUR 2011, 418 (421).

¹⁷ *EuGH*, NJW 2012, 2565 (2568).

¹⁸ Siehe dazu auch Kapitel „C. 3. b.“.

¹⁹ *LG Bielefeld*, U. v. 05.03.2013 - 4 O 191/11.

²⁰ *Hartmann*, GRUR-Int. 2012, 980 (987); *Hoeren/Försterling*, MMR 2012, 642 (647); *Schneider/Spindler*, CR 2012, 489 (497); *Stieper*, ZUM 2012, 668 (670).

Insoweit könnte sich die Heranziehung der Grundsätze der Parfumflakon-Entscheidung des *BGH* als fruchtbar erweisen.²¹ Wenn jedenfalls die Berechtigung zur Nutzung eines Werkes mit der Übertragung von Daten an eine andere Person übergehen soll, müssen Daten, parallel zu Sachen, einer Person rechtlich zugeordnet werden können. Es entspricht nicht dem System des Urheberrechts und auch sonst nicht der Systematik über den Erwerb von Rechten, dass die gewillkürte Übertragung einer Rechtsposition durch einen reinen Realakt geschehen muss. So sind neben der einfachen Konstellation, dass ein Ersterwerber ein Computerprogramm in Datenform an einen Zweiterwerber „verschickt“, weitaus kompliziertere Fälle denkbar, in denen die Zuordnung der Rechtsposition nicht so eindeutig feststellbar ist. Wie verhält es sich danach etwa, wenn der Zweiterwerber nicht Eigentümer des Datenträgers ist, auf den die Daten kopiert werden? Oder wie wirkt sich die Übermittlung von Daten in einem Mehrpersonenverhältnis aus?

Die Rechtsprechung des *EuGH* klärt zwar eine Frage zum Erschöpfungsgrundsatz im Urheberrecht, lässt dabei aber die Frage, wer eigentlich „Berechtigter“ an Daten sein kann, unbeantwortet.

4. „Eigentumsrechtliche“ Zuordnung von Daten

Ausgangspunkt der weiteren Überlegungen ist das Eigentum an dem jeweiligen Medium. Der Eigentümer eines solchen Mediums kann, soweit nicht Rechte Dritter entgegenstehen, mit dem Medium nach Belieben verfahren und Dritte von jeder Einwirkung ausschließen, § 903 BGB. Gleiches gilt für die physikalische Materialisierung von Daten auf dem Speichermedium. Der Eigentümer kann somit grundsätzlich jeden von der Nutzung von Daten und Datenspeichern ausschließen, soweit nicht das Gesetz oder die Rechte Dritter entgegenstehen. Insofern stellt sich die Frage, ob Dritte ein dem Eigentumsrecht an dem Speichermedium entgegenstehendes Recht geltend machen können. Eigentum zeichnet sich durch die rechtliche Zuweisung einer Sache zu einer Person im Sinne eines umfassenden Herrschaftsrechts aus.²² Ein irgendwie geartetes „Dateneigentum“ muss sich demnach zumindest durch diese Merkmale auszeichnen.

a) Daten als Sachen im Sinne von § 90 BGB

Das konstitutive Merkmal ist dabei die Sacheigenschaft. Eine direkte Anwendung des § 903 BGB kommt ausschließlich in Betracht, wenn körperliche Gegenstände im Sinne des § 90 BGB betroffen sind.²³

In dem Moment, in welchem die Daten auf dem Datenträger gespeichert werden, tritt eine physikalische Veränderung in materieller Hinsicht ein. Das allein ist allerdings nicht hinreichend. Im Hinblick auf die Eigenschaft von Sachen, drei Aggregatzustände annehmen zu können (fest, flüssig, gasförmig), werden Daten teilweise schon deshalb nicht als körperliche Gegenstände angesehen, weil sie aus elektrischen Spannungen bestünden, die

²¹ *Hoeren*, GRUR 2010, 665 (673).

²² *Klinck*, in: *Staudinger*, Eckpfeiler des Zivilrechts, Abschn. W Rn. 1-4; *Wendehorst*, MüKo, Art. 43 EGBGB Rn. 41.

²³ *Bassenge*, in: *Palandt*, § 903 Rn. 2; *Seiler*, in: *Staudinger*, Vorbemerkung zu § 903 Rn. 3.

diese Eigenschaft gerade nicht aufweisen.²⁴ Ohne tief in die physikalische Ebene einzudringen (etwa hinsichtlich der Frage, ob elektrische Spannung in Form von Elektronen nicht doch Materie ist²⁵), ist eine Anknüpfung an die molekulare Ebene von Materie ein Tretminenfeld für die juristische Glaubwürdigkeit. Rechtsbegriffe müssen generalisiert und abstrakt aufgefasst werden. Das Schicksal eines Rechtsinstituts wie dem Eigentum sollte nicht von umstrittenen und wandelbaren Modellen naturwissenschaftlicher Erkenntnisse abhängen. Allein die Orientierung an den Aggregatzuständen ist heikel. Aggregatzustände betreffen nämlich einen einzelnen chemischen Stoff, Sachen im juristischen Sinn können aber aus mehreren Stoffen zusammengesetzt sein. Der Begriff ist somit abstrakter als der naturwissenschaftliche Stoffbegriff.

Ferner stellen Daten auch nicht immer elektrische Spannungen dar, können sie vielmehr auch in magnetischer oder optischer Form oder als Oberflächenstruktur (z. B. bei einer CD) beschaffen sein. All diesen Erscheinungsformen ist jedoch gemein, dass sie *Eigenschaften* einer Sache – respektive des Datenträgers – sind. Daten sind damit als durch Zeichen vermittelte Information nicht eigenständig, sondern körperlich abhängig. Mit anderen Worten drückt dies die weitgehend unstreitige Erkenntnis aus, dass nicht Daten, sondern nur der Datenträger körperlich sein kann.²⁶ Daten können also nicht als Sachen im Sinne des § 90 BGB qualifiziert werden.

**b) Daten als wesentliche Bestandteile einer Sache im Sinne von
§ 93 BGB**

Im vorliegenden Fall stellt sich die Frage, ob gespeicherte Daten möglicherweise als wesentliche Bestandteile im Sinne von § 93 BGB angesehen werden können. Nach § 93 BGB sind Bestandteile einer Sache, die voneinander nicht getrennt werden können, ohne dass der eine oder andere zerstört oder in seinem Wesen verändert wird (wesentliche Bestandteile), nicht Gegenstand besonderer Rechte. Insofern würde dann das Eigentum am Datenträger auch das Eigentum an den darauf enthaltenen Daten umfassen. Dies erscheint aber insoweit problematisch, als dann „Dateneigentum“ und Sacheigentum nicht auseinander fallen könnten, was im Hinblick auf die tatsächlichen Verhältnisse so kaum praktikabel ist.

c) Eigentum am Datenträger

Die Rechtsprechung geht, wie oben bereits beschrieben, davon aus, dass ein Datenträger mit den darauf verkörperten Daten eigentumsfähig ist. So hat das Oberlandesgericht Karlsruhe²⁷ eine Eigentumsfähigkeit von Daten im Rahmen von § 823 Abs. 1 BGB bejaht.²⁸ Der Datenträger mit den darauf verkörperten Daten sei jedenfalls eine körperliche Sache.²⁹ Zwar könne die Information, die in diesen Daten repräsentiert ist, wegen ihrer immateriellen

²⁴ OLG Dresden, NJW-RR 2013, 87; LG Konstanz, NJW 1996, 2662.

²⁵ Vgl. Rüfner, in: HKK, §§ 90–103 Rn. 19; Stresemann, in: MüKo, § 90 Rn. 24.

²⁶ Stresemann, in: MüKo, § 90 Rn. 25.

²⁷ OLG Karlsruhe, NJW 1996, 200.

²⁸ Anderer Ansicht LG Konstanz, NJW 1996, 2662; siehe auch Gerstenberg, NJW 1956, 540.

²⁹ Jickeli/Stieper, in: Staudinger, § 90 Rn. 12; Redeker, NJW 2008, 2684 (2685).

Natur nicht Schutzgut einer Norm gegen die Beschädigung der materiellen Substanz sein. Jedoch sei deswegen von einer Verletzung des Eigentums auszugehen, weil der Eigentümer durch den Eingriff gehindert werde, mit der Sache seinem Wunsch entsprechend zu verfahren.

d) Recht am eigenen Datenbestand

Im Übrigen wird zum Teil vertreten, dass sich aus § 823 Abs. 1 BGB als sonstiges Recht auch ein Recht am eigenen Datenbestand herleiten lasse.³⁰

Ein solches Recht benötigt freilich zur Legitimation eine gewisse eigenständige Bedeutung neben anderen Rechten. Werden Daten auf einem Datenträger unbrauchbar gemacht, so ist – sogar weitergehend als nach der Ansicht des *OLG Karlsruhe*³¹ – von einer Substanzverletzung an dem Datenträger auszugehen, da dieser nachhaltig physisch verändert wurde.³² Mit diesem Rückgriff auf das Eigentum am Datenträger werden allerdings viele Probleme, die bei vernetzten Datenbeständen auftreten, nicht gelöst.³³ Bedeutsam sind besonders die Fälle, in denen der wirtschaftliche Schaden nicht beim Eigentümer des Datenträgers entsteht, weil nicht er, sondern ein Dritter den wirtschaftlichen Umgang mit der Information pflegt. Für solche Fälle hat das Recht am eigenen Datenbestand durchaus seine Berechtigung. Allerdings zeichnen sich „sonstige Rechte“ im Sinne des § 823 BGB dadurch aus, dass sie eine den ausdrücklich normierten Rechten entsprechende Ausschluss- und Nutzungsfunktion besitzen.³⁴ Genauer stellt sich die Frage, wer unter Ausschluss Dritter an einem Datenbestand berechtigt ist. Das Recht am eigenen Datenbestand ist daher eine Hilfskonstruktion, die zunächst mehr Probleme aufwirft als löst.

5. Datenschutzrecht

a) Grundlegendes

Das Datenschutzrecht schützt nur dem Namen nach „Daten“. Tatsächlich geht es um den dargestellten Inhalt, sprich die Information, welche sich mit einer Person in Beziehung bringen lässt. Dies lässt § 3 Abs. 1 BDSG erkennen, der personenbezogene Daten als „Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person“ definiert. Personenbezogene Daten im Sinne des Datenschutzrechts sind also „Angaben“ und damit Inhalte und dürfen nicht mit Daten im technischen Sinne verwechselt werden. Alle Daten, die einen Bezug zu einer natürlichen Person aufweisen oder die sich mit zusätzlichen Informationen auf diese beziehen lassen, begründen die Anwendbarkeit des BDSG. Es handelt sich im Ergebnis also um ein Persönlichkeitsrecht.

³⁰ Siehe dazu BGH, NJW 1996, 2924, 2925; ähnlich auch Meyer/Wehlau, NJW 1998, 1585 (1588).

³¹ OLG Karlsruhe, CR 1987, 19.

³² OLG Oldenburg, MDR 2012, 403; Meyer/Wehlau, NJW 1998, 1585 (1588).

³³ Meyer/Wehlau, NJW 1998, 1585 (1588).

³⁴ Wagner, in: MüKo, § 823 Rn. 143.

b) Abgrenzung zum „Dateneigentum“

Der Schutz personenbezogener Daten wird an vielen Stellen über die Beziehung zwischen der Information und dem Datum geregelt. Anders als im vorangegangenen Abschnitt entstehen durch das Datenschutzrecht keine über eine Information vermittelten Beziehungen einer Person zu einer Sache, sondern direkt zu Daten. So können beispielsweise nach § 35 BDSG Löschungs- oder Berichtigungsansprüche gegen eine verarbeitende Stelle unabhängig davon geltend gemacht werden, in welcher Form die Daten vorhanden sind. Auch Auskunftsansprüche (z. B. aus § 34 BDSG) gewähren dem Betroffenen immer nur das Recht, Daten vorgelegt zu bekommen, die mit den gespeicherten Daten inhaltsgleich sind. Das umfasst zwar beispielsweise auch die Auskunft über die Bezeichnung der Datei, in der die personenbezogenen Daten gespeichert sind,³⁵ der Betroffene erhält aber nicht das Recht auf Zugang zu dem Hauptspeichermedium, wie etwa dem Server der verarbeitenden Stelle. Das Datenschutzrecht schafft damit eine rechtliche Verantwortlichkeit für Daten. Es darf aber nicht dahingehend missverstanden werden, dass der Betroffene Ausschließlichkeitsrechte an den einzelnen Datensätzen im Sinne eines „Eigentums“ besitzt.

c) Datenschutzrecht als „Quasi-Eigentum“ an personenbezogenen Daten?

Einige Autoren argumentieren, es sei an der Zeit, den Datenschutz von seinem persönlichkeitsrechtlichen Ursprung zu lösen und im Hinblick auf die zunehmende Bedeutung personenbezogener Informationen, in ein „Quasi-Eigentum“ zu wandeln.³⁶ Im Ergebnis beziehen sich aber auch diese Positionen auf die rechtliche Stellung der Person zu „Daten“ im inhaltlichen Sinne. Die „eigentumsähnliche“ Position kommt dabei eher einer „urheberrechtsähnlichen“ Position nahe.

So unternahm bereits Paul Schwartz, orientiert am US-amerikanischen Recht, den Versuch, ein System des „Eigentums“ an personenbezogenen Daten zu schaffen. Dieses „Eigentum“ macht er an fünf Säulen fest: „inalienabilities, defaults, a right of exit, damages and institutions“.³⁷

d) Orientierung am KUG?

Der Ansatz von *Ladeur* orientiert sich an dem Recht am eigenen Bild aus § 22 KUG.³⁸ Dieses gibt dem Verletzten primär einen Anspruch auf Vernichtung (§ 37 KUG) und wird gleichzeitig von Unterlassungs- und Beseitigungsansprüchen flankiert.³⁹ Interessant ist aber auch das Recht auf Übernahme der Exemplare aus § 38 KUG. Mit der Absicht eine Parallelie zum Eigentum zu ziehen, ließe sich darin eine Art Herausgabeanspruch erblicken. Allerdings ist auch evident, dass das Eigentum an analogen Bildträgern eben nicht *ipso iure* auf den Verletzten übergeht, sondern nur ein Anspruch auf Übernahme besteht. Anders kann es sich bei einer Eigentumsposition an Daten, ungeachtet dessen, wie sie konkret ausgestaltet ist,

³⁵ HessVGH, RDV 1991, 187; Gola/Schomerus, § 34 Rn. 9.

³⁶ Ladeur, NJW 2000, 1977 (1980); vgl. auch Schwartz (2004), Harvard Law Review, p. 2055-2128.

³⁷ Schwartz (2004), Harvard Law Review, p. 2094.

³⁸ Ladeur, NJW 2000, 1977 (1980).

³⁹ Fricke, in: Wandtke/Bullinger, § 22 KUG Rn. 23; Loewenheim, in: Schricker/Loewenheim, § 4 Rn. 9.

nicht verhalten. Das Recht am eigenen Bild ist also keine wirkliche Eigentumsposition, sondern gibt als urheberrechtsähnliches Recht dem Abgebildeten die Möglichkeit zur Einwirkung auf fremdes Eigentum.

6. Der *sui-generis*-Schutz des Datenbankherstellers

Daneben findet sich im Urheberrecht ein weiterer Hinweis auf eine mögliche Zuordnung von Rechten an Daten: Die §§ 87a ff. UrhG, mit dem darin enthaltenen Leistungsschutzrecht für Datenbankhersteller. Wie bereits festgestellt, schützt das Urheberrecht primär Inhalte. Das gilt auch für Datenbankwerke, die als Unterfall der Sammelwerke nach § 4 Abs. 2 UrhG geschützt sind, wenn sie in ihrer Auswahl oder Anordnung persönliche geistige Schöpfungen darstellen.⁴⁰ Mit den §§ 87a ff. UrhG wurde aber in Umsetzung der Richtlinie 96/9/EG (EG-Datenbanken-Schutzrichtlinie) ein viel diskutierter Schutz *sui generis* geschaffen,⁴¹ auf den in den Ausführungen zum Urheberrecht noch genauer einzugehen sein wird.⁴² Der Datenbankhersteller wird jedenfalls vor ungenehmigter Vervielfältigung, Verbreitung und öffentlicher Wiedergabe geschützt.

a) Natur des Datenbankschutzes

Dass Datenbanken nicht gleichbedeutend mit den darin enthaltenen Daten sind, versteht sich bereits per Definition.⁴³ Es stellt sich aber die Frage, ob der Gesetzgeber mit dieser Regelung auch zum Ausdruck gebracht hat, dass er Daten nur im Umfang dieses *sui-generis*-Schutzes für schützenswert hält. Immerhin betonen die Erwägungsgründe der Datenbanken-Schutzrichtlinie, dass hinsichtlich des Schutzes des Datenbank-Marktes Handlungsbedarf bestand (EwG 1-4) und sich ein Ausschließlichkeitsrecht am besten am Urheberrecht orientieren könne (EwG 5). Der Richtliniengeber erkannte ferner, dass eine Lücke oder zumindest eine Unklarheit auf dem Gebiet des wirtschaftlichen Schutzes von Daten besteht, welcher der wachsenden Bedeutung nicht gerecht wird (EwG 10). In diesem Zuge wurde die Schutzwürdigkeit jedoch in der Investition (EwG 10-12) und nicht in einer eigentumsähnlichen Beziehung von Personen zu unterschiedlichen Daten gesehen.

b) Schutzworaussetzungen

Voraussetzung für den Schutz ist also nach § 87a Abs. 1 UrhG, dass mehrere Daten in einer bestimmten Systematik angeordnet werden und für die Beschaffung, Überprüfung oder Darstellung eine wesentliche Investition getätigt wird. Ungeschützt bleiben also nicht nur Einzeldaten, sondern auch „unsystematische Datenhaufen“.⁴⁴ Nach der kontroversen⁴⁵ Rechtsprechung des *EuGH* genügen Investitionen, die zur *Erzeugung* (nicht zur *Beschaffung*) der Daten dienen, jedoch nicht, um Schutz auszulösen.⁴⁶

⁴⁰ Bullinger, in: Wandtke/Bullinger, § 4 Rn. 8.

⁴¹ Gaster, in: Hoeren/Sieber, Teil 7.6 Rn. 1ff.; Sendrowski, GRUR 2005, 369.

⁴² Siehe Kapitel „C. 2. b)“.

⁴³ Sendrowski, GRUR 2005, 369 (372).

⁴⁴ KG, GRUR-RR 2001, 102 (102); OLG München, GRUR-RR 2001, 228 (229); Gaster, S. 37 Rn. 66.

⁴⁵ EuGH, MMR 2005, 29 (35) m. Anm. Hoeren; Lehmann, CR 2005, 10.

⁴⁶ EuGH, GRUR 2005, 252.

Jedenfalls findet auch dieser Streit auf der Ebene des Dateninhalts statt. Die Frage, ob auch solche „Daten“, zu deren Erzeugung eine wesentliche Investition getätigt wurde, den Schutz einer Datenbank begründen können und wann diese „Daten“ wesentliche Teile der Datenbank sind, knüpft an einen inhaltlichen Begriff von Daten an. Der in der Einleitung besprochene technische Datenbegriff erkennt jedoch gerade eine losgelöste Ebene an. Auch hier, ebenso wie im Datenschutzrecht, hat der Begriff „Daten“ eine eigenständige Bedeutung. Deshalb muss strikt unterschieden werden: Daten im inhaltlichen Sinne umfassen Informationen. Sie sind Inhalt einer Datenbank. Daten im technischen Sinne beinhalten Dateien und Datenstrukturen. Diese können sowohl die Darstellung der Elemente einer Datenbank als auch die Darstellung der gesamten Datenbank selbst sein.

c) **Schutz von Daten?**

Der als Ausschließlichkeitsrecht und subjektiv-rechtliche Rechtsposition ausgestaltete⁴⁷ Schutz des Datenbankherstellers begründet aber nur den Schutz vor Vervielfältigung, Verbreitung und öffentlicher Wiedergabe, nicht jedoch ein eigentumsähnliches Vollrecht. Auch die Erschöpfungswirkung an körperlichen Vervielfältigungsstücken der Datenbank⁴⁸ zeigt, dass gerade nicht der Schutz von Daten bezweckt wird, sondern der des Dateninhalts. Im Vordergrund stehen der Investitionsschutz und die wirtschaftliche Verwertung der Investition. Wie andere Leistungsschutzrechte oder gewerbliche Schutzrechte hat der Datenbankschutz nur eine begrenzte Schutzdauer (nämlich 15 Jahre, § 87d UrhG). Weiterhin umfasst er auch Vervielfältigungen von Daten, die sich durch gleichen Inhalt auszeichnen, schließt aber Dritte nicht von der bloßen Benutzung (Abfrage) der Informationen aus. Dies zusammen genommen zeigt, dass mit der Regelung für Datenbanken ein spezieller Investitionsschutz geschaffen werden sollte, der vor wirtschaftlicher Ausbeutung fremder Leistung schützt, nicht aber eine rechtliche Zuordnung von Daten zu einer Person. Der sui-generis-Schutz schafft eine zusätzliche abstrakte Ebene, die in einer systematischen Anordnung von Inhalten besteht, jedoch selbst lediglich inhaltlicher Natur ist.

7. **§ 303a StGB**

Möglicherweise können Grundgedanken aus dem Strafrecht zur Beantwortung der Frage, wem Daten gehören, beitragen.⁴⁹ Aus der oben in Bezug auf das Datenschutzrecht angesprochenen Idee des „Quasi-Eigentums“ an personenbezogenen Daten folgt zumindest im Strafrecht im Rahmen des § 303a StGB die Notwendigkeit, die Zuordnung der Daten zum Berechtigten zu definieren. Diese Zuordnung wird zum großen Teil so verstanden, dass an den Daten ein Vollrecht analog § 903 BGB entstehe.⁵⁰ Anknüpfungspunkte für die Zuordnung eines solchen Rechts sollen u. a. die nach den Regeln des Zivilrechts zu qualifizierenden dinglichen und obligatorischen Rechte an Datenträgern sein, also die durch das Herstellen der Daten vermittelte Urheberschaft (unter Rückgriff auf § 950 BGB sowie die Grundsätze in

⁴⁷ Vgl. BGH, MMR 1999, 470 (472); Gaster, in: Hoeren/Sieber, Teil 7.6 Rn. 75 f.

⁴⁸ Gaster, S. 131 Rn. 522 ff.; Dreier, in: Dreier/Schulze, § 87b Rn. 18; Thum, in: Wandtke/Bullinger, § 87b Rn. 58

⁴⁹ Vgl. hierzu eingehend Hoeren, MMR 2013, 486 (486 ff.).

⁵⁰ Hilgendorf, JuS 1996, 890 (890); Stree/Hecker, in: Schönke/Schröder, § 303a Rn. 3; Welp, IuR 1988, 443 (448).

§ 69a ff. UrhG) und die Inhaberschaft an den gespeicherten Originaldaten im Rahmen einer Auftragsdatenverarbeitung.⁵¹

a) Zuordnung nach der persönlichen Betroffenheit durch Daten

Dieser Ansatz wurde in der Literatur zu § 303a StGB schon kurze Zeit nach der Entstehung der Vorschrift diskutiert.⁵² Im Zuge der Entwicklung in der Wissenschaft wurde aber richtigerweise die Unzweckmäßigkeit des Kriteriums der Betroffenheit für die Herrschaftszuordnung erkannt. Zwar gibt eine Formulierung in den Gesetzesmaterialien eine entsprechende Auslegung her, doch soll im Hinblick auf das Rechtsgut des § 303a StGB gerade die Verfügungsbefugnis über die Integrität der Daten von ihrem Inhalt getrennt werden.⁵³ Das durch die Strafnorm geschützte Rechtsgut ist die Verwendbarkeit der Daten durch den Berechtigten.⁵⁴ Eine Zuordnung nach Betroffenheit würde aber gerade den Inhalt und die Zuordnung gleichstellen.

Eine Zuordnung, gestützt auf das Kriterium der Betroffenheit, ließe sich auch nicht mit dem Datenschutzrecht in Einklang bringen. Die datenschutzrechtlichen Vorschriften sollen die Einschränkungen, denen die datenverarbeitende Stelle im Umgang mit den Daten (im technischen Sinne) unterliegt, abschließend regeln. Nur in dem dort geregelten Umfang sollen Beschränkungen bestehen. Erfolgte die Zuordnung im Rahmen des § 303a StGB aufgrund der Betroffenheit, so würde die gesetzliche Erlaubnis zur Verarbeitung personenbezogener Daten einem Eingriff in das „Dateneigentum“ des Betroffenen gleichkommen. Auch ergibt sich die nähere Beziehung zum Sacheigentum im Gegensatz zum Persönlichkeitsrecht aus der systematischen Stellung des § 303a StGB in den Eigentumsdelikten, direkt nach der Sachbeschädigung.⁵⁵

b) Zuordnung nach Sacheigentum am Datenträger

Teilweise wird angenommen, die Zuordnung von Daten zu einer Person im Rahmen von § 303a StGB folge direkt den dinglichen Rechten am Datenträger.⁵⁶ Dieser Ansatz verkennt, dass der Eigentümer von Datenspeichern, etwa von Servern, regelmäßig in keiner Beziehung zu den Daten steht, weil er nur Dritten den Speicherplatz zur Verfügung stellt. Wer etwa als Host-Provider nur eingeschränkt für den Inhalt von Daten verantwortlich ist,⁵⁷ kann nicht einziger Verfügungsberechtigter über die Daten im Sinne eines strafrechtlichen Eigentumsdelikts sein. Schon aus der Laiensphäre kann die Wertung des § 303a StGB nicht bedeuten, dass der Nutzer eines Host-Dienstes die Verfügungsbefugnis über „seine“ Daten nur aufgrund der schuldrechtlichen Beziehung zum Serverbetreiber besitzt.⁵⁸ Das zeigt sich

⁵¹ Ähnlich *BayObLG*, CR 1993, 779; *Bär*, in: *Graf/Jäger/Wittig*, § 303a Rn. 13; *Cornelius*, in: *Kilian/Heussen*, Teil 10 Rn. 180; *Wolff*, in: *Leipziger Kommentar*, § 303a Rn. 10 ff.

⁵² *Welp*, IuR 1988, 443 (448) m.w.N in Fn. 46.

⁵³ *Hilgendorf*, JuS 1996, 890 (892); *Welp*, IuR 1988, 443 (448).

⁵⁴ *Hilgendorf*, JuS 1996, 890 (890); *Rengier*, § 26 Rn. 2; *Welp*, IuR 1988, 443 (448).

⁵⁵ *Haft*, NstZ 1987, 6 (10).

⁵⁶ *Stree/Hecker*, in: *Schönke/Schröder*, § 303a Rn. 3; *Wieck-Noodt*, in: *MüKo*, StGB, § 303a Rn. 10.

⁵⁷ *Hoeren*, in: *Hoeren/Sieber*, Teil 18.2 Rn. 124.

⁵⁸ Wie aber angedeutet von *Wieck-Noodt*, in: *MüKo*, StGB, § 303a Rn. 10.

auch im Vergleich mit der Situation zwischen dem Eigentümer einer Mietwohnung und den Gegenständen, die der Mieter in der Wohnung aufbewahrt.

Zudem soll § 303a StGB der Berechtigung an den Daten eine eigenständige Bedeutung geben, ansonsten hätte der Tatbestand auch an eine Sachbeschädigung durch Verändern von gespeicherten Daten anknüpfen können. Um die Interessen des Datenverkehrs zu wahren, müssen auch Datenträgereigentum und Dateneigentum auseinander fallen können.⁵⁹

Das Kriterium der Eigentümerschaft am Speichermedium kann deshalb höchstens als eines von mehreren Kriterien wirken, um die Eigentumsverhältnisse an Daten zu klären.

c) Zuordnung nach Schaffensprozess

Als Zuordnungskriterium wird weiterhin der Prozess der Entstehung von Daten diskutiert. In Frage kommen diesbezüglich die geistige oder die technische „Urheberschaft“.⁶⁰

Die geistige Urheberschaft knüpft an den Dateninhalt an. Gegen eine Beurteilung der Datenberechtigung auf der Grundlage geistiger Urheberschaft wird eingewandt, dass damit § 303a StGB in eine Erweiterung des Urheberrechtsschutzes verwandelt und so eine weitreichende Pönalisierung inhaltsverändernden Verhaltens außerhalb der §§ 69a ff., 106 ff. UrhG geschaffen würde, die vom Regime des Urheberrechts nicht vorgesehen ist.⁶¹ Abgesehen von diesem überzeugenden wertenden Argument erscheint das Kriterium der geistigen Urheberschaft kaum praktikabel. Für eine quasi-dingliche Rechtsposition ist es problematisch, vollständig auf ein Publizitätsmoment zu verzichten. Jede Kopie des Inhalts von Daten oder auch nur die Eingabe eines zuvor analog verkörperten Gedankeninhalts in digitale Daten würde ferner dem ursprünglichen Urheber des Inhalts zufallen. Parallel zu der Zuweisung nach dem Eigentum am Speichermedium würde wieder die Eigenständigkeit der Daten neben Medium und Inhalt in Frage gestellt. Damit ließe sich die spezifische Regelung für Daten nicht erklären.

Als dogmatisch einwandfreistes und operabelstes Kriterium gilt der Prozess der technischen Herstellung der Daten. Welp prägte dafür den Begriff des „Skripturakts“.⁶² „Skribent“ und damit originär Berechtigter an den Daten soll derjenige sein, der durch Eingabe oder Ausführung eines Programms Daten selbst erstellt.⁶³ Dieses Kriterium ist insofern dogmatisch und praktisch brauchbar, als es gerade an die spezifische Dateneigenschaft anknüpft. Der „Skribent“ ist der technische „Ersteller“ der Daten, zunächst unabhängig davon, auf wessen Medium die Speicherung geschieht und wer geistig den Inhalt geprägt hat. Auch innerhalb eines Arbeits- oder Dienstverhältnisses, in dem Daten im Auftrag erstellt werden, soll zunächst der Auftragnehmer Berechtigter sein, bis er die Daten ausgehändigt

⁵⁹ Hilgendorf, JuS 1996, 890 (893).

⁶⁰ Hilgendorf, JuS 1996, 890 (892).

⁶¹ Hilgendorf, JuS 1996, 890 (893).

⁶² OLG Nürnberg, B. v. 23.01.2013 - 1 Ws 445/12; Welp, IuR 1988, 443 (447).

⁶³ Welp, IuR 1988, 443 (447).

hat.⁶⁴ Letzteres ist zwar umstritten,⁶⁵ im Ergebnis aber wohl richtig, denn zum einen ist so die eindeutige Zuordnung gewährleistet und zum anderen soll § 303a StGB nicht zu einer weiten Kriminalisierung von Vertragsbrüchen führen. Werden hingegen übermittelte Daten im Auftrag verarbeitet, soll der ursprüngliche Inhaber der Originaldaten verfügbefugt bleiben.⁶⁶

d) Verhältnis von Medien- und Dateneigentum

Maßgebendes Kriterium für die Entstehung einer originären Zuordnung der Daten könnte also der Skripturakt sein.⁶⁷ Wie oben bereits angedeutet, könnten damit das Eigentum an dem Speichermedium und das „Dateneigentum“ auseinander fallen. Dieses Ergebnis ist mit Hinblick auf Hosting-Verhältnisse und Auftragsverhältnisse auch einleuchtend. Wenn aber die Speicherung ohne oder gegen den Willen des Medieneigentümers erfolgt, kommt es zu einer Konfliktsituation der Rechte.⁶⁸ Fraglich ist, ob diese direkt auf der Ebene des Entstehens des Dateneigentums oder im Nachhinein zu lösen ist.

(1) Skribent als Dateneigentümer

Da das Sacheigentum zunächst von der Rechtsordnung stärker und unmittelbarer geschützt ist, liegt es nahe, dass der Eigentümer eines Datenspeichers es nicht uneingeschränkt dulden muss, wenn sein Eigentum durch die Rechte Dritter an Daten beeinträchtigt wird.⁶⁹ Es erscheint aber systemwidrig, ihn in allen Fällen der Verletzung seines Eigentums durch fremde Datenskriptur als alleinigen Dateninhaber anzusehen.⁷⁰ Der Skribent könnte dann nämlich nur durch Willensakt des Eigentümers Dateneigentum durch Erzeugung entstehen lassen. Somit wäre wiederum das Medieneigentum das grundlegende Zuordnungskriterium. Konsequent wäre es daher, die Dateninhaberschaft auch in dem Fall immer dem Skribenten zuzuordnen, in dem er vorsätzlich auf fremden Speichermedien Daten ablegt, dem Eigentümer jedoch gleichzeitig nur zivilrechtliche Unterlassungs- und Beseitigungsansprüche aufgrund seines Sacheigentums zuzuerkennen.⁷¹ Dass im Endeffekt die Rechtekollision wohl grundsätzlich über derartige Ansprüche zu lösen sein wird, ist jedenfalls systematisch folgerichtig, wenn man dem „Dateneigentum“ einen dinglichen Charakter zuspricht.

(2) Ausnahmen

Allerdings müssen von diesem Grundsatz wertungsgerechte Ausnahmen gelten. Es kann nicht sein, dass das Recht aus § 903 BGB, mit der Sache nach Belieben zu verfahren, durch eine reine Abspeicherung einer Datei auf einen Datenträger massiv beeinträchtigt werden kann, ohne dass der Eigentümer die Entstehung der Daten nicht zumindest mit veranlasst hat. Man denke nur daran, diesen Grundsatz auf in ein System eingeschleuste Daten

⁶⁴ OLG Nürnberg, B. v. 23.01.2013 - 1 Ws 445/12.

⁶⁵ Fischer, § 303a Rn. 6.

⁶⁶ ähnlich BayObLG, CR 1993, 779; Bär, in: Graf/Jäger/Wittig, § 303a StGB Rn. 13; Cornelius, in: Kilian/Heussen, Teil 10 Rn. 180; Wolff, in: Leipziger Kommentar, § 303a Rn. 10 ff.

⁶⁷ Vgl. auch Kühl, in: Lackner/Kühl, § 303a Rn. 4; Rengier, § 26 Rn. 7.

⁶⁸ Welp, IuR 1988, 443 (448).

⁶⁹ Welp, IuR 1988, 443 (448).

⁷⁰ Welp, IuR 1988, 443 (448).

⁷¹ Hilgendorf, JuS 1996, 890 (893).

anzuwenden. Dann müsste etwa der Nutzer eines PCs vor Entfernen einer Schadsoftware den Angreifer zuvor erfolgreich verklagen, wenn er sich nicht auf Rechtfertigungsgründe (§ 904 BGB) berufen kann. In Ausnahmefällen muss die Verfügungsbefugnis an Daten also auch originär beim Eigentümer des Mediums entstehen. Ein solcher Ausnahmefall ist immer dann anzunehmen, wenn der Eigentümer die Skriptur nicht selbst in irgendeiner Form mitveranlasst hat. Dabei kommt es auf seinen Willen nicht an. Eine Mitveranlassung scheidet aber zumindest dann aus, wenn etwa die Sache abhandengekommen ist (§ 935 BGB) oder die Skriptur durch Einbruch in ein System erfolgt ist. Letzteres gebietet schon das Grundrecht auf Gewährleistung der Vertraulichkeit und Integrität informationstechnischer Systeme, das die Verfügungsgewalt des Systeminhabers sicherstellen soll.⁷²

(3) Automatische Skriptur

Als problematisch könnte sich noch die bereits angesprochene automatische Skriptur erweisen. Im Grundsatz ist eindeutig, dass die Zuordnung zum Skribenten auch dann erfolgen muss, wenn dieser Programme ausführt, die Daten erstellen oder eine Dateneinspeisung selbst bewirken.⁷³ Das gilt auch für profan-physische Akte, wie dem Betreiben einer Mikrowelle mit Datenspeicher oder dem Auslösen einer Digitalkamera. Fraglich ist aber, wie viel Anteil an der Skriptur durch einen vorprogrammierten Prozess eine Person noch hat, die nur einen unwesentlichen Teil dieses Prozesses, aber damit das Resultat beeinflusst. Als Beispiel sei eine umfangreiche Datenbank genannt, in der durch einen Bearbeiter unter vielen einzelne Bestandteile verändert oder eingebracht werden. In solchen Fällen prozessual festgelegter Skriptur muss nach Wesentlichkeit des Beeinflussungsmoments abgegrenzt werden, wobei auch die Grundsätze zur Veranlassung und dem Medieneigentum eine Rolle spielen können.

e) Reichweite der Datenverfügungsbefugnis

Die Datenverfügungsbefugnis erfasst zunächst jedes einzelne Datum und auch zusammenhängende Datenstrukturen. Anders als bei Sachen kann jedoch die Eingrenzung eines Datums nicht einfach an einem Stoff festgemacht werden. Wie das oben besprochene Urteil des *EuGH* zeigt, ist es vertretbar, dass der rechtliche Zuweisungsgehalt von Daten nicht endet, wenn deren Inhalt vervielfältigt wird. Das zeigt die Gleichstellung der Weitergabe von Werkstücken in körperlicher und unkörperlicher Form. Fraglich ist also, ob die Kopie einer Datei ebenfalls in den Zuweisungsbereich der ursprünglichen Datei fällt?

Hilgendorf sieht jedenfalls bezüglich § 303a StGB die Kopie nicht von der Verfügungsbefugnis des Berechtigten erfasst.⁷⁴ Allerdings soll kein Eingriff in das Rechtsgut des § 303a StGB vorliegen, wenn eine „fremde“ Datei auf einem eigenen Medium gelöscht wird und der Berechtigte eine inhaltsgleiche Kopie zurück behalten hat.⁷⁵ Das lässt sich damit erklären, dass der Berechtigte in die Veränderung der weitergegebenen Version eingewilligt hat oder

⁷² Polenz, in: Kilian/Heussen, Teil 13 Rn. 36.

⁷³ Welp, IuR 1988, 443 (447).

⁷⁴ Hilgendorf, JuS 1996, 890 (890).

⁷⁵ Hilgendorf, JuS 1996, 890 (893).

dass die Versagung einer solchen Einwilligung in einem Auftragsverhältnis treuwidrig sein kann. Das Dateneigentum an Kopien kann nicht generell anhand der Kopie des Inhalts festgelegt werden. Insoweit ist die Zuweisung der Datenberechtigung getrennt von den Grundsätzen der Erschöpfung bei Werkstücken zu betrachten. Vielmehr gelten auch hier die Kriterien der Skriptur und der Veranlassung. Werden Daten in demselben System und auf demselben Datenträger „kopiert“, wird wohl nicht der Kopierende, sondern der ursprüngliche Ersteller der Berechtigte bleiben. Werden sie auf ein anderes Medium kopiert, ist nach dem Kopievorgang zu unterscheiden. Entweder werden Daten direkt auf ein Medium kopiert, das dann weitergegeben wird (z. B. eine CD-Rom), oder in einem Netz verschoben. In diesen Fällen ist Skribent der Ersteller der Kopie bzw. der Absender. Mit der Überlassung verfügt er aber über sein Recht an den Daten zugunsten des Empfängers, sodass dieser mit Entgegennahme die Berechtigung erwirbt. Werden die Daten allerdings zum Abruf bereitgestellt und direkt vom Empfänger selbst kopiert, dann ist jedenfalls er selbst Skribent und damit originär Berechtigter.

Die Verfügungsbefugnis endet mit der Aufgabe der Daten, also wenn der Berechtigte über einen längeren Zeitraum kein Interesse mehr an Ihnen zeigt.⁷⁶ Dann wird die Inhaberschaft wohl am ehesten dem Medieneigentümer zufallen.

f) Zwischenfazit

Im Gegensatz zum Zivilrecht, dem in Bezug auf die Frage nach dem Zuweisungsgehalt des Eigentums eine viel größere Rolle zukommt, gibt es im Strafrecht die Regelung des § 303a StGB, die sich explizit auf Daten als geschütztes Gut bezieht. Zwar spricht einiges dafür, die Zuordnung der Daten im Rahmen des § 303a StGB anhand des Skripturaktes vorzunehmen. Wie die vorherige Darstellung allerdings verdeutlicht hat, herrscht diesbezüglich eine erhebliche Rechtsunsicherheit. Dies gilt vor allem im Zivilrecht, weil es dort keine Norm gibt, die sich explizit mit Daten befasst. Deshalb ist es umso bedeutender, im bestehenden Geflecht der zivilrechtlichen Normen Lösungen zu finden. Einige Ansätze hierzu werden im Folgenden erläutert.

8. § 453 BGB

a) Zivilrechtliche Einordnung der Datenberechtigung

Natürlich muss eine Disposition der Datenberechtigung möglich sein.⁷⁷ Für das Strafrecht genügt es insoweit, dass der originäre Dateninhaber einen anderen zur Bearbeitung der Daten berechtigt. Mit der Berechtigung entfällt das Tatbestandsmerkmal der „rechtswidrigen“ Löschung, Unterdrückung, Unbrauchbarmachung oder Veränderung.

Im Sinne einer einheitlichen Rechtsordnung ist derjenige, der eine Datenverfügungsbefugnis nach § 303a StGB unter den oben ausgeführten Bedingungen erlangt hat, auch nach dem Zivilrecht als Berechtigter an den Daten anzusehen. Das zeigt schon die Analogie, die in der

⁷⁶ Hilgendorf, JuS 1996, 890 (894).

⁷⁷ Hilgendorf, JuS 1996, 890 (892).

strafrechtlichen Literatur zu § 903 BGB gezogen wird.⁷⁸ Man mag sich fragen, ob eine solche Analogie dem Grunde nach überhaupt gerechtfertigt ist. Sicher ist, dass der historische Gesetzgeber des BGB die Einbeziehung von Daten in den Kontext der zivilen Rechtsordnung nicht eingeplant hat.⁷⁹ Trotz der allgemeinen Unklarheit auf diesem Gebiet und dem Rückstand der Gesetzgebung hinter dem Fortschritt der Technik kann nicht davon ausgegangen werden, dass eine solche Lücke absichtlich nicht geschlossen wurde. Auch der Hinweis auf das *sui-generis*-Recht für Datenbankhersteller hilft insoweit, wie im diesbezüglichen Abschnitt dargelegt, einer Regelungslücke nicht ab.

Ob die Lücke durch eine Analogie zu Ausschließlichkeitsrechten überhaupt praxisgerecht und hinreichend geschlossen werden kann,⁸⁰ ist eine andere Frage, die es im Folgenden zu klären gilt. Gerade weil die Datenmenge exponentiell anwächst, wie ein Blick auf das Big-Data-Projekt belegt, und Beziehungen in digitalen Netzen immer komplexer werden, sind Antworten auf diesem Gebiet notwendig. Zwar mag es auf den ersten Blick konstruiert und theoretisch erscheinen, neben den ohnehin schon zahlreichen Ebenen (Sacheigentum, Urheberrecht, Schutz des Datenbankherstellers, Datenschutz, schuldrechtliche Beziehungen etc.) noch eine weitere Ebene des eigenständigen Dateneigentums hinzuzufügen. Für Menschen, die täglich mit digitaler Technik in Berührung kommen, sind Daten aber bereits zu einem selbstverständlichen, fühlbaren Wert geworden. Zudem wird es auch juristischen Laien einleuchten, dass Daten, die etwa auf Speichern im Internet liegen, nicht dem Eigentümer des Servers „gehören“ und dieser deshalb nicht nach Belieben damit verfahren kann.

b) Datenkauf

Für ein zivilrechtliches Schutzsystem für Dateneigentum ist zunächst von Interesse, ob die Datenberechtigung ausschließlich schuldrechtliche oder auch dingliche Wirkung hat, das heißt, ob und inwieweit sie zeitlich oder inhaltlich begrenzt bzw. auch wieder entzogen werden kann.

Spricht man von einem „Vollrecht“ an Daten,⁸¹ so geht man von einer dinglichen Berechtigung aus. Dieses muss auch als Vollrecht dinglich übertragbar sein. Daneben kann konsequenterweise auch das Recht zur Benutzung in gewissem Umfang schuldrechtlich eingeräumt werden.⁸² Letzteres ist im Rahmen der Privatautonomie in *sui-generis*-Verträgen (§ 311 Abs. 1 BGB) rechtlich ohne weiteres zulässig.

(1) Einräumung der Dateninhaberschaft als kaufähnlicher Vertrag

Grundlage für eine Verpflichtung zur Einräumung der vollen Dateninhaberschaft muss folgerichtig ein kaufähnlicher Vertrag sein. Im *Used-Soft*-Urteil verwendet der *EuGH* zur

⁷⁸ Hilgendorf, JuS 1996, 890 (890).

⁷⁹ Duisberg, S. 38.

⁸⁰ Duisberg, S. 38.

⁸¹ Welp, IuR 1988, 443 (448).

⁸² Stree/Hecker, in: Schönke/Schröder, § 303a Rn. 3.

Auslegung der Computerprogramme-Richtlinie einen weiten Verkaufsbegriff, der auch die Übertragung von Eigentum an unkörperlichen Softwarekopien umfasst.⁸³ Dieser Begriff ist zunächst einmal ein autonom richtlinienspezifischer Begriff.⁸⁴ Nach dem deutschen § 433 BGB verpflichtet sich im Normalfall des Kaufvertrags der Verkäufer, dem Käufer das Eigentum an einer Sache i. S. d. § 90 BGB zu verschaffen. Befindet sich auf einem Datenträger Software und wird der Datenträger verkauft, so handelt es sich immer noch um einen Sachauf nach § 433 BGB.⁸⁵ Hardware und Software bilden zusammen eine einheitliche Kaufsache.⁸⁶ Wenn aber Daten unkörperlich veräußert werden sollen, so muss § 453 Abs. 1, 2. Alt. BGB angewandt werden.⁸⁷

(2) Daten als „sonstige Gegenstände“ im Sinne des § 453
BGB

Nach herrschender Meinung fallen unter den Begriff der sonstigen Gegenstände im Sinne von § 453 BGB einzelne unkörperliche Vermögenswerte, wie z. B. Know-how, Erwerbs- oder Gewinnchancen sowie Informationen.⁸⁸ Auf Daten als „sonstige Gegenstände“ findet nach dieser Vorschrift daher Kaufrecht entsprechende Anwendung. Das ist insofern interessant, als „sonstige Gegenstände“ in der Systematik des Kaufrechts neben Sachen und Rechten genannt werden. Sachen und Rechte besitzen beide eine abstrakte rechtliche Zuweisung: An Sachen besteht Eigentum nach § 903 BGB, über Rechte kann nach den §§ 398, 413 BGB nur abstrakt verfügt werden.⁸⁹ „Sonstige Gegenstände“ sind wie Rechte unkörperliche Vermögenswerte; anerkannt sind unter anderem auch Ideen und Informationen.⁹⁰ Nun sind Ideen nicht fassbar genug, um eine dingliche Zuweisungsposition festzumachen. Informationen sollen gerade nicht über die entsprechenden Schutzrechte hinaus mit Ausschließlichkeitsrechten geschützt werden. Daher kommt zumindest bei diesen „sonstigen Gegenständen“ eine Verweisung auf die Eigentumsverschaffung in § 433 Abs. 1 S. 1 BGB nicht in Betracht, vielmehr schuldet der Verkäufer nur die „Übergabe“. Der Käufer, der sich ferner im „Besitz“ der Information oder Idee befindet, kann Dritte – außerhalb der Grenzen des Urheberrechts oder gewerblicher Schutzrechte – nicht von der Benutzung ausschließen. Anders verhält es sich aber bei Daten. Ein entsprechendes Schutzregime ist zwar im Zivilrecht nicht geregelt, jedoch zeigt sich in § 303a StGB deutlich eine Zuordnung von Daten zu einer Person, die Dritte von den aufgezählten Handlungen ausschließt. Nur der Datenberechtigte darf also mit seinen Daten, was ihre spezifische Eigenschaft angeht (also unabhängig von Sache und Information), „nach Belieben verfahren und andere von jeder Einwirkung ausschließen“. Die Analogie zu § 903 BGB ist gerechtfertigt, obwohl der Schluss von in § 453 Abs. 1 BGB aufgeführten „sonstigen Gegenständen“ auf eine dingliche

⁸³ EuGH, NJW 2012, 2566 (2568).

⁸⁴ EuGH, NJW 2012, 2566 (2567).

⁸⁵ Hoeren, IT-Recht, S. 105.

⁸⁶ Beckmann, in: Staudinger, § 453 Rn. 55.

⁸⁷ Beckmann, in: Staudinger, § 453 Rn. 53.

⁸⁸ Beckmann, in: Staudinger, § 453 Rn. 20.

⁸⁹ Berger, in: Jauernig, § 453 Rn. 3.

⁹⁰ Beckmann, in: Staudinger, § 453 Rn. 20.

Rechtsposition nicht zwingend ist. Anders als Ideen, Informationen oder Know-how⁹¹ sind Daten abgrenzbar und es ist möglich, zwischen Besitzer und „Eigentümer“ zu unterscheiden. Während dem Verkäufer mit der „Übergabe“ anderer „Gegenstände“ im Sinne des § 453 Abs. 1 BGB sämtliche Positionen bereits verliert, verschafft der alleinige Besitz von Daten noch nicht das Vollrecht. Das zeigt sich gerade darin, dass Dateninhaberschaft und Eigentum an dem Datenträger auseinander fallen können.

(3) Vergleich zur Problematik bei elektrischer Energie

Die Tatsache, dass auch Strom sowohl strafrechtlich geschützt (§ 248c StGB) als auch „sonstiger Gegenstand“ im Sinne des § 453 Abs. 1, 2. Alt. BGB ist,⁹² macht den Schluss vom Strafrecht auf das Zivilrecht nicht unzulässig. Zwar weist auch § 248c StGB elektronische Energie einer Person zu („fremde elektrische Energie“), allerdings ist Strom nicht bleibend physisch vorhanden, sondern verbraucht sich in dem Moment, in dem er entsteht. Elektrische Energie ist also ein Zustand einer Sache. Im Grunde ist die Übertragung von Strom der Übertragung von Daten nicht unähnlich: Der Zustand einer Sache (Akkumulator, Batterie) wird durch physikalische (oder chemische) Abläufe dazu gebracht, den Zustand der Sache eines Erwerbers (Batterie, elektrisches Gerät) zu verändern. Ähnliches geschieht bei der Übertragung von Daten, etwa über ein Kabel oder auch nur durch Schreiben auf einer Festplatte. Jedoch ist Strom immer nur profan die Energie, die durch die Sache erzeugt wird. Daten besitzen neben ihrer physikalischen Eigenschaft eben die zu Anfang aufgezeigte, abstrakte Ebene. Das Merkmal der „Fremdheit“ bezieht sich also bei elektrischer Energie nur auf die Befugnis zur Entnahme,⁹³ bei Daten aber auf die Inhaberschaft.

(4) Geschuldete Leistung

Folgte man diesem durchaus reizvollen Konzept, so wäre bei einem Datenkauf nach § 453 Abs. 1, 2. Alt. BGB also neben der „Übergabe“, mit Verweis auf § 433 Abs. 1 S. 1 BGB, auch die Verschaffung von „Dateneigentum“ geschuldet. Dies kann nur über einen abstrakten Verfügungsakt geschehen, der sich von der reinen tatsächlichen Überlassung, etwa durch Zugangsgewährung, unterscheidet. Zur Einräumung des Vollrechts an Daten böte sich analog § 929 S. 1 BGB mithin das Erfordernis einer Einigung in dinglicher Hinsicht und einer Überlassung an.

c) Überlassungsverträge

Die Parallele zum Trennungsprinzip bedeutet aber nicht, dass die Überlassung von Daten im Verkehr immer dadurch eingeschränkt werden kann, dass anstatt von Kaufverträgen sonstige Überlassungsverträge geschlossen werden, die die Inhaberschaft bei dem ursprünglichen Datenherrn belassen sollen. In Frage kommt diese Konstellation überhaupt nur dann, wenn Daten weder auf einem Datenträger (dann einheitlicher Kauf) überlassen werden, noch von dem Empfänger durch Kopie (etwa Download) selbst erstellt werden. Übrig bleiben Fälle, in denen der Überlassende die Daten selbst auf einem Medium des

⁹¹ Beckmann, in: Staudinger, § 453 Rn. 47.

⁹² Beckmann, in: Staudinger, § 453 Rn. 50.

⁹³ Eser/Bosch, in: Schönke/Schröder, § 248c Rn. 3-5.

Empfängers erstellt, also durch Eingabe oder Manipulation (letzteres auch nur dann, wenn nicht vom Empfänger veranlasst). Dann aber ist eine eingeschränkte Überlassung ohne Verfügung über das Vollrecht in Formularverträgen jedenfalls nur unter Wahrung des Vertragszwecks (§ 307 Abs. 2 Nr. 2 BGB) möglich, der in den meisten Fällen wohl die volle Übertragung vorsieht.⁹⁴

Dass die Verfügungsbefugnis über Daten auch durch Rechte am Dateninhalt beschränkt werden kann, versteht sich ebenfalls. Dateninhalte unterliegen genauso den Beschränkungen des Urheber- oder Datenschutzrechts wie analoge Medien. Daher kann die Ausübung von Eigentum an Daten, die personenbezogene Inhalte haben, durch die Vorschriften des Datenschutzrechts beschränkt sein.

9. § 950 BGB

Die Analogie zum dinglichen Eigentum kann selbstverständlich nicht grenzenlos sein. Das sachenrechtliche System dient nur insoweit der Einordnung von Dateneigentum, wie es wesensgemäß anwendbar und sachdienlich ist. Der starke zivilrechtliche Schutz des (Sach-) Eigentums im Gegensatz zu anderen Rechten, die nur relativ gelten, wird durch die zwingenden Vorschriften des Sachenrechts gewährleistet, das nicht ohne weiteres ausgedehnt werden darf. Daher ist das Sacheigentum dem Dateneigentum grundsätzlich vorrangig. Da Dateneigentum und Eigentum an dem Trägermedium auseinander fallen können, ist es nicht möglich, dass das Dateneigentum durch Verbindung mit dem Medium nach § 947 BGB auf dessen Eigentümer übergeht.

Die Rechtsprechung der Instanzgerichte sieht das Aufspielen von Daten auf einen Rechner als einen Verarbeitungsvorgang im Sinne von § 950 BGB an.⁹⁵ Dies wird von der Literatur zum Teil allerdings kritisiert, weil die Datenspeicherung auf dem Datenträger jederzeit wieder rückgängig gemacht werden kann.⁹⁶ Ein Urteil des *LAG Chemnitz* aus dem Jahr 2008 ist insofern für das Verhältnis von Daten zu einer Sache interessant. Das Gericht entschied darin, dass das Eigentum an einer Programmversion von Microsoft Outlook nach § 950 BGB auf den „Hersteller“ übergeht.⁹⁷ Durch Aufspielen des Programms auf ein Notebook würden die Daten verarbeitet, so dass Hersteller der Verarbeiter oder, im Falle eines Besitzdieners bzw. Besitzmittlungsverhältnisses, der (mittelbare) Besitzer sei.⁹⁸

Das Gericht scheint § 950 BGB zunächst ohne Bedenken auf Daten direkt anwenden zu wollen. Wenn aber das Aufspielen von Daten auf eine Festplatte die Daten grundsätzlich verarbeitet, entsteht wieder das Ergebnis, dass das Dateneigentum dem Sacheigentum folgt. In diesem Zusammenhang verweist an entscheidender Stelle das *LAG Chemnitz* auf ein Urteil des *OLG Karlsruhe* aus dem Jahre 1986. In dem Karlsruher Urteil ging es um den gesetzlichen

⁹⁴ Vgl. Hoeren, IT-Recht S. 105.

⁹⁵ OLG Karlsruhe, CR 1987, 19; LAG Chemnitz, MMR 2008, 416 (417).

⁹⁶ Füller, in: MüKo, § 950 Rn. 9.

⁹⁷ LAG Chemnitz, MMR 2008, 416.

⁹⁸ LAG Chemnitz, MMR 2008, 416 (417); Füller, in: MüKo, § 950 Rn. 18.

Eigentumserwerb an einer Diskette durch Verarbeitung.⁹⁹ Das OLG hatte damals festgestellt, dass grundsätzlich eine neue Sache im Sinne des § 950 BGB dadurch entstehen kann, dass ein Programm auf einer leeren Diskette erstellt wird.¹⁰⁰ Aus dem Urteil geht jedoch nicht eindeutig hervor, ob es dabei um das Eigentum an der Diskette oder an dem Programm geht. Selbst wenn man es so versteht, dass Eigentum an dem Programm entsteht, ist die maßgebliche Verarbeitungshandlung ausweislich der Entscheidung nicht das Speichern auf einer fremden Diskette, sondern vielmehr das Herstellen des Programms. Das Eigentum an der Diskette war nur für die Ermittlung des Indizes der *fremden Mittel zur Herstellung* („mit von ihr vorgelieferten Stoffen“) in einem Abhängigkeitsverhältnis relevant.¹⁰¹

Um die Trennung von Dateneigentum und Sacheigentum zu wahren, kann das alleinige Speichern auf einem fremden Medium nicht generell nach § 950 BGB zu einer Übertragung des Dateneigentums auf den Sacheigentümer führen. Wie oben ausgeführt, ist das Abspeichern nämlich gerade der die Datenherrschaft begründende Skripturakt. Eigentümer an den Daten würde danach etwa der Eigentümer eines Laptops, wenn er selbst die Kopie erstellt oder die Kopie unter Einigung mit dem Dateninhaber empfängt.

Dennoch sind die Grundsätze nicht gänzlich außer Acht zu lassen. Auch der Skripturakt kann in einem Verhältnis entstehen, das dem Abhängigkeitsverhältnis entspricht, welches Verarbeiter und Hersteller in § 950 BGB auseinander fallen lässt. In diesem Fall ist es sachgerecht, die Grundsätze von § 950 BGB entsprechend auf die originäre Entstehung von Dateneigentum anzuwenden und die Berechtigung der übergeordneten Person zuzusprechen.

10. Ausgestaltung des zivilrechtlichen Schutzes

Wenn also zivilrechtliches Dateneigentum als Ausschließlichkeitsrecht anerkannt wird, stellt sich die Frage, wie zivilrechtlicher Schutz ausgestaltet sein kann.

a) Herausgabe

Konstitutives Recht eines dinglichen Eigentümers ist das Recht auf Herausgabe, § 985 BGB. So kann auch der Dateneigentümer von dem Datenbesitzer in gewisser Weise Herausgabe verlangen. Allerdings gestaltet sich eine solche Konstruktion nicht unproblematisch. Liegen die Daten in einem für den Berechtigten grundsätzlich zugänglichen Speicher, könnte die Herausgabe dadurch realisiert werden, dass alle Zugangshindernisse (Verschlüsselung, Passwörter, Sperren etc.) zugunsten des Berechtigten entfernt werden. Eine Pflicht zur Herausgabe des Mediums müsste aber wohl daran scheitern, dass das Sacheigentum aus den oben genannten Erwägungen gegenüber dem Dateneigentum wohl vorrangig ist.

Es wurde bereits festgestellt, dass sich das Dateneigentum nicht auf eine Kopie erstreckt. Allerdings ist in einem Zustand, der das Dateneigentum verletzt, die Herausgabe in Form einer Kopie möglich. Zunächst klingt dieser Ansatz systemwidrig, denn die Kopie ist ein

⁹⁹ OLG Karlsruhe, CR 1987, 19 f.

¹⁰⁰ OLG Karlsruhe, CR 1987, 19 f.

¹⁰¹ OLG Karlsruhe, CR 1987, 19 f.

neues eigenständiges Datum, das bloß den gleichen Inhalt trägt. Das Besondere an digitalen Daten ist aber, dass sie beliebig und ohne Abnutzung vervielfältigt werden können. Verletzt werden kann zwar nur das ursprüngliche Datum, als Folge der Verletzung kann jedoch, um die Sacheigentümerschaft des Medieneigentümers nicht zu weit zu strapazieren, die Herausgabe einer Kopie anstatt des Mediums geltend gemacht werden.

Der Herausgabeanspruch der Kopie ähnelt insoweit dem Anspruch aus § 38 KUG. Der Datenberechtigte hat den Aufwand für die Herstellung der Kopie demnach zu tragen, er kann aber eventuell im Rahmen eines Schadensersatzanspruchs vom Verletzer Befreiung erlangen.

b) Unterlassung und Beseitigung

Im Bereich des Dateneigentums ist die Grenze zwischen Herausgabe auf der einen und Beseitigungs- und Unterlassungsansprüchen (§§ 1004, 823 BGB) auf der anderen Seite nicht eindeutig trennbar. So kann das Entfernen einer Zugangssperre sowohl Herausgabe als auch Beseitigung einer Beeinträchtigung bedeuten. Auch die Herausgabe einer Kopie kann als Beseitigungsanspruch als Folge einer Verletzung verstanden werden. Die unberechtigte Kopie stellt insofern eine Verletzung des Ausschließlichkeitsrechts dar, die über die Grenzen des Datums hinausgeht. Weitere Ausprägungen von Unterlassungs- und Beseitigungsansprüchen können etwa die Löschung oder das Rückgängigmachen einer Veränderung darstellen.

c) Deliktsrecht

Folgt man der hier angedachten Lösung und qualifiziert das Dateneigentum als quasidingliches Recht mit Ausschlussfunktion, so ist es auch als „sonstiges Recht“ im Sinne des § 823 Abs. 1 BGB zu qualifizieren.¹⁰² Verletzungen können dann eine Schadensersatzpflicht nach § 823 Abs. 1 BGB, sowie gegebenenfalls nach § 823 Abs. 2 BGB i. V. m. § 303a StGB auslösen. Damit wäre auch die angedachte Schöpfung eines Rechts am eigenen Datenbestand als sonstiges Recht i. S. d. § 823 Abs. 1 BGB nicht zwingend erforderlich.

d) Entgegenstehende Rechte

Das Recht am Dateneigentum kann mit dem Recht am Trägermedium und mit Rechten am Dateninhalt kollidieren. Als Schnittstelle zwischen Medium und Inhalt unterliegen Daten deren rechtlichen Beziehungen. Ist also aufgrund eines Rechts am Inhalt die Einwirkung auf eine Sache zu gestatten, die Sache herauszugeben oder Ähnliches, so gilt das auch für Daten. Andersherum kann sich etwa ein Urheberrecht auch dann erschöpfen, wenn ein Vervielfältigungsstück nur in Form von Daten in den Verkehr gebracht wird.¹⁰³ Das Recht an Daten kann nur in Ausnahmefällen zur Herausgabe des Trägermediums berechtigen.

Ansprüche aus Dateneigentum sind nicht zuletzt immer subsidiär gegenüber sonstigen sachenrechtlichen Ansprüchen sowie zu Ansprüchen, die die Information selbst betreffen,

¹⁰² Vgl. Wagner, in: MüKo, § 823 Rn. 143.

¹⁰³ Siehe oben Kapitel „A. 2. b.“ zur *Used-Soft*-Entscheidung des EuGH.

also etwa solche aus dem klassischen Immateriagüter- oder dem Datenschutzrecht. Nur wenn diese nicht zum Tragen kommen, muss auf die spezifische Eigenschaft von Daten zurückgegriffen werden.

11. Fazit

Die rechtliche Zuordnung von Daten kann auf verschiedenen Wegen erfolgen, beispielsweise durch die enge rechtliche Bindung an das Speichermedium oder durch die Ausdehnung des urheberrechtlichen Erschöpfungsgrundsatzes. Alle Möglichkeiten müssen sorgfältig und umfassend auf ihre Umsetzbarkeit geprüft werden. Besonders vielversprechend erscheint allerdings, ein Dateneigentum durch Analogie zu § 903 BGB zu konstruieren. Die Frage ist, ob sich eine solche Analogie durch die Erforderlichkeit rechtfertigt, eine planwidrige Regelungslücke zu schließen. Da sich aber im Hinblick auf manche Fragen der Rechtsdogmatik das Problem der Selbstständigkeit von Daten nicht umgehen lässt, ist der hier aufgezeigte Ansatz eine Möglichkeit, sich der Beziehung zwischen Datenträger, Dateninhalt und Daten zu nähern. Die Reichweite dieser Analogie bedarf dabei ebenso der näheren Untersuchung wie die damit einhergehenden Kollisionen mit Rechten anderer. Daten sind im Vergleich zu Sachen nicht einfach stofflich abgrenzbar. Auch in der spezifischen Verletzung und in der Beweisbarkeit von Aspekten des Dateneigentums liegen Probleme. Mit ähnlichen Problemen haben aber auch andere Schutzrechte zu kämpfen. Die aufgezeigten Kriterien zur Zuordnung von Dateneigentum können in der Praxis einige dieser Probleme lösen.

Mit der zunehmenden Ubiquität von Rechenprozessen¹⁰⁴ werden Daten immer mehr zu fühlbaren Gegenständen. Ein Schutz des Inhalts über die Grenzen des Urheberrechts hinaus wäre ein unangemessen scharfes Schwert, die Rechtsordnung muss aber auf die Interessen an gesellschaftlichen Werten Rücksicht nehmen. Dies wurde durch die Einführung des § 303a StGB für das Strafrecht bereits erkannt. Da Daten Vermögenswerte sein können, muss auch die zivile Rechtsordnung einen entsprechenden Schutz bereithalten. Auch wenn es nach hier vertretener Auffassung gute Gründe für die Existenz eines solchen Rechtsinstituts gibt, ist festzuhalten, dass es nach heutigem Stand noch völlig unklar ist, ob die Rechtsfigur eines „Dateneigentums“ existiert und ob sich diese in Zukunft wird durchsetzen können. Hierzu erscheinen eingehendere Untersuchungen dringend geboten, um den in der Informationsgesellschaft Tätigen die notwendigen rechtlichen Rahmenbedingungen zur Verfügung zu stellen.

B. Insolvenzrecht

Regelmäßig werden Drittanbieter von Firmen nicht nur mit der technischen Speicherung von Daten beauftragt, sondern im Rahmen des Outsourcings auch mit der Abwicklung ganzer Dienstleistungsvorgänge, z.B. mit dem E-Mail-Marketing.¹⁰⁵ Kommt es bei einem der

¹⁰⁴ Weiser, S. 3.

¹⁰⁵ OLG Düsseldorf, U. v. 27.09.2012 - I-6 U 241/11 = NJW-Spezial 2012, 759; Heckmann, jurisPR-ITR 23/2012 Anm. 2.

beteiligten Big-Data-Unternehmen zum Eintritt des Insolvenzfalles, ist fraglich, was mit den Daten geschieht, die sich im Machtbereich des insolventen Unternehmens befinden.

In diesem Zusammenhang hätte die Beantwortung der im vorhergehenden Abschnitt aufgeworfenen Frage nach dem Eigentum an Daten direkte Auswirkungen auf das Schicksal der Daten in der Insolvenz. Würde das Institut des „Dateneigentums“ anerkannt, könnten nämlich den Insolvenzverwalter eigene Aussonderungs- und Herausgabepflichten treffen.¹⁰⁶ Daneben ist ebenfalls zu klären, auf welcher Grundlage sich ansonsten eine Aussonderungs- und Herausgabepflicht ergeben könnte. Besonderes Augenmerk ist zudem auf eine insolvenzfeste Vertragsgestaltung zu legen, die keinen Zweifel daran lässt, ob Daten vom Auftragnehmer erhalten (§ 667 1. Alt. BGB) oder erlangt (§ 667 2. Alt. BGB) wurden; dieser Unterscheidung kommt im Zweifel zentrale Bedeutung für eine etwaige Aussonderungspflicht zu.¹⁰⁷

Wichtig ist aus datenschutzrechtlicher Sicht zudem die Frage, wie die im Machtbereich des insolventen Unternehmens befindlichen Daten verwertet werden dürfen. Im Falle der „Quelle-Insolvenz“ beispielsweise sah das zuständige Landesamt für Datenschutzaufsicht keine rechtlichen Probleme bezüglich des Verkaufs der Daten.¹⁰⁸ Solange ein Widerspruch des Betroffenen nicht vorliege, könnten personenbezogene Daten übermittelt werden. Ob diese Auffassung haltbar ist, bleibt jedoch abzuwarten. Es ist nicht ausgeschlossen, dass im Insolvenzfall den im Nachhinein abgesendeten Widerspruchsanträgen überhaupt nicht mehr nachgekommen wird. Jedenfalls sollte sich der von der Datenverarbeitung Betroffene bereits beim Vertragsschluss darüber Gedanken machen, dass er die Möglichkeit hat, die Übermittlung seiner Daten im Insolvenzfall zu verhindern. Ob trotz Widerrufs der Einwilligung dann noch wegen Eingreifens gesetzlicher Erlaubnisgründe eine Übermittlung zulässig ist, ist bislang nicht geklärt.¹⁰⁹ Klarer scheint dagegen die Situation für Unternehmen, die Daten an Geschäftspartner übermitteln, damit diese ihrerseits vertraglich versprochene Dienstleistungen erfüllen können. Im Insolvenzfall des Geschäftspartners jedenfalls besteht nach Ansicht des *OLG Düsseldorf* die Pflicht des Insolvenzverwalters, die überlassenen Daten auszusondern und herauszugeben.¹¹⁰

Insolvenzrechtlich problematisch ist der Fall, dass ein Big-Data-Unternehmen Software im Wege des „Cloud Computing“ bezieht. Diese Verträge werden grundsätzlich nach mietrechtlichen Vorschriften beurteilt. Als Dauerschuldverhältnis richtet sich das weitere Verfahren nach § 103 InsO.¹¹¹ Danach kann der Insolvenzverwalter wählen, ob er in den Mietvertrag eintritt oder die Erfüllung verweigert. Im letzteren Falle sind die Ansprüche des

¹⁰⁶ Bäuerle, in: Braun, § 47 Rn. 20.

¹⁰⁷ Heckmann, jurisPR-ITR 23/2012 Anm. 2.

¹⁰⁸ SZ vom 17.05.2010, abrufbar unter <http://www.sueddeutsche.de/wirtschaft/ausverkauf-bei-quelle-kundendaten-auf-dem-wuehltisch-1.131329> (zuletzt abgerufen am 6.9.2013).

¹⁰⁹ Gola/Schomerus, § 4 Rn. 16.

¹¹⁰ OLG Düsseldorf, U. v. 27. 9. 2012 - I-6 U 241/11; NJW-Spezial 2012, 759.

¹¹¹ Selk, ITRB 2012, 201 (201).

Vertragspartners nicht mehr durchsetzbar.¹¹² Der Kunde könnte die Software also ggf. innerhalb kürzester Zeit nicht mehr nutzen. Um zumindest dennoch bis zum Ende der normalen Kündigungsfristen die Weiternutzung zu ermöglichen, sollte im Rahmen der Vertragsgestaltung darauf geachtet werden, die urheberrechtliche Nutzungsrechteeinräumung im Sinne von § 31 UrhG dinglich auszugestalten. Diese Lösung zugunsten des Softwarenutzers versagt jedoch, wenn die Software nur noch per Fernzugriff über das Internet genutzt wird. Ist nämlich weder die Installation der Software auf der Hardware des Nutzers, noch ein Laden in den Arbeitsspeicher erforderlich, so erfolgt bei diesem keinerlei urheberrechtlich relevante Nutzungshandlung.¹¹³

Im Rahmen zukünftiger Forschung gilt es die Auswirkungen der Insolvenz auf das Big-Data-Management zu untersuchen. Darüber hinaus ist ein besonderes Augenmerk auf die Frage zu richten, wie Nutzungsrechte zweiter oder späterer Stufe im Falle der Insolvenz des Einräumenden zu behandeln sind. Relevant ist dies, wenn ein Big-Data-Unternehmen ein Nutzungsrecht an urheberrechtlich geschützter Software erhält, das vom Inhaber eines ausschließlichen Nutzungsrechts und nicht vom Urheber selbst eingeräumt wird.

Die Schlüsselfrage stellt auch hier die dingliche Wirkung und damit einhergehend die potentielle Insolvenzfestigkeit des eingeräumten Nutzungsrechts späterer Stufe dar. In der Entscheidung „Reifen Progressiv“ aus dem Jahr 2009¹¹⁴ entschied der BGH, dass das einfache Nutzungsrecht, ebenso wie das ausschließliche Nutzungsrecht, keinen schuldrechtlichen, sondern dinglichen Charakter aufweist.¹¹⁵ Das Enkelrecht sei nach seiner Abspaltung vom Tochterrecht von dessen Fortbestand unabhängig. Dieses Urteil führte dazu, dass teilweise von der Insolvenzfestigkeit abgeleiteter Nutzungsrechte als dingliche Rechte im Sinne des § 47 InsO ausgegangen wurde.¹¹⁶ Abschließend geklärt wurde diese Frage bislang jedoch nicht.

In den Parallelentscheidungen „M2Trade“¹¹⁷ und „Take Five“¹¹⁸ vom 19.07.2012 berief sich der BGH ausdrücklich auf seine in „Reifen Progressiv“ vertretene Linie. Ein Bruch in der Kette der eingeräumten Nutzungsrechte führe in der Regel nicht zum Erlöschen der eingeräumten Nutzungsrechte späterer Stufe.¹¹⁹ Allerdings erwähnte das Gericht den dinglichen Charakter der einfachen Lizenz nicht mehr.¹²⁰ Er stützte seine Ansicht vielmehr auf eine umfassende Interessensabwägung, die im Ergebnis zum Schutz des Inhabers des Nutzungsrechts späterer Stufe ausfiel. Künftig ist daher ausführlich zu klären, welche Kriterien im Rahmen dieser Interessensabwägung Berücksichtigung finden und ob sich eine Tendenz zugunsten des Inhabers des Nutzungsrechts späterer Stufe abzeichnen wird. Zudem ist zu klären, ob

¹¹² Kreft, in: MüKO, § 103 InsO Rn. 16.

¹¹³ Selk, ITRB 2012, 201 (203).

¹¹⁴ BGH, U. v. 26.3.2009, I ZR 153/06 = GRUR 2009, 946.

¹¹⁵ BGH, U. v. 26.3.2009, I ZR 153/06 = GRUR 2009, 946 (948).

¹¹⁶ Scholz, GRUR 2009, 1107 (1111).

¹¹⁷ BGH, U. v. 19.7.2012, I ZR 70/10 = GRUR 2012, 916.

¹¹⁸ BGH, U. v. 19.7.2012, I ZR 24/11 = GRUR 2012, 914.

¹¹⁹ Vgl. den zweiten amtlichen Leitsatz in BGH, U. v. 19.7.2012, I ZR 70/10 = GRUR 2012, 916.

¹²⁰ Trimborn, MarkenR 2012, 460 (462).

aufgrund dieser Interessensabwägung stets von einer „Insolvenzfestigkeit“ der Nutzungsrechte späterer Stufe auszugehen ist.

C. Urheberrecht

Der Umgang mit Big Data wirft eine Vielzahl urheberrechtlicher Fragen auf. Berührungspunkte von Big-Data-Management und Urheberrecht bestehen dabei auf verschiedenen Ebenen. Eine allgemeingültige und für jede Big-Data-Lösung einheitliche Begutachtung urheberrechtlicher Fragen ist nicht möglich. Die konkreten Probleme ergeben sich stets aus der Ausgestaltung des Verfahrens im Einzelfall. Im folgenden Abschnitt werden die urheberrechtlichen Grundlagen sowie deren konkreter Bezug zum Umgang mit Big Data dargestellt. Hierbei wird auf ungeklärte Rechtsfragen und sich daraus ergebende Forschungsperspektiven hingewiesen.

1. Anwendbarkeit des UrhG

Bei Big-Data-Lösungen stellt sich zunächst die Frage, ob das deutsche Urheberrechtsgesetz (UrhG) Anwendung findet. Scheidet die Anwendbarkeit aus, kommt es auf sich anschließende Fragen wie die Schutzfähigkeit von Datenbanken, betroffene Verwertungsrechte und die Auslegung urheberrechtlicher Schranken des nationalen Urheberrechts nicht mehr an. Die Anwendbarkeit des UrhG hängt von der Big-Data-Lösung im Einzelfall ab.

Der räumliche Anwendungsbereich des UrhG ist durch das Territorialitätsprinzip auf das Gebiet der Bundesrepublik Deutschland beschränkt.¹²¹ Demzufolge findet in Deutschland grundsätzlich kein ausländisches Recht Anwendung. Umgekehrt findet deutsches Recht im Ausland keine Anwendung. Die weltweite Nutzung und Verbreitung urheberrechtlich geschützter Werke schränkt die territoriale Wirkung des UrhG ein. Nach dem Schutzlandprinzip findet das Recht des Staates Anwendung, für dessen Gebiet Schutz beansprucht wird.¹²² Der persönliche Anwendungsbereich des UrhG richtet sich nach den §§ 120–128 UrhG, die an die Staatsangehörigkeit des Betroffenen anknüpfen. Der urheberrechtliche Schutz gebührt zunächst deutschen Staatsangehörigen, § 120 Abs. 1 S. 1 UrhG. Ausländer genießen den Schutz des UrhG, wenn das Werk im räumlichen Geltungsbereich des UrhG erscheint oder sich der Schutz aus Staatsverträgen ergibt.

Diese Grundsätze sind bei der rechtlichen Begutachtung der verschiedenen Konstellationen beim Umgang mit Big Data zu beachten. Neben der Frage, ob sich das datenverarbeitende Unternehmen selbst auf das UrhG berufen kann, kann es auch auf die Frage ankommen, ob die verarbeiteten Daten fremder Personen dem nationalen urheberrechtlichen Werkschutz unterfallen. Grundsätzlich wird bei Big-Data-Lösungen, die innerhalb Deutschlands genutzt werden, auch deutsches Urheberrecht zu beachten sein.

¹²¹ BGH, Urt. v. 16.06.1994 - I ZR 24/92 = GRUR 1994, 798; Dreier, in: Dreier/Schulze, Vor § 120 Rn. 1; Kotthoff, in: Dreyer/Kotthoff/Meckel, § 120 Rn. 4; v. Welser, in: Wandtke/Bullinger, Vor §§ 120 ff. Rn. 5.

¹²² BGH, Urt. v. 15. 2. 2007 - I ZR 114/04 = GRUR 2007, 871; Kotthoff, in: Dreyer/Kotthoff/Meckel, § 120 Rn. 8; v. Welser, in: Wandtke/Bullinger, Vor §§ 120 ff. Rn. 4.

Dennoch ergeben sich bei Internet-Sachverhalten, etwa im Bereich des Cloud-Computing, aufgrund der Ubiquität des Internets und der Tatsache, dass sich Handlungen nicht stets einem bestimmten Schutzterritorium zuordnen lassen, kollisionsrechtliche Fragestellungen.¹²³ Die Anwendung des Schutzlandprinzips auf Datenbankabfragen im Internet würde dazu führen, dass sich Urheberrechtsverletzungen nach einer Vielzahl von Rechtsordnungen bestimmen lassen.¹²⁴ Ein Festhalten am Schutzlandprinzip führt daher zu wenig praktikablen Lösungen.¹²⁵ Ebenso wenig liefern die Anknüpfung an die Rechtsordnung des Serverstandortes oder die „Bogsch“-Theorie, die im Rahmen des Senderechts auch das Recht des Empfangslandes für anwendbar erklärt,¹²⁶ in allen Fällen gerechte Ergebnisse. Mangels einer höchstrichterlichen Entscheidung ist die Bestimmung des anwendbaren Rechts bei Internet-Sachverhalten bislang umstritten und eröffnet eine für Big-Data-Anwendungen bedeutsame Forschungsperspektive. Angesichts der Widersprüchlichkeit der territorial beschränkten Anknüpfungsregel des Schutzlandprinzips und der Globalität des Internet empfiehlt es sich, nach anderen Anknüpfungsmöglichkeiten zu suchen.¹²⁷ Andernfalls wird die bestehende Rechtsunsicherheit die Entwicklung des Big-Data-Managements, das damit verbundene Innovations- und Investitionspotenzial sowie die wirtschaftliche Leistungsfähigkeit des Standorts Deutschland hemmen.

2. Urheberrecht an und Werkqualität von Daten

Die urheberrechtliche Beurteilung einer Big-Data-Lösung erfordert insbesondere eine Auseinandersetzung mit dem Datenbankrecht. Möglicherweise erfolgt der Datenumgang mit Hilfe eines Datenbankwerks oder es greift der sui-generis-Datenbankschutz. Beachtenswert erscheint aber auch die Frage, ob die Daten, mit denen im Rahmen des Big-Data-Managements umgegangen wird, möglicherweise selbst urheberrechtlich geschützt sind. Entgegenstehende Urheberrechte könnten den Datenumgang zusätzlich erschweren.

Das Urheberrecht schützt Werke der Literatur, Wissenschaft und Kunst, insbesondere in den in § 2 Abs. 1 UrhG genannten Formen. Es muss sich um persönliche geistige Schöpfungen handeln, die über eine ausreichende Gestaltungshöhe verfügen, § 2 Abs. 2 UrhG. Das Merkmal der Gestaltungshöhe erfordert dabei einen bestimmten Mindestgrad an Individualität des Werkes und dient folglich der Aussonderung einfacher Alltagserzeugnisse, die keinen Urheberrechtsschutz genießen.¹²⁸ Wann eine ausreichende Gestaltungshöhe erreicht ist, ist umstritten und differiert zwischen den einzelnen Werkarten.¹²⁹ Der Rechtsprechung folgend, darf ebenjener Grad an Gestaltungshöhe jedoch nicht zu gering angesetzt werden, so dass das Werk eine erhebliche individuelle Prägung besitzen muss.¹³⁰

¹²³ Nägele/Jacobs, ZUM 2010, 281 (284 f).

¹²⁴ Hoeren, in: Hoeren/Sieber, Teil 7.8 Rn. 33.

¹²⁵ Nägele/Jacobs, ZUM 2010, 281 (284).

¹²⁶ Adolphsen/Mutz, GRUR Int 2009, 789 (794); Hoeren, in: Hoeren/Sieber, Teil 7.8 Rn. 39; Nägele/Jacobs, ZUM 2010, 281 (285).

¹²⁷ Hoeren, in: Hoeren/Sieber, Teil 7.8 Rn. 35.

¹²⁸ Bullinger, in: Wandtke/Bullinger, § 2 Rn. 23.

¹²⁹ Bullinger, in: Wandtke/Bullinger, § 2 Rn. 24f.

¹³⁰ Bullinger, in: Wandtke/Bullinger, § 2 Rn. 24.

Eine solche kommt einzelnen Daten regelmäßig nicht zu.¹³¹ Denn Daten und Fakten, wie sie z.B. in Datenbanken gespeichert werden, erfüllen nicht die Voraussetzungen des § 2 Abs. 2 UrhG, weil sie entweder nur eine grundsätzlich freizuhaltende Tatsache enthalten oder ihnen die erforderliche Individualität fehlt.¹³² Entscheidend ist auch nicht, ob es sich um fremde oder eigene Daten handelt, mit denen im Rahmen des Big-Data-Managements umgegangen wird. Urheberrechtsverstöße sind beim Umgang mit einzelnen Daten deshalb zunächst nicht zu befürchten.

Die Big-Data-Lösung muss je nach ihrer konkreten Ausgestaltung auch fremde Urheberrechte beachten. Dies kann etwa bei der Auswertung nutzergenerierter Inhalte (UGC)¹³³ Dritter in sozialen Netzwerken, denen unter Umständen Schutz als Sprachwerk (§ 2 Abs. 1 Nr. 1 UrhG) zukommt, der Fall sein.¹³⁴ Gleiches gilt für nach § 72 UrhG geschützte Lichtbilder oder Lichtbildwerke i. S. d. § 2 Abs. 1 Nr. 5 UrhG, die von Nutzern sozialer Netzwerke eingestellt werden. Die Auswertung solcher Werke wird dann regelmäßig das Vervielfältigungsrecht und das Recht der öffentlichen Zugänglichmachung berühren.¹³⁵

Erfolgt im Rahmen der Big-Data-Lösung ein Zugriff auf fremde Datenbestände um deren Inhalte auszuwerten, hängt die Zulässigkeit des Vorgehens zunächst davon ab, ob die fremden Datenbestände urheberrechtlichem Schutz unterliegen. Datenbanken können gem. § 4 Abs. 2 UrhG als Datenbankwerke geschützt sein oder gem. §§ 87a ff. UrhG dem sui-generis-Schutz für Datensammlungen unterfallen.

a) Datenbankwerk

Ein Datenbankwerk nach § 4 Abs. 2 UrhG ist ein Sonderfall des Sammelwerks, dessen Elemente systematisch oder methodisch angeordnet und einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich sind. Ein Schutz als Datenbankwerk kommt nur in Betracht, wenn die Auswahl oder Anordnung der enthaltenen Elemente auf einer schöpferischen Leistung beruht.¹³⁶ Eine schöpferische Leistung kann insbesondere dann angenommen werden, wenn hinsichtlich der Auswahl der aufzunehmenden Daten ein Entscheidungsspielraum besteht.¹³⁷ Unter Umständen kann daher sogar Webseiten urheberrechtlicher Schutz zukommen.¹³⁸ Es ist im Einzelfall zu beurteilen, ob die Big-Data-Anwendung beim Zugriff auf fremde Datenbestände fremde Datenbankwerke berührt. In den meisten Fällen wird jedoch ein Zugriff auf Datenbestände erfolgen, deren Inhalt nach

¹³¹ Schulze, in: Dreier/Schulze, § 2 Rn. 130; Spindler/Hillegoist, in: Handbuch Forschungsdatenmanagement, S. 63 f.

¹³² Nordemann, in: Fromm/Nordemann, § 2 Rn. 76; Schulze, in: Dreier/Schulze § 20 Rn. 130.

¹³³ Siehe dazu auch Kapitel „L. 2.“.

¹³⁴ Solmecke/Wahlers, ZD 2012, 550 (553 f.).

¹³⁵ Solmecke/Wahlers, ZD 2012, 550 (554).

¹³⁶ Dreier, in: Dreier/Schulze, § 4 Rn. 1; Marquardt, in: Wandtke/Bullinger, § 4 Rn. 8; Czychowski, in: Fromm/Nordemann, § 4 Rn. 13.

¹³⁷ Czychowski, in: Fromm/Nordemann, § 4 Rn. 13; Kotthoff, in: Dreyer/Kotthoff/Meckel, § 4 Rn. 8.

¹³⁸ Vgl. zum Streit Marquardt, in: Wandtke/Bullinger, § 4 Rn. 14 ff; Kotthoff, in: Dreyer/Kotthoff/Meckel, § 4 Rn. 17; Schack, MMR 2001, 9.

logischen Gesichtspunkten geordnet und auf Vollständigkeit ausgerichtet ist, so dass jedenfalls der Schutz als Datenbankwerk nicht entgegensteht.¹³⁹

b) Sui-generis-Datenbankschutz

Wesentlich häufiger als der Schutz einer Datensammlung als Datenbankwerk wird der sui-generis-Schutz für Datenbanken gem. §§ 87a ff. UrhG greifen. Bereits bei der Frage des Eigentums an Daten wurde auf den urheberrechtlichen Datenbankschutz eingegangen. Die wesentlichen Rechtsprobleme sollen in der Folge noch einmal kurz skizziert werden.

Generiert die Big-Data-Anwendung Informationen aus fremden Datenbeständen, ist stets zu prüfen, ob sie das Recht des Datenbankherstellers berührt. Dieses gewährt einen Investitionsschutz für denjenigen, der zwar keine eigene schöpferische Leistung erbracht hat, jedoch das Investitionsrisiko für die Erstellung einer Datenbank trägt.¹⁴⁰ In Abgrenzung zum Datenbankwerk fordert der sui-generis-Datenbankschutz für die systematische und methodische Anordnung der Elemente keine besondere Schöpfungshöhe, sondern eine wesentliche Investition.¹⁴¹ Das sui-generis-Recht des Datenbankherstellers hat von Beginn an zu schwierigen Interpretationsfragen geführt.¹⁴² Die Frage, ob eine Investition als nach „Art oder Umfang wesentlich“ zu bezeichnen ist, muss durch eine Abwägung im Einzelfall und eine Gesamtbetrachtung aller Umstände ermittelt werden.¹⁴³ Nach h. M. statuiert der unbestimmte Begriff der „Wesentlichkeit“ eine niedrige Schutzzschwelle, die nur ganz unbedeutende Investitionen ausschließen will.¹⁴⁴ Datensammlungen unterfallen daher regelmäßig dem sui-generis-Schutz. Dies gilt insbesondere für soziale Netzwerke und ähnliche Plattformen mit hohen Hosting-Kosten.¹⁴⁵

Etwas anderes gilt jedoch dann, wenn die wesentlichen Kosten schon im Rahmen der Datenerhebung und nicht erst im Rahmen der Verwaltung der Datenbank anfallen.¹⁴⁶ Diese Unterscheidung zwischen der Erstellung und dem Betrieb der Datenbank einerseits und der Erzeugung der Daten andererseits traf der EuGH in seiner „The British Horseracing Board“-Entscheidung¹⁴⁷. Das Gericht hielt zur Frage, welche Investitionen im Rahmen des § 87a UrhG berücksichtigungsfähig sind, fest, dass zwar die Erstellung der Datenbank als solche, nicht aber die Erzeugung von Daten zu berücksichtigen sei. Nach diesem „Zweiphasenmodell“ sind nur die Investitionen für die Datenstrukturierung, allerdings nicht diejenigen für die Datengenerierung, berücksichtigungsfähige Investitionen.

¹³⁹ Spindler/Hilleggeist, in: Handbuch Forschungsdatenmanagement, S. 64; Zieger/Smirra, MMR 2013, 418 (420).

¹⁴⁰ Dreier, in: Dreier/Schulze, Urheberrecht, § 4 Rn. 3; Thum, in: Wandtke/Bullinger, Vor §§ 87a ff. Rn. 1.

¹⁴¹ Vogel, in: Schricker/Loewenheim, § 87a Rn. 26; Dreier, in: Dreier/Schulze, Vor §§ 87a ff. Rn. 1.

¹⁴² Hoeren, MMR 2005, 29 (35).

¹⁴³ Czychowski, in: Fromm/Nordemann, § 87a Rn. 15; Dreier, in: Dreier/Schulze, § 87a Rn. 15.

¹⁴⁴ Thum, in: Wandtke/Bullinger, § 87a Rn. 55 ff; Dreier, in: Dreier/Schulze, § 87a Rn. 14; Gaster, Der Rechtsschutz von Datenbanken Rn. 476.

¹⁴⁵ Zieger/Smirra, MMR 2013, 418 (420).

¹⁴⁶ Spindler/Hilleggeist, in: Handbuch Forschungsdatenmanagement, S. 64.

¹⁴⁷ EuGH, Urt. v. 9. 11. 2004 - C-203/02 The British Horseracing Board Ltd u.a./William Hill Organization Ltd = GRUR 2005, 244.

Die Abgrenzung der Datenerzeugungsphase von der Datenbankaufbereitungsphase ist in vielen Fällen praktisch nicht möglich. Die Datenerzeugung kann mit der Datenbankerzeugung zusammenfallen oder zu den ohnehin anfallenden Aufgaben des Datenbankherstellers zählen. Das Urteil des *EuGH* lässt viele Fragen offen, deren Lösung auch im Hinblick auf das Big-Data-Management von enormer Bedeutung ist. Ein Eingriff in das *sui-generis*-Leistungsschutzrecht durch die Big-Data-Anwendung kommt nur dann in Betracht, wenn im Hinblick auf die jeweilige Quelle auch der Anwendungsbereich der §§ 87a ff. UrhG eröffnet ist. Gerade dies kann aber aufgrund der unklaren gesetzgeberischen Vorgaben und der Rechtsprechung des *EuGH* in vielen Fällen nicht zuverlässig beurteilt werden. Die Klärung dieser Rechtsfrage würde ein erhebliches Problem in Bezug auf das Big-Data-Management lösen und die Rechtssicherheit auf diesem Gebiet in erheblichem Maße erhöhen.

Der Schutzmfang des *sui-generis*-Schutzes bestimmt sich nach § 87b UrhG. Dem Rechteinhaber stehen das Vervielfältigungs- und Verbreitungsrecht sowie das Recht zur öffentlichen Wiedergabe zu. Satz 1 beschränkt die Rechte hierbei zunächst auf die Nutzung der gesamten Datenbank sowie eines wesentlichen Teils einer Datenbank. Nach § 87c UrhG werden dem Schutz durch § 87b UrhG bestimmte Schranken gesetzt. In diesen Fällen übersteigt das allgemeine bzw. öffentliche Interesse an der Datenbanknutzung regelmäßig das Amortisationsinteresse des Datenbankherstellers. Im Falle des Big-Data-Managements wird keine der in § 87c UrhG genannten Schranken eingreifen. Da der Datenbankhersteller das ausschließliche Recht hat, die Datenbank insgesamt oder einen wesentlichen Teil der Datenbank zu vervielfältigen, zu verbreiten und öffentlich wiederzugeben (§ 87b Abs. 1 S. 1 UrhG), dürfen zur Datengewinnung ohne Zustimmung des Datenbankherstellers nur unwesentliche Teile der fremden Datenbank übernommen werden. Wie lange von unwesentlichen Teilen einer Datenbank ausgegangen werden kann, unterliegt der Beurteilung im Einzelfall und ist daher ebenfalls mit einer gewissen Rechtsunsicherheit behaftet. Werden jeweils nur einzelne Teile einer Datenbank abgefragt und liegt dieser Anteil bei unter 10%, wird keine Wesentlichkeit gegeben sein.¹⁴⁸

Der *sui-generis*-Schutz des Datenbankherstellers beruht auf der EU-Datenbankrichtlinie. Ein Leistungsschutzrecht für Datenbankhersteller findet sich daher allein im harmonisierten Gemeinschaftsrecht. Datenbankhersteller aus dem Ausland können sich dann auf die §§ 87a ff. UrhG berufen, wenn sie die Voraussetzungen des § 127a Abs. 1, 2 UrhG erfüllen oder eine Reziprozitätsvereinbarung zwischen der EU und dem entsprechenden Drittstaat vorliegt, § 127 Abs. 3 UrhG.¹⁴⁹ Solche internationalen Abkommen, Staatsverträge oder Gegenseitigkeitsabkommen bestehen bislang nicht, wenngleich diesbezügliche Bestrebungen andauern.¹⁵⁰ Dies ist im Hinblick auf das Big-Data-Management und die

¹⁴⁸ *BGH*, Urt. v. 22.6.2011- I ZR 159/10 = MMR 2012, 544; *BGH*, Urt. v. 1.12.2010 - I ZR 196/08 = MMR 2011, 676.

¹⁴⁹ *Dreier*, in: *Dreier/Schulze*, Vor §§ 87a ff. Rn. 12; *Vogel*, in: *Schricker/Loewenheim*, Vor §§ 87a ff. Rn. 53; *Thum*, in: *Wandtke/Bullinger*, Vor §§ 87a ff. Rn. 49.

¹⁵⁰ *Dreier*, in: *Dreier/Schulze*, Vor §§ 87a ff. Rn. 12; *Schricker*, § 127a Rn. 6; *Thum*, in: *Wandtke/Bullinger*, Vor §§ 87a ff. Rn. 50.

Ubiquität des Internet von besonderer Relevanz. Weder die schweizerische noch die US-amerikanische Rechtsordnung kennen ein Leistungsschutzrecht für Datenbankhersteller. Schweizer oder US-Amerikaner können sich auch nicht auf die §§ 87a ff. UrhG berufen, da sie die Voraussetzungen des § 127a UrhG nicht erfüllen. Dieser uneinheitliche Rechtsrahmen senkt die Attraktivität des Big-Data-Managements für Investoren aus Drittstaaten und hemmt damit dessen Entwicklung.

Die vorstehenden Ausführungen machen deutlich, dass bei der Gewinnung von Daten fremde Rechte auch aus urheberrechtlicher Sicht zu beachten sind. Die Gefahr der Verletzung dieser Rechte sollte daher in diesen Fällen keinesfalls außer Acht gelassen und bei der Entwicklung einer Big-Data-Lösung berücksichtigt werden.

c) Eigener Schutz der Big-Data-Anwendung

Nicht nur die Quelle der genutzten Daten, sondern auch die Big-Data-Anwendung selbst kann nach dem UrhG geschützt sein. Hierbei kommt es erneut auf die konkrete Ausgestaltung dieser Anwendung an. Denkbar ist zum einen, dass die Anwendung als Datenbankwerk geschützt wird oder dem sui-generis-Schutz für Datenbanken unterfällt.¹⁵¹ Naheliegend ist zudem der urheberrechtliche Schutz der Big-Data-Lösung als Computerprogramm (§§ 69a ff. UrhG). Computerprogramme in diesem Sinne sind Programme in jeder Gestalt (§ 69a Abs. 1 UrhG). Sie werden geschützt, wenn sie das Ergebnis eigener geistiger Schöpfung ihres Urhebers sind, wobei es insbesondere nicht auf qualitative oder ästhetische Kriterien ankommt, § 69a Abs. 3 UrhG.

d) Computer-generated work

Im Zusammenhang mit dem urheberrechtlichen Datenbankschutz und Big Data ergibt sich unter dem Stichwort „computer-generated work“ eine interessante Fragestellung und Forschungsperspektive. Hierbei geht es zum einen um die Frage, ob maschinell erzeugte Werke oder Ergebnisse überhaupt Gegenstand des Urheberrechts sein können. Zum anderen erscheint fraglich, wem etwaige Rechte zustehen.

Reine Maschinenerzeugnisse, an denen der Mensch nicht mitgewirkt hat, stellen kein schutzwürdiges Werk dar.¹⁵² Ein Werk erfordert eine menschlich-gestalterische Tätigkeit.¹⁵³ Für das Vorliegen einer persönlichen geistigen Schöpfung ist es deshalb entscheidend, inwiefern die Person den jeweiligen Vorgang beherrscht und somit selbst als Urheber gelten kann.¹⁵⁴ Menschliches Schaffen liegt noch vor, wenn eine Maschine als Hilfsmittel eingesetzt wird und Anweisungen des Menschen umsetzt.¹⁵⁵ Als Urheber von „computer-generated work“ gilt derjenige, der das Werk mit Hilfe der Software geschaffen hat und nicht der

¹⁵¹ Vgl. Beispiele bei *Kieun*, Database Technology for Large Scale Data, <http://www.cubrid.org/blog/web-2-0/database-technology-for-large-scale-data/> (zuletzt abgerufen am 6.9.2013).

¹⁵² *Loewenheim*, in: Schricker/Loewenheim, § 2 Rn. 12; *Schulze*, in: Dreier/Schulze, § 2 Rn. 8; *Dreyer*, in: Dreyer/Kotthoff/Meckel, § 2 Rn. 5

¹⁵³ *Loewenheim*, in: Schricker/Loewenheim, § 2 Rn. 11 m.w.N.

¹⁵⁴ *Schulze*, in: Dreier/Schulze, § 2 Rn. 8; *Dreyer*, in: Dreyer/Kotthoff/Meckel, § 2 Rn. 26.

¹⁵⁵ *Loewenheim*, in: Schricker/Loewenheim, § 2 Rn. 12; *Nordemann*, in: Fromm/Nordemann, § 2 Rn. 21; *Bullinger*, in: Wandtke/Bullinger, § 2 Rn. 16.

Programmierer der Software.¹⁵⁶ Eine Unterscheidung ist in der Praxis nur schwerlich zu bewerkstelligen, da sich die Ergebnisse menschlichen Schaffens und maschineller Tätigkeit zuweilen kaum unterscheiden.¹⁵⁷

Die rechtliche Einordnung eines Big-Data-Tools kann im Einzelfall mit erheblicher Rechtsunsicherheit verbunden sein und deshalb die Entwicklung in diesem Bereich hemmen. Gerade im Hinblick darauf, dass ein Großteil des Datenumgangs automatisiert erfolgt, ist die Klärung der Frage, ob in diesen Fällen überhaupt urheberrechtlicher Schutz denkbar ist, von grundlegender Bedeutung. Es ist deshalb Aufgabe der Forschung, verbindliche Kriterien herauszuarbeiten, die festlegen, ob und inwieweit computer-generierte Datenbanken dem Urheberrechtsschutz unterfallen.

e) *Screen-Scraping*

Die Akquirierung großer Datenmengen erfolgt vielfach im Wege des sog. „screen-scraping“. Hierbei handelt es sich um Verfahren, die die Gewinnung von Daten durch gezieltes Auslesen von Texten auf Webseiten ermöglichen.¹⁵⁸ Typische Beispiele sind Preisvergleichsportale oder News-Aggregatoren. Die rechtliche Zulässigkeit dieser Verfahren beschäftigt seit längerer Zeit die Gerichte und wird auch heute noch kontrovers diskutiert. Problematisch ist „screen-scraping“ vor allem im Hinblick auf die Rechte des Datenbankherstellers, das wettbewerbsrechtliche Behinderungsverbot und die allgemeine zivilrechtliche Deliktshaftung.

Der *BGH* urteilte im Jahr 2011, dass eine Suchmaschine, die verschiedene Automobil-Onlinebörsen anhand einer konkreten Suchanfrage durchsucht, grundsätzlich rechtlich zulässig sei.¹⁵⁹ Solche Onlinebörsen für Kraftfahrzeuge sind Datenbanken im Sinne des § 87a Abs. 1 S. 1 UrhG. Der Nutzer der Suchmaschine vervielfältigt Daten aus den entsprechenden Datenbanken, indem er das Ergebnis einer Suchanfrage im Arbeitsspeicher seines Computers abspeichert. Dadurch werde die Datenbank jedoch weder insgesamt noch zu einem wesentlichen Teil vervielfältigt. Eine Verletzung der Rechte des Datenbankherstellers (§ 87b Abs. 1 UrhG) liege deshalb nicht vor. Ebenso verneinte der *BGH* wettbewerbswidriges Verhalten aufgrund unlauterer Behinderung gem. § 4 Nr. 10 UWG. Die Tatsache, dass dem Betreiber der Onlinebörsen möglicherweise Werbeeinnahmen entgingen und er keine Begleitdienstleistungen anbieten könne, sei hinzunehmen, weil er sein Angebot im konkreten Fall bewusst ohne technische Schutzmaßnahmen öffentlich zugänglich gemacht habe. Ebenso wenig konnte der Kläger eine wettbewerbsrechtliche Behinderung aufgrund einer durch die Suchsoftware hervorgerufenen Störung nachweisen. Schließlich verneinte das Gericht einen deliktischen Anspruch wegen Verletzung des Rechts am eingerichteten und ausgeübten Gewerbebetrieb gem. § 823 Abs. 1 BGB. Es liege allein ein Eingriff in die Datenbank und kein Eingriff in das konkrete Unternehmen vor. Hierfür seien die §§ 87a ff.

¹⁵⁶ Schulze, in: Dreier/Schulze, § 2 Rn. 14.

¹⁵⁷ Loewenheim, in: Schricker/Loewenheim, § 2 Rn. 12.

¹⁵⁸ Deutsch, GRUR 2009, 1027 (1027).

¹⁵⁹ BGH, Urt. v. 22. 6. 2011 - I ZR 159/10 – Automobil-Onlinebörse = GRUR 2011, 1018.

UrhG die entsprechenden Spezialregelungen. Der *BGH* entschied sich damit für die prinzipielle Zulässigkeit des „screen-scraping“. Dieses Ergebnis stütze das Gericht im konkreten Fall allerdings maßgeblich darauf, dass der Betreiber der Onlinebörsen keine technischen Schutzvorkehrungen (Registrierung, Login) gegen das unbefugte Auslesen der Datenbank getroffen hatte. Daneben waren die Nutzungsbedingungen der Onlinebörsen, die ein solches Vorgehen untersagten, mangels ausdrücklicher Einwilligung unwirksam. Das Gericht schloss sich damit der Meinung des erstinstanzlich zuständigen *LG Hamburg* an.¹⁶⁰ Das *OLG Hamburg*¹⁶¹ hatte in zweiter Instanz wettbewerbsrechtliche Ansprüche des Onlinebörsen-Betreibers bejaht.

Das *OLG Frankfurt a.M.* hatte bereits im Jahr 2009 einen vergleichbaren Fall zu entscheiden, in dem es um das Auslesen von Flugangeboten auf den Websites der verschiedenen Fluggesellschaften ging.¹⁶² Auch dies hielt das Gericht damals für rechtlich zulässig. Wie umstritten und einzelfallabhängig die rechtliche Beurteilung des „screen-scraping“ ist, verdeutlicht eine Entscheidung des *OLG Hamburg*¹⁶³ aus dem vergangenen Jahr. Das Gericht blieb seiner bereits in der „Automobil-Onlinebörsen“-Entscheidung eingeschlagen Linie treu. Es verneinte rechtliche Ansprüche des Datenbankherstellers, gewährte den Fluggesellschaften, deren Websites im Wege des „screen-scraping“ ausgelesen werden, allerdings einen wettbewerbsrechtlichen Unterlassungsanspruch aus §§ 3, 4 Nr. 8, 10 UWG. Hierbei verwies das *OLG Hamburg* in seiner Entscheidung auf mehrere Faktoren, die die wettbewerbswidrige Behinderung begründen. Vor allem habe die Fluggesellschaft in ihren Nutzungsbedingungen den Weiterverkauf und die Weitervermittlung ihrer Flugdaten ausdrücklich untersagt. Diese AGB müsse der Endkunde der Fluggesellschaft auch ausdrücklich annehmen. Die Entscheidung des *OLG Hamburg* ist gegenwärtig beim *BGH* anhängig. Inwiefern diese der rechtlichen Überprüfung standhält, darf mit Spannung erwartet werden. Ein umfassendes „virtuelles Hausrecht“, das Websites gegen den Zugriff von außen schützt, widerspräche wohl der Funktionsweise des Internets.¹⁶⁴ Vielmehr scheint es zumutbar, den Anbieter der Website auf seine vertragsrechtlichen Regelungsmöglichkeiten zu verweisen. Will er eine bestimmte Nutzung „seiner“ Daten ausschließen, so muss er den Zugriff technisch einschränken und die Nutzung von einer wirksamen vertraglichen Zustimmung zu einer entsprechenden Beschränkung abhängig machen.¹⁶⁵ Festzuhalten ist bislang, dass die rechtliche Zulässigkeit des „screen-scraping“ stets im Einzelfall zu beurteilen ist, was sicherlich mit einer erhöhten Rechtsunsicherheit einhergeht.

¹⁶⁰ *LG Hamburg*, Urt. v. 9. 4. 2009 - 310 O 39/08 = BeckRS 2009, 20109.

¹⁶¹ *OLG Hamburg*, Urt. v. 18. 8. 2010 - 5 U 62/09 = GRUR 2011, 728.

¹⁶² *OLG Frankfurt a.M.*, Urt. v. 5.3.2009 - 6 U 221/08 = MMR 2009, 400.

¹⁶³ *OLG Hamburg*, Urt. v. 24.10.2012 - 5 U 38/10 = BeckRS 2012, 22946 = GRUR-Prax 2013, 344430.

¹⁶⁴ Ohly, in: Piper/Ohly/Sosnitza, § 4 Rn. 10.65; Deutsch, GRUR 2009, 1027 (1032).

¹⁶⁵ Zieger/Smirra, MMR 2013, 418 (421).

3. Betroffene Verwertungsrechte

Werden fremde Leistungen durch Urheber- oder Leistungsschutzrechte geschützt, stellt der Zugriff auf diese regelmäßig einen Eingriff in fremde Rechte dar. Welche Verwertungsrechte allerdings speziell betroffen sind, hängt maßgeblich von der konkreten Ausgestaltung der Big-Data-Anwendung ab.

Bei der Übernahme urheberrechtlich geschützter Texte, zur Archivierung oder im Rahmen einer sonstigen Vorgehensweise zum Big-Data-Management, wird regelmäßig das Vervielfältigungsrecht des Rechtsinhabers betroffen sein. Jede Speicherung auf einem Datenträger stellt eine urheberrechtlich relevante Vervielfältigung dar.¹⁶⁶ Dies gilt auch dann, wenn die Daten aufgrund einer Kodierung nur noch mittelbar wahrnehmbar, allerdings wieder dekodierbar sind.¹⁶⁷ Konkret ist daher zum Beispiel im Umgang mit Big Data die Arbeit eines Tools, das Datenblöcke mehrfach innerhalb eines Clusters ablegt, als grundsätzlich unerlaubte Vervielfältigung zu werten.¹⁶⁸

Daneben ist vor allem das Recht der öffentlichen Zugänglichmachung (§ 19a UrhG) von besonderer Bedeutung für das Big-Data-Management. Dies ist das Recht, das Werk drahtgebunden oder drahtlos der Öffentlichkeit in einer Weise zugänglich zu machen, die es Mitgliedern der Öffentlichkeit erlaubt, von beliebigen Orten und zu Zeiten ihrer Wahl Zugang zu dem Werk erlangen zu können. Beispiele für Eingriffe in § 19a UrhG sind das „Ins-Netzstellen“ von Werken, das Einstellen von Werken auf Servern zum Abruf, die Nutzung einer Datenbank oder CD-ROM im Intranet oder On-Demand-Dienste.¹⁶⁹ Auch hier ergeben sich allerdings regelmäßig größere Abgrenzungsschwierigkeiten, da der Begriff der Öffentlichkeit nicht klar definiert ist und somit im Einzelfall Raum für Unklarheiten lässt. Eine Übermittlung von Daten über das Internet an Dritte, wie es etwa von Monitoring-Anbietern praktiziert wird und auch bei einer Big-Data-Anwendung denkbar ist, verletzt damit regelmäßig das Recht der öffentlichen Zugänglichmachung.¹⁷⁰

a) Die „Used-Soft“-Entscheidung des EuGH

Ein lange erwartetes Grundsatzurteil zum Verbreitungsrecht und dessen Erschöpfung beim Onlinevertrieb „gebrauchter“ Software fällte der *EuGH* in seiner „UsedSoft“-Entscheidung,¹⁷¹ die bereits im Kapitel „Eigentum an Daten“ angesprochen wurde. Das Gericht urteilte zur umstrittenen Frage der Anwendbarkeit des urheberrechtlichen Erschöpfungsgrundsatzes auf den Onlinebezug von Computerprogrammen, dass sich das Recht auf ausschließliche Verbreitung der Programmkopien mit dem Erstverkauf erschöpfe. Stelle ein Konzern seinem Kunden eine Kopie zur Verfügung und gewähre diesem über einen Lizenzvertrag das unbefristete Nutzungsrecht dieser Kopie, so erlösche sein Recht auf ausschließliche Verbreitung. Durch ein solches Geschäft werde das Eigentum an dieser Kopie übertragen.

¹⁶⁶ Schulze, in: Dreier/Schulze, § 16 Rn. 7; Dustmann, in: Fromm/Nordemann, § 16 Rn. 12.

¹⁶⁷ Heerma, in: Wandtke/Bullinger, § 16 Rn. 2; Solmecke/Wahlers, ZD 2012, 550 (554).

¹⁶⁸ Hackenberg, Datenschutz setzt Big Data Grenzen, Computerwoche vom 24.1.2013.

¹⁶⁹ Bullinger, in: Wandtke/Bullinger, § 19a Rn. 22 ff.; Dreier, in: Dreier/Schulze, § 19a Rn. 6.

¹⁷⁰ Solmecke/Wahlers, ZD 2012, 550 (554).

¹⁷¹ EuGH, U. v. 3. 7. 2012, Az. C-128/11 = MMR 2012, 586 m. Anm. Heydn.

Das gelte auch für verbesserte und aktualisierte Fassungen, weil sie Bestandteil der Kopie seien. Auf eine Unterscheidung von verkörperter und unverkörperter Kopie kommt es nach der Entscheidung des *EuGH* nicht mehr an. Anderer Auffassung waren zuvor die deutsche Rechtsprechung und Teile der rechtswissenschaftlichen Literatur.¹⁷²

Im Hinblick auf die dogmatische Einordnung des Onlinebezugs als „Erstverkauf“ i.S.d. Softwareschutzrichtlinie bekräftigt der *EuGH* die bisherige Rechtsprechung des *BGH* zur Software als Sachkauf und der vertragstypologischen Einordnung des ASP-Vertrages (Application Service Providing) als Mietvertrag.¹⁷³ Die Erschöpfung bezieht sich allerdings nicht auf Dienstleistungsverträge oder die Einräumung eines zeitlich befristeten Nutzungsrechts.

Die Grundsatzentscheidung des *EuGH* eröffnet wichtige Fragen für das Big-Data-Management. Die Forschung wird klären müssen, inwieweit sich das Urteil auf andere digitale Güter als Software übertragen lässt.¹⁷⁴ Eine Abschottung nachgelagerter Märkte durch eine weite Anwendung des Erschöpfungsgrundsatzes könnte verhindert werden, wenn dieser gerade auch für andere digitale Inhalte wie eBooks oder Musikfiles anerkannt wird.¹⁷⁵ Unter wirtschaftlichen Gesichtspunkten wird es keinen großen Unterschied machen, ob es sich im konkreten Fall um Software oder eBooks handelt. Allerdings hat der *EuGH* in seiner Entscheidung sehr stark mit Bezugnahme auf die Besonderheiten der EU-SoftwareschutzRL (2009/24/EG) argumentiert.¹⁷⁶ Dies nahm jüngst das *LG Bielefeld*¹⁷⁷ in einer ersten Entscheidung zu dieser Frage zum Anlass, die Argumentation des *EuGH* zum Handel mit Gebrauchtsoftware nicht auf andere Werkarten zu übertragen. Der Weiterverkauf von eBooks sei nicht vom Erschöpfungsgrundsatz gedeckt.

b) Cloud Computing

Auch im Hinblick auf betroffene Verwertungsrechte, etwa im Rahmen des Cloud Computing, ergeben sich Rechtsfragen, die in Zusammenhang mit der „UsedSoft“-Entscheidung zu sehen sind. Grundsätzlich gilt es beim Cloud Computing die Rechtsverhältnisse des Cloud-Anbieters zum Hersteller der Applikationssoftware einerseits und das Verhältnis des Cloud-Anwenders zum Cloud-Anbieter andererseits zu unterscheiden.¹⁷⁸ Im ersten Fall stellt die Nutzung der Software durch den Cloud-Anbieter eine Vervielfältigung (§ 69c Nr. 1 UrhG) dar. Die Frage, ob die Softwarenutzung durch den Cloud-Anbieter auch das Verbreitungs- und Vermietungsrecht (§ 69c Nr. 3 UrhG) des Softwareherstellers berührt oder ob eine körperliche Überlassung der Software hierzu nötig ist, war vor dem Urteil des *EuGH* umstritten.

¹⁷² Vgl. *BGH*, GRUR 2011, 418 (421); *OLG Düsseldorf*, ZUM 2010, 60; *OLG Frankfurt a.M.*, MMR 2009, 544 sowie *Hoeren/Försterling*, MMR 2012, 642 m.w.N.

¹⁷³ *BGH*, MMR 2007, 243.

¹⁷⁴ *Hartmann*, GRUR-Int 2012, 980 (981 ff.); *Hoeren/Försterling*, MMR 2012, 642 (647).

¹⁷⁵ *Hoeren/Försterling*, MMR 2012, 642 (647).

¹⁷⁶ Siehe dazu auch bereits Kapitel „A. 2. b.“.

¹⁷⁷ *LG Bielefeld*, GRUR-RR 2013, 281.

¹⁷⁸ Hierzu und zu den folgenden Aussagen: *Nägele/Jacobs*, ZUM 2010, 281 (285 ff.).

Auch die vertragstypologische Einordnung von Cloud-Verträgen ist nicht endgültig geklärt.¹⁷⁹ Das Mietrecht halten verschiedene Autoren für nicht anwendbar. Zwar verlange das Mietrecht nicht zwingend eine Überlassung der Mietsache, aber doch zumindest die Möglichkeit eines (wenn auch zeitlich begrenzten) tatsächlichen Zutritts zur Sache. Davon könne man im Falle des Cloud Computing, das eine bloße Nutzungsmöglichkeit bietet, nicht sprechen. Hierbei will der Provider auch regelmäßig nicht den „Erfolgsgedanken“ des Werkvertragsrechts (§§ 631 ff. BGB) übernehmen. Mithin ist an eine Anwendung der Vorschriften über den Dienstvertrag (§§ 611 ff. BGB) zu denken.

Der *BGH* hat in seinem Urteil zu ASP-Verträgen die Einordnung als Mietvertrag, der die entgeltliche Gebrauchsüberlassung einer beweglichen Sache zum Gegenstand hat, angenommen.¹⁸⁰ Typischerweise sei die Gewährung der Onlinenutzung von Software für eine begrenzte Zeit im Mittelpunkt der vertraglichen Pflichten. Der Anwendung des Mietrechts stehe nicht entgegen, dass es sich bei der Software nicht um eine Sache i.S.d. § 90 BGB handele. Der *BGH* habe wiederholt entschieden, dass eine auf einem Datenträger verkörperte Standardsoftware als bewegliche Sache anzusehen ist, auf die je nach der vereinbarten Überlassungsform Miet- oder Kaufrecht anwendbar ist. Diese Rechtsprechung wurde nun vom *EuGH* bekräftigt.

Eine verbindliche Erforschung und Klärung der aufgeworfenen Rechtsfragen ist – vor allem auch im Hinblick auf das Cloud Computing und die Pflichten des Cloud-Providers – dringend geboten. Die Erschöpfung im Onlinebereich hat mithin auch für das Big-Data-Management weitreichende Auswirkungen. Die Analyse der noch nicht vollständig absehbaren rechtlichen und wirtschaftlichen Folgen des *EuGH*-Urteils in Sachen „Used-Soft“, das bisherigen Ansichten in Rechtsprechung und Schrifttum teilweise widerspricht und kontrovers diskutiert wird, erscheint deshalb dringend geboten.

4. Schranken des Urheberrechts

Eingriffe in urheberrechtliche Verwertungs- oder Leistungsschutzrechte sind zulässig, wenn sie durch entsprechende Schrankenregelungen ausdrücklich legitimiert sind. Regelmäßig werden der Umgang mit großen Datenmengen und der damit verbundene Eingriff in fremde Verwertungsrechte allerdings nicht legitimiert sein.¹⁸¹ Insbesondere erscheinen weder das Zitatrecht (§ 51 UrhG) noch das Recht, Vervielfältigungen zum privaten und sonstigen eigenen Gebrauch anzufertigen (§ 53 UrhG), als einschlägige Schranken.

Es drängt sich damit eine weitere Forschungsperspektive auf: Muss das UrhG um eine weitere Schranke ergänzt werden und wie ist diese ggf. auszugestalten, um den neuen Dimensionen des Datenverkehrs gerecht zu werden? Eine solche Schranke müsste insbesondere die urheberrechtlichen Interessen in ausreichendem Maße berücksichtigen und sollte dennoch die gewinnbringende Auswertung von Daten ermöglichen. Zu beachten

¹⁷⁹ *Eichler*, Anwalt 5/2002, 20; *Sedlmaier/Kolk*, MMR 2002, 75 (78 f.); *v. Westerholt/Berger*, CR 2002, 81 (84); aus urheberrechtlicher Sicht *Czychowski/Bröcker*, MMR 2002, 81.

¹⁸⁰ *BGH*, MMR 2007, 243 = NJW 2007, 2394.

¹⁸¹ *Zieger/Smirra*, MMR 2013, 418 (419).

ist, dass der Gesetzgeber bei der Schaffung gesetzlicher Schranken nicht völlig frei agieren kann, sondern durch das nationale und internationale Recht beeinflusst wird. Zudem ist bei Bejahung der obigen Frage zu erwägen, ob eine maßnahmenorientierte Schranke oder eine generalklauselartige Formulierung, wie sie etwa aus dem Bereich des Common Law bekannt ist, vorzugswürdig wäre. Eine Erforschung dieser Rechtsfrage könnte nicht nur zu einer lang geforderten Modernisierung urheberrechtlicher Schranken führen. Sie würde das Urheberrecht vollständig an das digitale Zeitalter, insbesondere die Big-Data-Technologien, anpassen. In diesem Fall wäre eine gewinnbringende wirtschaftliche Nutzung der Big-Data-Technologien gewährleistet.

Eine aus dem Januar stammende Entscheidung des *EGMR*¹⁸² in Straßburg eröffnet eine weitere Forschungsperspektive. Abweichend von der bisherigen Rechtsprechung des *BGH* entschied der *EGMR*, dass die EMRK das geistige Eigentum nicht uneingeschränkt schütze. Das Recht des Urhebers müsse in jedem Einzelfall mit der Meinungs- und Informationsfreiheit der Medien und ihrer Nutzer aus Art. 10 EMRK abgewogen werden. Dies bedeutet, dass die urheberrechtlichen Schrankenregelungen nicht stets abschließend sind, sondern dass auch ohne das Eingreifen einer urheberrechtlichen Schranke dem Informationsinteresse der Allgemeinheit und der Pressefreiheit Vorrang einzuräumen sein kann. Zwar hat die EMRK nur den Rang einfachen Bundesrechts, dennoch haben die nationalen Gerichte die Entscheidungen des *EGMR* im Rahmen ihrer Rechtsprechung zu beachten. Das Bundesverfassungsgericht zieht die EMRK als Auslegungshilfe für einfaches Recht in Betracht und nutzt sie, um Inhalt und Reichweite der Grundrechte zu bestimmen.¹⁸³ Der Rechtsprechung des *EGMR* kommt auf EU-Ebene deshalb große Bedeutung zu, wie sich auch aus Art. 6 Abs. 2 EUV und Art. 52 Abs. 3 der EU-Grundrechte-Charta ergibt.

Die bereits oben aufgezeigte Problematik der urheberrechtlichen Schrankenregelungen wird durch das Urteil des *EGMR* verstärkt. Fraglich ist nicht allein, ob die bestehenden Schranken noch zeitgemäß sind oder ob es einer neuen Schranke bedarf. Die Rechtsprechung des *EGMR* liefert vielmehr wesentliche Neuerungen im Hinblick auf die Auslegung urheberrechtlicher Schranken, die Abwägung von Urheberrecht und Informationsinteresse und bringt daher die bisherige nationale Rechtsprechung ins Wanken.

5. Einräumung von Nutzungsrechten

Greifen keine legitimierenden Schranken, hängt die Zulässigkeit der urheberrechtlich relevanten Handlung von der Zustimmung des Rechteinhabers ab. Dieser kann seine Zustimmung entweder durch Einwilligung oder vertraglich erteilen.

Im Sinne der Rechtssicherheit erscheint der rechtsgeschäftliche Erwerb der Nutzungsrechte aufgrund ausdrücklicher vertraglicher Regelung dienlich. Das UrhG regelt in den §§ 31 ff. die Einräumung von Nutzungsrechten einfacher oder ausschließlicher Art. Im Rahmen eines entsprechenden Vertragsschlusses können verschiedene praktische Probleme auftreten. Die

¹⁸² *EGMR*, Urt. v. 10.1.2013, Nr. 36769/08 - Ashby Donald/Frankreich.

¹⁸³ BVerfGE 111, 307 (329); 128, 326 (367 ff.).

neuen Technologien des Internetzeitalters führen dazu, dass fast zwangsläufig in zahllose fremde Ausschließlichkeitsrechte eingegriffen wird.¹⁸⁴ Je nach Einzelfall wird auch das Big-Data-Tool seine Daten aus verschiedensten Quellen generieren. Es müsste unter Umständen eine erhebliche Zahl vertraglicher Vereinbarungen getroffen werden, was in der Praxis enorme Probleme aufwerfen würde. Regelmäßig wird es mithin im Rahmen des Big-Data-Managements Fälle geben, in denen weder legitimierende Schranken greifen noch eine vertragliche Einräumung von Nutzungsrechten stattgefunden hat.

In vergleichbaren Konstellationen bediente sich der *BGH* der Konstruktion einer schlichten konkludenten Einwilligung, um zu interessengerechten und praxisorientierten Lösungen zu gelangen.¹⁸⁵ Das Gericht entschied in seinen beiden „Vorschaubilder“-Entscheidungen¹⁸⁶, dass in dem Einstellen von Inhalten ins Internet dann eine (schlichte) Einwilligung in die Verwendung der Inhalte durch Suchmaschinen zu sehen ist, wenn der Betreffende von bestehenden Schutzmöglichkeiten keinen Gebrauch macht. Dies gelte selbst dann, wenn die Abbildung nicht vom Urheber selbst, sondern von einem Dritten ins Internet eingestellt wurde. Hier sollen jedoch nicht nur jene Abbildungen erfasst werden, die nach Zustimmung des Urhebers eingestellt wurden, sondern auch unberechtigterweise ins Internet eingestellte Kopien.

Hieraus ergibt sich unmittelbar die Frage, inwiefern sich die Grundsätze des *BGH* auch auf das Big-Data-Management übertragen lassen. Kann die Auswertung einer fremden Datenbank durch eine konkludente Einwilligung legitimiert sein? Im Rahmen der Beantwortung dieser Frage kommt es auch darauf an, wie tief die Analyse und Auswertung der Daten in fremde Rechte eingreift.

Von grundsätzlicherer Bedeutung ist allerdings die Frage, ob die Konstruktion der schlichten Einwilligung überhaupt haltbar ist. Große Teile des Schrifttums üben Kritik an der Rechtsprechung des *BGH* und äußern dogmatische Bedenken gegenüber der schlichten Einwilligung.¹⁸⁷ Zwar führe die Annahme einer schlichten Einwilligung zu interessengerechten und wünschenswerten Ergebnissen.¹⁸⁸ Sie sei allerdings allenfalls eine „Notlösung“¹⁸⁹ oder „Krücke“¹⁹⁰. Die (schlichte) Einwilligung könne zu einer Umgehung des bestehenden Schrankensystems führen. Es komme zu Unstimmigkeiten und Verwerfungen im Urheberrechtssystem, da durch die Einwilligung Eingriffe legitimiert werden, die als zulässig gewünscht sind, aber nicht von den bestehenden Schranken erfasst werden.¹⁹¹ Insbesondere breche der *BGH* mit der zivilrechtlichen Dogmatik der Willenserklärung, was

¹⁸⁴ *Klass*, ZUM 2013, 1.

¹⁸⁵ *Zieger/Smirra*, MMR 2013, 418 (419).

¹⁸⁶ BGHZ 185, 291 = NJW 2010, 2731 – Vorschaubilder I; *BGH*, Urt. v. 19. 10. 2011 – I ZR 140/10 = NJW 2012, 1886 – Vorschaubilder II.

¹⁸⁷ Vgl. etwa *Conrad*, ZUM 2010, 585; *Götting*, LMK 2010, 309481; *Hüttner*, WRP 2010, 1008 ff.; *Spindler*, GRUR 2010, 785.

¹⁸⁸ Vgl. etwa *Klass*, ZUM 2013, 1 (8); *Thum*, GRUR-Prax 2012, 215.

¹⁸⁹ *Götting*, LMK 2010, 309481.

¹⁹⁰ *Spindler*, GRUR 2010, 785 (791).

¹⁹¹ *Klass*, ZUM 2013, 1 (8).

im Hinblick auf das Selbstbestimmungsrecht des Einzelnen bedenklich sei.¹⁹² Auf diese Weise führt die Einwilligung unter Umgehung des Schrankensystems zu einer nicht normierten „Quasi-Fair-Use-Regelung“ im deutschen Recht.¹⁹³

Diesen (unbefriedigenden) Zustand gilt es zu beheben. Ein erhöhtes Maß an Rechtssicherheit kann als Treiber für Big-Data-Technologien wirken. Hierzu kann eine Lösung über das urheberrechtliche Schrankensystem erfolgen, in dem eine neue Schranke etabliert oder eine bestehende Schranke erweitert wird.¹⁹⁴ Zielführend könnte die Einführung einer „Fair-Use-Klausel“ nach US-Vorbild sein.¹⁹⁵ Diese wäre in der Lage, neuartige Entwicklungen flexibel zu erfassen, und unterläge im Konflikt zwischen verschiedenen Lobbygruppen nicht der potenziellen Blockade des Gesetzgebers.¹⁹⁶

Losgelöst von der Problematik der Einräumung einzelner Nutzungsrechte im Rahmen des Datenumgangs muss ein Big-Data-Nutzer grundsätzlich auch Nutzungsrechte für das konkrete Big-Data-Tool erwerben. Dies eröffnet regelmäßig keine besonderen Schwierigkeiten, da es sich bei der Mehrzahl der Big-Data-Tools, wie etwa dem Framework „Apache Hadoop“, um frei herunterladbare „open source“-Software handelt. „Apache Hadoop“ steht unter einer „Apache License 2.0“, welche die freie Verwendung, Modifizierung und Verteilung der Software erlaubt. Im Gegensatz zur „General Public License (GNU/GPL)“ handelt es sich bei der „Apache-Softwarelicense“ um keine „Copyleft“-Lizenz, so dass kein Zwang besteht, vollendete Versionen der Software auf der Basis von Open Source-Produkten wiederum unter einer Open Source-Lizenz weiterzugeben.

D. Datenschutzrecht

Auch das Verhältnis von Datenschutz und Big-Data-Management ist nicht unproblematisch. Während Vertreter der IT-Branche die Vorteile des Umgangs mit Big Data preisen,¹⁹⁷ warnen Datenschützer im Gegenzug vor den Gefahren von Big Data.¹⁹⁸ Unter anderem wird kritisiert, dass schon die Quantität der gesammelten Informationen wegen der Möglichkeit einer umfassenden Profilbildung gefährlich sei. Zudem wird die Sicherheit der Datenbestände angezweifelt, deren Schutz nicht mit dem Wachstum des Datenvolumens Schritt halte. Inwiefern diese Bedenken berechtigt sind und ob das Datenschutzrecht diesen Gefahren effektiv begegnen kann, soll im Folgenden näher untersucht werden.

¹⁹² Klass, ZUM 2013, 1 (5f.); Spindler, GRUR 2010, 785 (790).

¹⁹³ Spindler, GRUR 2010, 785 (790); Thum, GRUR-Prax 2012, 215.

¹⁹⁴ Klass, ZUM 2013, 1 (9f.) mit Verweis auf den Vorschlag von Schaefer, Urheberrechtliche Rahmenbedingungen für Bildersuchmaschinen de lege lata und de lege ferenda, S. 265 ff.

¹⁹⁵ Kleinemenke, CR 2009, 55 (56); Ott, ZUM 2009, 345 (352ff.).

¹⁹⁶ Spindler, GRUR 2010, 785 (792).

¹⁹⁷ Vgl. z.B. Handelsblatt vom 11.03.2012, abrufbar unter <http://www.handelsblatt.com/technologie/it-tk/cebit-special-2012/hintergrund/big-data-wie-aus-daten-ein-wettbewerbsvorteil-wird/6287194.html> (zuletzt abgerufen am 6.9.2013).

¹⁹⁸ Vertiefend dazu: ARTICLE 29 DATA PROTECTION WORKING PARTY, Opinion 03/2013 on purpose limitation, S. 45.

1. Datenschutzrechtliche Grundsätze

Das Datenschutzrecht wird von verschiedenen Grundsätzen geprägt, deren Zweck der Schutz des im Volkszählungsurteil¹⁹⁹ entwickelten Grundrechts auf informationelle Selbstbestimmung (Art. 2 Abs. 1, Art. 1 Abs. 1 GG) ist. Es handelt sich um „vorverlagerten Grundrechtsschutz“²⁰⁰, der integraler Bestandteil eines „Codex Digitalis Universalis“ sein soll.²⁰¹ Das Verständnis dieser Grundsätze ist daher auch Ausgangslage für die Lösung datenschutzrechtlicher Fragen in Zusammenhang mit Big Data. Neben den allgemeinen Rechtsgrundsätzen, wie dem *Gebot der Normenklarheit* oder dem *Verhältnismäßigkeitsgrundsatz*, gelten im Datenschutzrecht speziell die Grundsätze der *Zweckbindung*, der *Transparenz*, der *Direkterhebung* und der *Datenvermeidung* bzw. der *Datensparsamkeit*. Zudem steht die Datenverarbeitung nach dem BDSG unter einem *Verbot mit Erlaubnisvorbehalt*, worauf in der Folge noch gesondert eingegangen wird. Die speziellen datenschutzrechtlichen Grundsätze haben in unterschiedlicher Form ihren Eingang in das BDSG gefunden. In Bezug auf Datenvermeidung und Datensparsamkeit normiert § 3a BDSG ausdrücklich, dass jeder Datenumgang mit dem Ziel zu erfolgen hat, so wenig personenbezogene Daten wie möglich zu erheben und diese möglichst zu anonymisieren oder zu pseudonymisieren. Der Grundsatz der Direkterhebung findet sich in § 4 Abs. 2 S. 1 BDSG wieder, wonach personenbezogene Daten beim Betroffenen zu erheben sind. Der Zweckbindungsgrundsatz, der für Big Data von besonderer Wichtigkeit ist²⁰², besagt wiederum, dass die Daten grundsätzlich nur zu dem Zweck verarbeitet und genutzt werden dürfen, zu dem sie ursprünglich erhoben worden sind. Schließlich gewährleistet der Transparenzgrundsatz jedem Betroffenen das Recht, Einsicht in seine gespeicherten personenbezogenen Daten zu nehmen, was einerseits durch Auskunftsrechte des Betroffenen und andererseits durch Informationspflichten der verarbeitenden Stelle sichergestellt wird. Nur in begrenzten Ausnahmefällen, etwa aufgrund eines überwiegenden Allgemeininteresses, kann die Zweckbindung entfallen.²⁰³

2. Anwendbarkeit des BDSG - Kollisionsrechtliche Fragen bei grenzüberschreitendem Bezug

Grundsätzlich gilt im Datenschutzrecht das Territorialprinzip, das heißt ausländische Stellen müssen das BDSG beachten, wenn sie personenbezogene Daten in Deutschland verarbeiten oder nutzen.²⁰⁴ Dieses Prinzip gilt nur dann nicht, sofern eine in einem anderen Mitgliedstaat der EU oder EWR-Vertragsstaat belegene verantwortliche Stelle personenbezogene Daten im Inland erhebt, verarbeitet oder nutzt. In Abweichung vom Territorialprinzip gilt dann das

¹⁹⁹ *BVerfG*, U. v. 15.12.1983, Az. 1 BvR 209, 269, 362, 420, 440, 484/83.

²⁰⁰ *Schneider*, MMR 2009, VII, IX.

²⁰¹ *Weichert*, ZD 2013, 251 (251).

²⁰² *ARTICLE 29 DATA PROTECTION WORKING PARTY*, Opinion 03/2013 on purpose limitation, S. 46 f.

²⁰³ *Weichert*, ZD 2013, 251 (256).

²⁰⁴ *RegE* zum BDSG, BT-Drs. 14/4329, 31 f.; *Spies*, MMR 2009, XI.

Niederlassungsprinzip (auch: modifiziertes Sitzprinzip), das heißt, es kommt entscheidend auf die Niederlassung der verantwortlichen Stelle an.²⁰⁵

Beim Umgang mit personenbezogenen Daten durch Stellen, die in Drittstaaten (Staaten außerhalb der EU / des EWR) angesiedelt sind, ist § 1 Abs. 5 S. 2 BDSG als kollisionsrechtliche Vorschrift zu beachten. Das BDSG findet Anwendung, sofern die in einem Drittstaat belegene verantwortliche Stelle personenbezogene Daten im Inland erhebt, verarbeitet oder nutzt.²⁰⁶

Besonders im Zusammenhang mit Big-Data-Anwendungen auf ausländischen Servern ist noch völlig ungeklärt, wann ein ausländischer Anbieter deutsches Datenschutzrecht zu beachten hat.

Klärungsbedarf besteht zunächst im Hinblick auf das Merkmal „Erheben“. Gemäß der Legaldefinition in § 3 Abs. 3 BDSG ist „Erheben“ das Beschaffen von Daten über den Betroffenen. Es handelt sich um einen mehrstufigen Prozess, bei dem der entsprechende Betroffene Daten preisgibt und die verantwortliche Stelle diese aufnimmt.²⁰⁷ Fraglich ist, wie der Ort zu bestimmen ist, an dem die Erhebung der personenbezogenen Daten stattfindet. Es wird u.a. vorgeschlagen auf den Sitz der verantwortlichen Stelle, den Standort des Servers dieser Stelle oder auch den Standort des Betroffenen abzustellen.²⁰⁸

Eine inländische Erhebung von Daten soll vorliegen, wenn die verantwortliche Stelle hierzu auf im Inland belegene Computer der Nutzer zurückgreift und mit Erhebungswillen Kenntnis von deren Daten erhält.²⁰⁹ Eine Erhebung von Daten als verantwortliche Stelle soll demnach anzunehmen sein, wenn die verantwortliche Stelle im Inland spezielle Einwahlknoten oder andere Zugangseinrichtungen für Kunden betreibt.²¹⁰

Zu einer Anwendbarkeit soll es nach einer Ansicht jedoch nicht kommen, wenn ein Nutzer lediglich auf einer Website ein Anmeldeformular oder eine andere Datenmaske ausfüllt, weil der Betreiber der Website keinen Einfluss darauf hat, wer wann von seinem Angebot Gebrauch macht.²¹¹ Zu den Begriffen „verarbeiten“ und „nutzen“ finden sich in § 3 Abs. 4, 5 BDSG Legaldefinitionen.

Nach obergerichtlicher Rechtsprechung soll das Datenschutzrecht dagegen dann Anwendung finden, wenn keine im Inland belegenen Speicher oder Datenleitungen der verarbeitenden Stelle vorhanden sind. Eine Verarbeitung fände auch dann im Inland statt, wenn personenbezogene Daten zur Einsicht oder zum Abruf bereitgehalten werden.²¹² Können

²⁰⁵ Bergmann/Möhrle/Herb, § 1 Rn. 36; Ellger, Datenschutz, S. 604; Gola/Schomerus, § 1 Rn. 27.

²⁰⁶ Gola/Schomerus, § 1 Rn. 29.

²⁰⁷ Dammann, in: Simitis, § 1 Rn. 222; Jotzo, MMR 2009, 232 (235).

²⁰⁸ Jotzo, MMR 2009, 232, (235).

²⁰⁹ Jotzo, MMR 2009, 232, (237).

²¹⁰ Gabel in: Taeger/Gabel, § 1 Rn. 59.

²¹¹ Dammann, in: Simitis, § 1 Rn. 223.

²¹² OLG Hamburg, ZUM 2012, 405 (406).

Daten also bspw. über eine Homepage oder ein sonstiges Big-Data-Tool in Deutschland abgerufen werden, findet auch hier eine Verarbeitung statt.²¹³

Nach teilweise vertretener Ansicht bestimmt sich das anwendbare Recht direkt danach, an wen die im Drittstaat belegene verantwortliche Stelle ihre Leistungen, insbesondere ihre Internetpräsenz, richtet.²¹⁴ Indizien für den Willen, sich an deutsche Kunden zu richten, sollen etwa die Verwendung der deutschen Sprache auf einer Homepage, eine länderspezifische Domain oder die Tatsache sein, dass das Angebot auf sonstige Weise äußerlich erkennbar an deutsche Nutzer gerichtet ist.²¹⁵ Entscheidend ist jedoch nach dem Wortlaut von § 1 Abs. 5 BDSG vielmehr der Sitz der verarbeitenden Stelle. Insofern gilt für eine Datenverarbeitung durch eine US-Gesellschaft nicht das deutsche und europäische Datenschutzrecht, wenn diese Gesellschaft als Hauptniederlassung des Konzerns in den USA keine Daten im europäischen Inland erhebt, verarbeitet oder nutzt.

Im Interesse der rechtssicheren Ausübung von Big Data besteht hier großes Forschungspotential für die Entwicklung rechtlicher und institutioneller Strukturen, die einerseits eine eindeutige Zuordnung von Daten zu einem (Server-)Standort garantieren und andererseits Lösungen zum Umgang mit dem ungeklärten Begriff der Erhebung entwickeln. Die derzeitige Rechtslage jedenfalls führt zu großer Verunsicherung bei Bürgern und Unternehmen, da eine eindeutige Bestimmung des anwendbaren Rechts vielfach nicht möglich ist und so die Wirtschaftstätigkeit gehemmt wird.

Aufgrund des die Anwendbarkeit des BDSG regelnden Territorialprinzips stellt sich im Zusammenhang mit Big-Data-Anwendungen zudem grundsätzlich die Frage, inwieweit sich der Standort bzw. ein Standort der Datenmengen überhaupt verlässlich bestimmen lässt. Auf den räumlichen Anwendungsbereich des deutschen Datenschutzrechts haben vertragliche Rechtswahlklauseln grundsätzlich keinen Einfluss, so dass eine Lösung auf diesem Wege nicht möglich ist.²¹⁶ Durch Cloud Computing sind Daten weltweit auf verschiedene Standorte verteilt,²¹⁷ die großen Datenmengen sind zudem flüchtig und der einmal ermittelte Standort kann sich binnen kürzester Zeit erneut ändern. Die Anwendbarkeit einer bestimmten Rechtsordnung kann folglich nur schwer aufgrund des Standortes der Daten ermittelt werden,²¹⁸ so dass das Territorialitätsprinzip im Rahmen vieler Big-Data-Anwendungen versagt.

Perspektivisch ist daher herauszuarbeiten, ob und inwieweit überhaupt eine rechtsfeste Bestimmung des Standortes großer Datenmengen möglich ist und wie den durch

²¹³ OLG Hamburg, ZUM 2012, 405 (406).

²¹⁴ Dammann, in: Simitis, § 1 Rn. 220; Jotzo, MMR 2009, 232 (236).

²¹⁵ Dammann, in: Simitis, § 1 Rn. 220; Jotzo, MMR 2009, 232 (237).

²¹⁶ Gola/Schomerus, § 1 Rn. 16; Weichert, in: Däubler/Klebe/Wedde/Weichert, § 1 Rn. 5.

²¹⁷ Nägele/Jacobs, ZUM 2010, 281 (289 f.).

²¹⁸ Nägele/Jacobs, ZUM 2010, 281 (289 f.).

gleichzeitige Anwendbarkeit mehrerer Rechtsordnungen entstehenden Missbrauchsmöglichkeiten (z.B. sog. „Forum Shopping“) begegnet werden kann.²¹⁹

Auf europäischer Ebene wurde erkannt, dass die bisherigen Regelungen nicht mit der technischen Entwicklung mithalten können. Die reformbedürftigen derzeit geltenden Regelungen werden als fragmentiert erachtet und mitverantwortlich dafür gemacht, dass Online-Aktivitäten von Nutzern als risikobehaftet wahrgenommen werden.²²⁰ Die Unterschiede in der Umsetzung der Datenschutz-Richtlinie stellen dabei ein Hemmnis für den grenzüberschreitenden Datenverkehr dar und somit nach Ansicht der EU-Kommission auch eine Wettbewerbsbeschränkung.²²¹

Die geplante europaweit einheitliche Neuregelung des Datenschutzrechts durch die Datenschutz-Grundverordnung, auf deren Entwurf sich der zuständige Ausschuss des EU-Parlaments im Oktober 2013 einigte, soll für die Bestimmung der Anwendbarkeit des europäischen Rechts wesentliche Änderungen mit sich bringen: Für in der Union niedergelassene Verantwortliche gilt nunmehr europaweit einheitlich die Datenschutz-Grundverordnung. Für nicht in der Union niedergelassene Verantwortliche müssen europäisches Recht beachten, sofern eine Verarbeitung personenbezogener Daten von in der Union ansässigen Betroffenen dazu dient, a) diesen Personen in der Union Waren oder Dienstleistungen anzubieten, oder b) ihr Verhalten zu beobachten. Mit der Einführung der neuen Begriffskategorien soll eine starke Ausweitung des Anwendungsbereiches verbunden sein.²²²

Im Zuge weiterer Forschung ist dringend klärungsbedürftig, was unter einem „Angebot von Waren oder Dienstleistungen“ zu verstehen ist und was der Begriff „Beobachtung von Verhalten“ beinhaltet; zu denken ist hier insbesondere an Big-Data-Anwendungen/-Lösungen als mögliche Dienstleistungen. Nicht geklärt ist ferner, ob der Begriff der Beobachtung von Verhalten auch die Analyse desselben umfasst – wäre dem nicht so, könnten erhebliche Teile von Big-Data-Anwendungen/-Lösungen aus dem Anwendungsbereich ausgenommen sein. Nur durch weitere Forschung auf diesem Gebiet kann geklärt werden, ob es nunmehr gelungen ist, eine zweifelsfreie Regelung der Anwendbarkeit zu erreichen oder ob weiterhin Zweifelsfälle bestehen.

3. Telemediengesetz (TMG)

Das TMG gilt für alle elektronischen Informations- und Kommunikationsdienste, soweit sie nicht Telekommunikation oder Rundfunk im engeren Sinne sind und damit den Regelungen von Telekommunikationsgesetz (TKG) oder Rundfunkstaatsvertrag (RStV) unterfallen.²²³ Die

²¹⁹ Nägele/Jacobs, ZUM 2010, 281 (289 f.); Spies, MMR 2009, XI, XII.

²²⁰ Hoeren/Giurgiu, NWB 2012, 1599 (1599 f.).

²²¹ Hoeren/Giurgiu, NWB 2012, 1599 (1600).

²²² Hoeren/Giurgiu, NWB 2012, 1599 (1601).

²²³ Müller-Broich, § 1 Rn. 6.

internationale Anwendbarkeit des TMG bestimmt sich ebenfalls nach den Regelungen in § 1 Abs. 5 BDSG.²²⁴

Zu den Telemedien zählen bspw. auch „Online-Dienste, die Instrumente zur Datensuche, zum Zugang zu Daten oder zur Datenabfrage bereitstellen“²²⁵, womit auch zahlreiche Big-Data-Anwendungen erfasst sind.

Die datenschutzrechtlichen Grundsätze des TMG gem. § 12 decken sich mit den allgemeinen Grundsätzen des BDSG. Abweichend von § 4a BDSG kann die Einwilligung im Bereich der Telemedien auch elektronisch erteilt werden, § 13 Abs. 2 TMG. Dabei ist nicht nur sicherzustellen, dass der Nutzer seine Einwilligung bewusst und eindeutig erteilt hat, sondern darüber hinaus muss die Einwilligung protokolliert werden und ihr Inhalt muss für den Nutzer jederzeit abrufbar sein, § 13 Abs. 2 TMG.

Nutzern von Telemedien ist grundsätzlich auch eine Nutzung und Bezahlung der Telemedien anonym oder unter Pseudonym zu ermöglichen, soweit dies technisch möglich und zumutbar ist, § 13 Abs. 6 TMG. Für Big-Data-Anwendungen ist grundsätzlich zu klären, inwieweit überhaupt eine anonymisierte oder pseudonymisierte Nutzung im Einzelfall in Betracht kommt. Daneben ist von Seiten der Rechtswissenschaft weiter zu klären, wann die Bereitstellung einer solchen Nutzungsmöglichkeit für den Diensteanbieter noch zumutbar und wann die Grenze der Unzumutbarkeit erreicht ist. Dabei wird auch bei Big-Data-Anwendungen von der Forschung zu berücksichtigen sein, dass die Möglichkeit derartiger Nutzung auch der Sicherung der Meinungsfreiheit gem. Art. 5 Abs. 1 S. 1 GG dient.²²⁶

Probleme im Zusammenhang mit Big-Data-Anwendungen können sich gerade aus der frühzeitigen Unterrichtungspflicht gem. § 13 Abs. 1 TMG ergeben. Der Nutzer ist zu Beginn des Nutzungsvorgangs über Art, Umfang und Zweck der Erhebung seiner Daten zu unterrichten; erfolgt die Verarbeitung der Daten außerhalb der EU bzw. des EWR, ist der Nutzer auch darüber zu unterrichten.

Der Nutzer soll so in die Position versetzt werden, die Folgen der Datenverarbeitung und Verbreitung seiner Daten abzuschätzen.²²⁷ In Hinblick auf Big-Data-Anwendungen ist zunächst zu klären, inwieweit von vornherein sicher bestimmt werden kann, ob eine Verarbeitung nur innerhalb von EU / EWR stattfindet, oder ob aufgrund der hohen Standortflexibilität von Daten (insbesondere Cloud Computing) generell eine Aufklärung über eine außereuropäische Datenverarbeitung ratsam ist. Dabei stellt sich für die weitere Forschung auch die Frage, inwieweit eine solche prophylaktische Aufklärung überhaupt zulässig ist.

²²⁴ Müller-Broich, § 11 Rn. 8; Spindler/Nink, in: Spindler/Schuster, 12. Teil, § 11 Rn. 15.

²²⁵ Holznagel/Ricke, in: Spindler/Schuster, 12. Teil, § 1 Rn. 10.

²²⁶ Müller-Broich, § 14 Rn. 10.

²²⁷ Spindler/Nink , in: Spindler/Schuster, 12. Teil, § 13 Rn. 3.

Besondere Erlaubnisnormen finden sich in den §§ 14, 15 TMG, welche die sog. „Bestands- und Nutzungsdaten“ bei Telemediendaten betreffen. Bestandsdaten gem. § 14 TMG sind die Daten, die für die Begründung, inhaltliche Ausgestaltung oder Änderung eines Vertragsverhältnisses zwischen dem Diensteanbieter und dem Nutzer über die Nutzung von Telemedien erforderlich sind. Sie dürfen vom Diensteanbieter erhoben und verarbeitet werden, soweit sie zur Durchführung des Vertragsverhältnisses erforderlich sind. Erforderlich sollen dabei nur diejenigen Daten sein, die unverzichtbar für den konkreten Vertrag sind.²²⁸

Ferner können gem. § 15 TMG Nutzungsdaten erhoben und verwendet werden. Nutzungsdaten sind personenbezogene Daten, die anlässlich der Nutzung entstehen und die erforderlich sind, um die Inanspruchnahme von Telemedien zu ermöglichen und abzurechnen. Der Maßstab der Erforderlichkeit bestimmt sich wie bei § 14 TMG.

Diensteanbieter dürfen gem. § 15 Abs. 3 TMG für Zwecke der Werbung, der Marktforschung und zur bedarfsgerechten Gestaltung der Telemedien Nutzungsprofile bei Verwendung von Pseudonymen erstellen, sofern kein Widerspruch des Nutzers vorliegt. Dabei ist der Nutzer auf sein Widerspruchsrecht hinzuweisen und die Nutzungsprofile dürfen nicht mit Daten über den Träger des Pseudonyms zusammengeführt werden.

4. Datenverarbeitung – Verbot mit Erlaubnisvorbehalt

Das deutsche Datenschutzrecht ist geprägt durch das Volkszählungsurteil²²⁹ des Bundesverfassungsgerichts, das dem Einzelnen das Recht zusichert, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen. Dieses Urteil wirkt sich auf die Regelungsgrundsätze des Bundesdatenschutzgesetzes (BDSG) aus. Einer der zentralen Grundsätze ist das sog. „Verbot mit Erlaubnisvorbehalt“ für den Umgang mit personenbezogenen Daten. Dieser Grundsatz ist in § 4 Abs. 1 BDSG kodifiziert und besagt, dass die Erhebung, Verarbeitung und Nutzung personenbezogener Daten nur zulässig ist, soweit dies gesetzlich gestattet ist oder der Betroffene eingewilligt hat. Für den Bereich der Telemedien gilt die inhaltsgleiche Bestimmung des § 12 Abs. 1 TMG.

Grundsätzlich kommt derzeit ein Datenumgang in Betracht, wenn der Betroffene seine Einwilligung hierzu erteilt hat oder eine gesetzliche Ermächtigungsgrundlage vorliegt. Als Ermächtigungsgrundlage des Datenumgangs kommen „alle materiellen Rechtsnormen mit unmittelbarer Außenwirkung“²³⁰ in Betracht. Dies gilt ebenso für „Satzungen der bundesmittelbaren Körperschaften, Anstalten und Stiftungen des öffentlichen Rechts“²³¹ wie für den „normativen Teil von Tarifverträgen und Betriebsvereinbarungen“²³².

²²⁸ Müller-Broich, § 14 Rn. 3.

²²⁹ BVerfG, NJW 1984, 419.

²³⁰ Sokol, in: Simitis, § 4 Rn. 9.

²³¹ Taeger, in: Taeger/Gabel, § 4 Rn. 22.

²³² Gola/Schomerus, § 4 Rn. 10; Sokol, in: Simitis, § 4 Rn. 11.

Die Regelung des Verbots mit Erlaubnisvorbehalt wird zunehmend als untauglich kritisiert. Das Gesetz geht dabei immer noch von der zentralen Datenverarbeitung einer Datei mit personenbezogenen Daten in Großrechnern aus, der sog. alten PC-Welt.²³³ Tatsächlich wird heute potentiell jedermann zum Datenverarbeiter.²³⁴ Jede Äußerung und Lebensregung, die im Netz stattfindet, ist schon ein unmittelbar verarbeitetes Datum.²³⁵ Durch die „Ubiquität der Datenverarbeitung“²³⁶ in Smartphones etwa findet Datenverarbeitung beinahe in jeder Lebensphase statt.

Das Verbot wird folglich „sowohl als zu scharf als auch gleichzeitig zu weich“²³⁷ bezeichnet und sogar als „Hemmnis für einen modernen Datenschutz“²³⁸ ausgemacht. Die Grundannahme des Gesetzes, nach der zunächst alle Verarbeitungen verboten sind, berücksichtigt nicht sämtliche Grundrechte der Beteiligten, sondern stellt den Schutz der informationellen Selbstbestimmung über alles weitere, um dann aber sehr viele und zum Teil sehr weite Erlaubnistratbestände vorzusehen.²³⁹ Die Ausnahmen werden somit zur eigentlichen Regel.²⁴⁰ Bei einem generellen Verbot bleibt unberücksichtigt, dass die moderne Informationsgesellschaft die Autonomie des Einzelnen erfordert, damit dieser „selbst seine Daten verarbeiten, verbreiten und ebenso auch die Daten anderer verarbeiten und verbreiten“²⁴¹ kann. Im Rahmen weiterer Forschung zu Big-Data-Anwendungen ist zu überprüfen, inwieweit sich bei einer effektsteigernden Modernisierung des Datenschutzes das Verbotsprinzip noch als tauglich erweist.²⁴² Dabei wird auch zu erforschen sein, inwieweit das Konzept der Privatsphäre bzw. Privacy als eigentlicher Schutzgegenstand stärker betont werden kann.²⁴³

Die USA, Sitzland zahlreicher Internetfirmen, stehen dazu in deutlichem Gegensatz und setzen der Datenauswertung durch Unternehmen nur geringe rechtliche Grenzen.²⁴⁴

a) Datenqualität – Personenbezogene Daten

Der Begriff des „personenbezogenen Datums“ ist zentral im Zusammenhang mit Big-Data-Anwendungen: Sobald ein Umgang mit personenbezogenen Daten stattfindet, ist der Anwendungsbereich des BDSG eröffnet, § 1 Abs. 1 u. 2 BDSG.²⁴⁵

²³³ Hoeren, ZD 2011, 145 (145); Roßnagel/Pfitzmann/Garstka, Modernisierung des Datenschutzrechts, S. 22; Schneider, AnwBI 2011, 233 (236).

²³⁴ Schneider, AnwBI 2011, 233 (236).

²³⁵ Roßnagel/Pfitzmann/Garstka, Modernisierung des Datenschutzrechts, S. 22.

²³⁶ Roßnagel/Pfitzmann/Garstka, Modernisierung des Datenschutzrechts, S. 22.

²³⁷ Hoeren, ZD 2011, 145 (145).

²³⁸ Schneider, AnwBI 2011, 233 (233).

²³⁹ Hoeren, ZD 2011, 145 (145); Schneider, AnwBI 2011, 233 (233).

²⁴⁰ Schneider, AnwBI 2011, 233 (233).

²⁴¹ Schneider, AnwBI 2011, 233 (236).

²⁴² Schneider, AnwBI 2011, 233 (234).

²⁴³ Schneider, AnwBI 2011, 233 (236).

²⁴⁴ Weichert, ZD 2013, 251 (252, 254).

²⁴⁵ Gola/Schomerus, § 1 Rn. 22; Kühling/Seidel/Sivridis, S. 79; Sokol, in: Simitis, § 1 Rn. 57.

Eine große Zahl von Big-Data-Anwendungen/-Lösungen beinhaltet potentiell den Umgang mit derartigen Daten. Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person, § 3 Abs. 1 BDSG. Sie entstehen insbesondere in sozialen Netzwerken, dem sog. Web 2.0.²⁴⁶

Im Rahmen einer Big-Data-Anwendung ist stets im Einzelfall zu fragen, ob eine Verwendung personenbezogener Daten stattfindet. Das Bundesverfassungsgericht hat in seiner „Volkszählungsentscheidung“ festgestellt, dass „es unter den Bedingungen der automatisierten Datenverarbeitung kein ‚belangloses‘ Datum mehr“²⁴⁷ gibt. Im Sinne eines effektiven Grundrechtsschutzes sind somit alle eine Person betreffende Daten vom Begriff des personenbezogenen Datums erfasst.²⁴⁸

Die Daten müssen den Bezug zu einer bestimmten oder bestimmbaren natürlichen Person zulassen. Lässt sich aus einer Information die eindeutige Identität einer Person ermitteln, so handelt es sich um ein personenbezogenes Datum einer bestimmten Person.²⁴⁹ Kann aus einer Information allein noch kein Personenbezug hergestellt werden, ergibt sich dieser aber im Zusammenhang mit weiteren Angaben, so handelt es sich um personenbezogene Daten einer bestimmbaren Person.²⁵⁰

Vielfach erfolgt bei Big-Data-Anwendungen ein Umgang mit anonymisierten Daten, so dass vermeintlich kein Personenbezug vorliegt. Dabei steigt mit der Menge an vorhandenen Daten auch die Identifizierbarkeit einer bestimmten Person.²⁵¹ So sollen bereits mit drei einfachen Merkmalen wie Geschlecht, Postleitzahl und Geburtsdatum 61 – 87 Prozent der US-amerikanischen Bevölkerung identifizierbar sein.²⁵² Es wird daher weiterer Untersuchungen bedürfen, ob Big Data als Ansammlung großer Datenmengen trotz Anonymisierung immer zu einer Bestimmbarkeit führt²⁵³ oder ob dieser Schluss zu weit geht. Dabei wird vorgeschlagen, bereits die Quelldaten zu anonymisieren, um so das Risiko der De-Anonymisierung zu verringern.²⁵⁴

Für die Anwendbarkeit des BDSG ist diese Unterscheidung zwischen der bestimmten oder bestimmbaren Person ohne Belang.²⁵⁵ Kein Personenbezug ist gegeben, wenn die Bestimmbarkeit einer Person objektiv ausgeschlossen ist.²⁵⁶

Im Rahmen der Bestimmbarkeit kommt es daher maßgeblich auf das Zusatzwissen der verarbeitenden Stelle an.²⁵⁷ Soweit nicht die verarbeitende Stelle selbst, sondern Dritte über

²⁴⁶ Weichert, ZD 2013, 251 (252).

²⁴⁷ BVerfG, NJW 1984, 419 (422).

²⁴⁸ Dammann, in: Simitis, § 3 Rn. 7; Kühling/Seidel/Sivridis, S. 79; Sokol, in: Simitis, § 1 Rn. 57.

²⁴⁹ Kühling/Seidel/Sivridis, S. 80.

²⁵⁰ Kühling/Seidel/Sivridis, S. 81.

²⁵¹ Baeriswyl, digma 2013, 14 (15); Mayer-Schönberger/Cukier (2013), p.155; Weichert, ZD 2013, 251 (257).

²⁵² Baeriswyl, digma 2013, 14 (15).

²⁵³ Baeriswyl, digma 2013, 14 (15); Mayer-Schönberger/Cukier (2013), p. 154; Weichert, ZD 2013, 251 (258).

²⁵⁴ Weichert, ZD 2013, 251 (254).

²⁵⁵ Kühling/Seidel/Sivridis, S. 81; Plath/Schreiber, in: Plath, § 3 Rn. 12.

²⁵⁶ Buchner in: Taeger/Gabel, § 3 Rn. 13.

relevantes Zusatzwissen verfügen, ist dieses zu berücksichtigen, „soweit es ‚vernünftigerweise‘ eingesetzt wird“.²⁵⁸

b) Der Datenumgang

Das BDSG unterscheidet im Bereich des Datenumgangs zwischen der Erhebung, Verarbeitung und Nutzung von Daten, § 1 Abs. 2 BDSG. Die konkrete Ausgestaltung des Datenumgangs orientiert sich am Gesetzeszweck, nämlich den Einzelnen davor zu schützen, dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird, § 1 Abs. 1 BDSG. Daraus folgt auch das zuvor beschriebene Verbot mit Erlaubnisvorbehalt des Datenumgangs.

(1) Erhebung

Mit Erhebung bezeichnet das BDSG das Beschaffen von Daten über den Betroffenen, § 3 Abs. 3 BDSG. Eine Erhebung liegt nur dann vor, wenn die erhebende Stelle selbst aktiv tätig wird.²⁵⁹ Keine Erhebung liegt hingegen vor, wenn Informationen der Stelle ohne eigene Veranlassung zuwachsen.²⁶⁰ Gemäß § 4 Abs. 2 S. 1 BDSG sind personenbezogene Daten grundsätzlich beim Betroffenen zu erheben, sog. „Grundsatz der Direkterhebung“.²⁶¹ Im Falle der Direkterhebung ist der Betroffene über die Identität der verantwortlichen Stelle (§ 4 Abs. 3 Nr. 1 BDSG), die Zweckbestimmung der Erhebung, Verarbeitung oder Nutzung (§ 4 Abs. 3 Nr. 2 BDSG) und die Kategorien von Empfängern, soweit der Betroffene nach den Umständen des Einzelfalles nicht mit der Übermittlung an diese rechnen muss (§ 4 Abs. 3 Nr. 3 BDSG), zu unterrichten.

Eine Erhebung ohne Mitwirkung ist nur möglich, wenn eine Rechtsvorschrift dies vorsieht oder zwingend voraussetzt (§ 4 Abs. 2 Nr. 1 BDSG) oder die zu erfüllende Verwaltungsaufgabe ihrer Art nach oder der Geschäftszweck eine Erhebung bei anderen Personen oder Stellen erforderlich macht (§ 4 Abs. 2 Nr. 2 lit. a) BDSG). Ferner ist die Direkterhebung entbehrlich, wenn sie einen unverhältnismäßigen Aufwand erfordern würde (§ 4 Abs. 2 Nr. 2 lit. b) BDSG). Schließlich dürfen keine Anhaltspunkte dafür bestehen, dass überwiegende schutzwürdige Interessen des Betroffenen beeinträchtigt werden, § 4 Abs. 2 a.E. BDSG.

Für den Bereich der Big-Data-Anwendungen ist somit zu prüfen, ob im Einzelfall eine Erhebung direkt beim Betroffenen zumutbar möglich ist und inwieweit eine Information über den Verarbeitungsvorgang und die verantwortliche Stelle erforderlich ist. Erst sofern eine Direkterhebung nicht in Betracht kommt, kann auf andere Wege der Erhebung zurückgegriffen werden. Erweiterter Forschungsbedarf besteht hier insbesondere für die

²⁵⁷ Dammann, in: Simitis, § 3 Rn. 26; Gola/Schomerus, § 3 Rn. 10; Kühling/Seidel/Sivridis, S. 81 f; Plath/Schreiber, in: Plath, § 3 Rn. 15.

²⁵⁸ Buchner in: Taeger/Gabel, § 3 Rn. 12.

²⁵⁹ Gola/Schomerus, § 3 Rn. 24; Kühling/Seidel/Sivridis, S. 88 f; Plath/Schreiber, in: Plath, § 3 Rn. 30.

²⁶⁰ Dammann, in: Simitis, § 3 Rn. 103; Gola /Schomerus, § 3 Rn. 24; Kühling/Seidel/Sivridis, S. 89; Plath/Schreiber, in: Plath, § 3 Rn. 30.

²⁶¹ Kühling/Seidel/Sivridis, S. 89.

Bestimmung des „unverhältnismäßigen Aufwands“, der eine Direkterhebung im Rahmen von Big-Data-Anwendungen entbehrlich macht, vgl. § 4 Abs. 2 Nr. 2 lit. b) BDSG.

(2) Verarbeitung

Soweit das BDSG von „Verarbeitung“ spricht, ist damit kein einheitlicher Begriff verbunden.²⁶² Gem. § 3 Abs. 4 BDSG ist Verarbeiten das Speichern, Verändern, Übermitteln, Sperren und Löschen personenbezogener Daten. Speichern ist das Erfassen, Aufnehmen oder Aufbewahren personenbezogener Daten auf einem Datenträger zum Zweck ihrer weiteren Verarbeitung oder Nutzung (§ 3 Abs. 4 Nr. 1 BDSG). Verändern ist das inhaltliche Umgestalten gespeicherter personenbezogener Daten (§ 3 Abs. 4 Nr. 2 BDSG). Übermitteln ist das Bekanntgeben gespeicherter oder durch Datenverarbeitung gewonnener personenbezogener Daten an einen Dritten in der Weise, dass die Daten an den Dritten weitergegeben werden oder der Dritte zur Einsicht oder zum Abruf bereitgehaltene Daten einsieht oder abruft (§ 3 Abs. 4 Nr. 3 BDSG). Ferner bezeichnet Sperren das Kennzeichnen gespeicherter personenbezogener Daten, um ihre weitere Verarbeitung oder Nutzung einzuschränken (§ 3 Abs. 4 Nr. 4 BDSG) und Löschen das Unkenntlichmachen gespeicherter personenbezogener Daten (§ 3 Abs. 4 Nr. 5 BDSG).

(3) Nutzung

Nutzung ist jede Verwendung personenbezogener Daten, soweit es sich nicht um Verarbeitung handelt, § 3 Abs. 5 BDSG. Der Begriff der Nutzung ist weit angelegt. Es ist nicht entscheidend, zu welchem Zweck und von wem die Daten genutzt werden.²⁶³ Schon die Kenntnisnahme ist eine Nutzung i.S.d. § 3 Abs. 5 BDSG.²⁶⁴ Big-Data-Anwendungen scheinen somit jedenfalls immer dieser Kategorie des Datenumgangs zu unterfallen.

c) § 6a BDSG – Verbot der automatisierten Einzelentscheidung

Das BDSG bestimmt in § 6a, dass Entscheidungen, die für den Betroffenen eine rechtliche Folge nach sich ziehen oder ihn erheblich beeinträchtigen, nicht ausschließlich auf eine automatisierte Verarbeitung personenbezogener Daten gestützt werden dürfen, die der Bewertung einzelner Persönlichkeitsmerkmale dienen. Eine ausschließlich auf eine automatisierte Verarbeitung gestützte Entscheidung liegt insbesondere dann vor, wenn eine inhaltliche Bewertung und darauf gestützte Entscheidung durch eine natürliche Person nicht stattgefunden hat.

Dabei bleibt die automatisierte Einzelentscheidung zulässig, wenn die Entscheidung im Rahmen des Abschlusses oder der Erfüllung eines Vertragsverhältnisses oder eines sonstigen Rechtsverhältnisses ergeht und dem Begehr des Betroffenen stattgegeben wurde (§ 6a Abs. 2 Nr. 1), oder die Wahrung der berechtigten Interessen des Betroffenen durch geeignete Maßnahmen gewährleistet ist und die verantwortliche Stelle dem Betroffenen die Tatsache des Vorliegens einer automatisierten Einzelentscheidung mitteilt sowie auf

²⁶² Gola/Schomerus, § 3 Rn. 25; Kühling/Seidel/Sivridis, S. 90.

²⁶³ Dammann, in: Simitis, § 3 Rn. 189; Kühling/Seidel/Sivridis, S. 95.

²⁶⁴ Gola/Schomerus, § 3 Rn. 42; Plath/Schreiber, in: Plath, § 3 Rn. 54.

Verlangen die wesentlichen Gründe dieser Entscheidung mitteilt und erläutert (§ 6a Abs. 2 Nr. 2). Die Vorschrift zielt insgesamt auf „die elektronische Auswertung von Persönlichkeitsprofilen“.²⁶⁵

Im Zusammenhang mit Big Data ist vor allem zu erforschen, wann ein Verbot der automatisierten Einzelentscheidung greift und wann eine Fallgestaltung vorliegt, nach der eine solche zulässig ist. Zunächst ist zu beachten, dass der Betroffene nicht zum bloßen Objekt einer automatisierten Verarbeitung degradiert werden darf.²⁶⁶ Sobald Merkmale einen besonderen Bezug zum Persönlichkeitsrecht aufweisen, soll die Verantwortung für den Inhalt einer Entscheidung auch bei einer natürlichen Person liegen.²⁶⁷ Die menschliche Bearbeitung darf dabei keine bloße Formalie sein,²⁶⁸ die letztendliche Entscheidung kann sich allerdings „auf einen automatisch erzeugten Entscheidungsvorschlag stützen.“²⁶⁹ Von großer praktischer Bedeutung im Rahmen der automatisierten Einzelentscheidung sind die sog. Scoring-Verfahren.²⁷⁰ Die Erhebung und Verwendung von Wahrscheinlichkeitswerten im Scoring-Verfahren hat in § 28b BDSG eine eigene Regelung gefunden.

Eine automatisierte Einzelentscheidung kann jedoch zulässig sein, wenn dem Begehrten des Betroffenen stattgegeben wird oder sofern die berechtigten Interessen des Betroffenen gewährleistet werden.

Eine besondere Forschungsperspektive für Big-Data-Anwendungen besteht in der Ausarbeitung von Verfahrensgestaltungen, die einerseits die Möglichkeiten großer Datenmengen wirtschaftlich ausschöpfen und andererseits die Wahrung der berechtigten Interessen der Betroffenen garantieren und somit auch im Rahmen einer automatisierten Einzelentscheidung Bestandsfähigkeit der Entscheidung und Rechtssicherheit bieten. Von Bedeutung in diesem Zusammenhang ist die Frage, inwieweit überhaupt noch eine menschliche Überprüfung getroffener Entscheidungen stattzufinden hat bzw. auf welche Ereignisse hin (bspw. Beschwerde eines Betroffenen) eine solche erforderlich ist, um die „personale Verantwortung“²⁷¹ sicherzustellen und die Interessen des Betroffenen zu wahren.²⁷²

Grundsätzlich sollen sogar den Betroffenen beschwerende Maßnahmen möglich sein, die ohne „Zwischenschaltung menschlicher Urteilsbildung“²⁷³ auskommen. Im Rahmen

²⁶⁵ *Gola/Schomerus*, § 6a Rn. 7; *Kamlah*, in: *Plath*, § 6a Rn. 5.

²⁶⁶ *Scholz*, in: *Simitis*, § 6a Rn. 3.

²⁶⁷ *Mackenthun* in: *Taeger/Gabel*, § 6a Rn. 1.

²⁶⁸ *Kamlah*, in: *Plath*, § 6a Rn. 12; *Scholz*, in: *Simitis*, § 6a Rn. 14.

²⁶⁹ *Kamlah*, in: *Plath*, § 6a Rn. 5; *Scholz*, in: *Simitis*, § 6a Rn. 16.

²⁷⁰ *Gola/Schomerus*, § 6a Rn. 15; *Kamlah*, in: *Plath*, § 6a Rn. 11; *Möller/Florax*, MMR 2002, 806 (806); *Scholz*, in: *Simitis*, § 6a Rn. 17.

²⁷¹ *Scholz*, in: *Simitis*, § 6a Rn. 32.

²⁷² *Kamlah*, in: *Plath*, § 6a Rn. 12; *Scholz*, in: *Simitis*, § 6a Rn. 34.

²⁷³ *Kamlah*, in: *Plath*, § 6a Rn. 20.

zukünftiger Forschung wird zu klären sein, welche Maßnahmen (neben einer möglichen Stellungnahme des Betroffenen) zur Interessenwahrung in Betracht kommen.²⁷⁴

Beachtlich ist, dass sich gemäß § 6a Abs. 3 BDSG das Recht des Betroffenen auf Auskunft auch auf den logischen Aufbau der automatisierten Verarbeitung der ihn betreffenden Daten bezieht. Im Rahmen der Big-Data-Anwendungen ist somit besonderer Wert auf eine ordnungsgemäße Dokumentation zu legen.

5. Erlaubnisgründe

Die Tatsache, dass die Verarbeitung personenbezogener Daten unter einem Verbot mit Erlaubnisvorbehalt steht, führt dazu, dass den verschiedenen Erlaubnisgründen für die Zulässigkeit des Big-Data-Managements eine erhebliche Bedeutung zukommt.²⁷⁵ Eines Erlaubnisgrundes bedarf es alleine dann nicht, wenn es sich bei den verarbeiteten Daten um nicht-personenbezogene Daten handelt.

a) Einwilligung

Besondere Schwierigkeiten stellen sich bei Big Data im Zusammenhang mit der Einwilligung gem. § 4a BDSG, die in der Praxis die am häufigsten gewählte Möglichkeit der Erlaubnisgründe darstellt.²⁷⁶ Das massenhafte Einholen von Einwilligungen kann dabei schon logistische Probleme verursachen:²⁷⁷ Die Einwilligung ist vor der Datenverarbeitung regelmäßig in Schriftform einzuholen und muss auf der freien Entscheidung des Betroffenen beruhen. Der Betroffene ist umfassend zu unterrichten.²⁷⁸

Der Betroffene muss in die Lage versetzt werden, im Vorfeld „Anlass, Ziel und Folgen der Verarbeitung korrekt abzuschätzen“.²⁷⁹ Die Einwilligung muss inhaltlich so bestimmt sein, dass sich ein eindeutiger konkretisierter Verarbeitungsvorgang mit ihr verbindet.²⁸⁰ Wichtig ist, dass dem Einwilligenden die Folgen der Erteilung stets klar vor Augen geführt werden – ist dies nicht oder nicht ordnungsgemäß geschehen, folgt daraus die Unwirksamkeit der Einwilligung.²⁸¹

Im Bereich der Telemediendienste kann die Einwilligung auch elektronisch erklärt werden, § 13 Abs. 2 TMG. Dabei ist zusätzlich sicherzustellen, dass die Einwilligung protokolliert und der Nutzer deren Inhalt jederzeit abrufen kann.

Gesetzliche Erlaubnistratbestände und Einwilligung stehen nicht beliebig zueinander: Die Einwilligung soll Ausdruck einer echten Wahlmöglichkeit des Betroffenen sein und ihm seine tatsächliche Handlungsmacht deutlich machen.²⁸² Keine zulässige Einwilligung des

²⁷⁴ Gola/Schomerus, § 6a Rn. 14b.

²⁷⁵ Weichert, ZD 2013, 251 (255).

²⁷⁶ Mayer-Schönberger/Cukier (2013), p. 153.

²⁷⁷ Weichert, ZD 2013, 251 (255).

²⁷⁸ Plath, in: Plath, § 4a Rn. 31; Simitis, § 4a Rn. 70.

²⁷⁹ Gola/Schomerus, § 4a Rn. 25; Plath, in: Plath, § 4a Rn. 31.

²⁸⁰ Taeger, in: Taeger/Gabel, § 4a Rn. 29.

²⁸¹ Plath, in: Plath, § 4a Rn. 82; Sokol, in: Simitis, § 4 Rn. 7.

²⁸² Kühling/Seidel/Sivridis, S. 107.

Betroffenen liegt folglich in den Fällen vor, in denen keine tatsächliche Wahlmöglichkeit besteht, etwa im Falle von Verarbeitungspflichten.²⁸³ Dort wo ausreichende gesetzliche Ermächtigungsgrundlagen bestehen, kommt die Einholung der Einwilligung nur in Betracht, wenn bei Verweigerung der Einwilligung auch tatsächlich auf den Datenumgang verzichtet und nicht hilfsweise auf den gesetzlichen Erlaubnistanstbestand zurückgegriffen wird.²⁸⁴

Im Rahmen von Big-Data-Anwendungen ist genau zu prüfen, in welchen Fällen sich ein Rückgriff auf gesetzliche Erlaubnistanstbestände anbietet und wann eine Einwilligung einzuholen ist.

Kennzeichnend für Big Data ist, dass bereits vorhandene Daten auf neue Weise analysiert, bzw. die vorhandenen Daten mit weiteren verfügbaren Daten zusammengeführt werden, um neue Erkenntnisse zu erhalten.²⁸⁵ Ein solches Vorgehen ist derzeit immer erst nach der Einholung einer neuen Einwilligung möglich – vorausgesetzt, es existiert kein gesetzlicher Erlaubnistanstbestand.²⁸⁶ Diese Praxis hemmt die Entwicklung von Big Data deutlich.

Soweit es sich um Abhängigkeitsverhältnisse handelt, ist besonders darauf zu achten, auf welche Weise die Art der Einholung erfolgt:²⁸⁷ In der derzeitigen Situation kann ein Ungleichgewicht zwischen Betroffenem und verantwortlicher Stelle dazu führen, dass diese „auf Grund ihrer Machtposition ihre Verarbeitungswünsche ohne Rücksicht“²⁸⁸ auf den Betroffenen durchsetzt. Die Freiwilligkeit der Erteilung ist jedoch zu wahren und darf nicht unzulässigerweise mit anderen Entscheidungen verknüpft werden, sog. Kopplungsverbot.²⁸⁹

Aus Big-Data-Sicht ist es daher besonders beachtenswert, dass gem. Art. 7 Abs. 4 DSGVO-E eine Einwilligung als Rechtsgrundlage für die Verarbeitung ausgeschlossen sein soll, „wenn zwischen der Position der betroffenen Person und des für die Verarbeitung Verantwortlichen ein erhebliches Ungleichgewicht besteht.“²⁹⁰

Die Bestimmung der genauen Ansatzpunkte zur Ermittlung eines erheblichen Ungleichgewichtes ist entscheidend für die Durchführung von Big-Data-Anwendungen/-Lösungen, da nur so rechtssicher zu bestimmen ist, wann die Einholung einer Einwilligung ausgeschlossen ist. Ein Forschungsschwerpunkt wird die Frage sein, ob bereits die Zuordnung einer verantwortlichen Stelle zu einem bestimmten Wirtschaftssektor (Grundversorgung, Gesundheitswesen usw.) zur Annahme eines erheblichen Ungleichgewichtes führt,²⁹¹ oder ob hierfür (auch) eine gewisse Marktmacht erforderlich ist.

²⁸³ Sokol, in: Simitis, § 4 Rn. 7.

²⁸⁴ Kühling/Seidel/Sivridis, S. 107; Sokol, in: Simitis § 4 Rn. 6.

²⁸⁵ Weichert, ZD 2013, 251 (256).

²⁸⁶ Mayer-Schönberger/Cukier (2013), p. 153.

²⁸⁷ Plath, in: Plath, § 4a Rn. 27; Sokol, in: Simitis § 4 Rn. 7.

²⁸⁸ Roßnagel/Pfitzmann/Garstka, S. 91.

²⁸⁹ Kühling/Seidel/Sivridis, S. 106; Plath, in: Plath, § 4a Rn. 30; Weichert, ZD 2013, 251 (256).

²⁹⁰ KOM (2012) 11 endg.

²⁹¹ Gola/Schomerus, § 4a Rn. 20; Roßnagel/Pfitzmann/Garstka, S. 93.

Parallelen zum Verbraucherschutz, welches dem Schwächeren ebenfalls einen Grundschutz zusichern will, verdienen dabei besondere Beachtung.²⁹²

Gleichzeitig wird zu erforschen sein, welche Vorkehrungen etwa in Form zusätzlicher Verfahren zu schaffen sind, um die freie Selbstbestimmung des Betroffenen zu sichern.²⁹³

Im Zusammenhang mit den durch Big-Data-Anwendungen verfügbaren großen Datenmengen ist besonderes Augenmerk auf die Behandlung Minderjähriger zu legen.²⁹⁴ Die Einwilligung ist das Instrument des Persönlichkeitsschutzes.²⁹⁵ Eine effektive Ausübung dieser Schutzfunktion setzt jedoch die Einwilligungsfähigkeit voraus.²⁹⁶ Minderjährige sind jedoch oftmals zu gutgläubig und besitzen noch nicht die notwendige Erfahrung, um die Reichweite ihrer datenschutzrechtlichen Einwilligung abzusehen.²⁹⁷ Ein Lösungsweg könnte ein Verfallsdatum für Daten Minderjähriger sein, sofern diese die Verwendung ihrer Daten nicht mit Eintritt der Volljährigkeit bestätigen.²⁹⁸ Zukünftige Forschung wird sich der Frage widmen müssen, ab wann Minderjährige die notwendige Fähigkeit besitzen, die Reichweite ihrer Entscheidung angesichts der großen Datenmengen zu begreifen und ob und ggf. inwieweit ein Verfallsdatum für Daten ein Lösungsweg sein kann. Ohne eine Lösung ist die Persönlichkeitsentfaltung von Kindern gefährdet.

International wurden bereits verschiedene Lösungen gefunden. Die USA bestimmen etwa für den Online-Bereich mit dem Children Online Privacy Protection Act, dass bei unter 13-jährigen Kindern keine Daten ohne die Einwilligung der Eltern erhoben werden dürfen.²⁹⁹ In Europa findet sich etwa im niederländischen Datenschutzgesetz eine Regelung, die eine Zustimmung der gesetzlichen Vertreter vorsieht, bis die Minderjährigen das 16. Lebensjahr erreicht haben.³⁰⁰ In jedem Fall wird zukünftige Forschung gerade im Angesicht der besonderen Herausforderungen mit großen Datenmengen einen Ausgleich finden müssen, der einerseits auch effektiven Grundrechtsschutz Minderjähriger gewährleistet und andererseits Rechtssicherheit für die verantwortliche Stelle bietet.³⁰¹

Datenschutzrechtliche Einwilligungen als eine Art „Handelsgut“ erlangen bei Big-Data-Anwendungen große Bedeutung. Unentgeltliche Online- und Offline-Dienste sind keinesfalls kostenlos, sondern werden vielmehr im Rahmen eines Tauschgeschäfts (Einwilligung gegen Dienstleistung) abgewickelt.³⁰² Die Einwilligung gilt mittlerweile als „Kommerzialisierungsinstrument“³⁰³, das den spezifischen wirtschaftlichen Wert

²⁹² Roßnagel/Pfitzmann/Garstka, S. 91.

²⁹³ Roßnagel/Pfitzmann/Garstka, S. 91.

²⁹⁴ Roßnagel/Pfitzmann/Garstka, S. 95.

²⁹⁵ Rogosch, S. 47.

²⁹⁶ Plath, in: Plath, § 4a Rn. 8; Rogosch, S. 47.

²⁹⁷ Rogosch, S. 48.

²⁹⁸ Schneider, AnwBI 2011, 233 (239).

²⁹⁹ Rogosch, S. 51; Roßnagel/Pfitzmann/Garstka, S. 95.

³⁰⁰ Roßnagel/Pfitzmann/Garstka, S. 95.

³⁰¹ Rogosch, S. 51.

³⁰² Buchner, DuD 2010, 39 (39); Rogosch, S. 41.

³⁰³ Buchner, DuD 2010, 39 (39).

personenbezogener Daten einerseits widerspiegelt und andererseits auch abschöpft,³⁰⁴ denn in Zeiten von „Big Data“ sind Daten an sich bereits wertvoll.³⁰⁵ Zukünftige Forschung wird sich somit auch der Frage annehmen müssen, inwieweit ein solches Tauschgeschäft mit der Idee der Freiwilligkeit der Einwilligung vereinbar ist und ob und ggf. wie die Freiwilligkeit im Zuge eines immer sorgloseren Umgangs der Nutzer mit Einwilligungen sichergestellt werden kann. Daten und die Einwilligung zur Datenverarbeitung erweisen sich als modernes Zahlungsmittel,³⁰⁶ bei dem durch Transparenz über Datenbestände die informierte Selbstbestimmung der Verbraucher sichergestellt werden muss.³⁰⁷ Eine besondere Forschungsperspektive bei Big-Data-Anwendungen besteht daher in der Frage, welche gesetzgeberischen Konsequenzen sich aus der veränderten Natur der Einwilligung ergeben und wie sich die notwendige Transparenz, bspw. mit Hilfe einer Art Kontoauszug, für die Betroffenen herstellen lässt.³⁰⁸

Zurzeit sieht die DSGVO eine Einwilligung in Form einer expliziten Willensbekundung vor, Art. 4 Abs. 8 DSGV-E. Die Voraussetzung „explizit“ stellt allerdings höhere Anforderungen an die Erklärung der Einwilligung, als sie in dem derzeitig verwandten Ausdruck „ohne jeden Zweifel“ zu verstehen ist.³⁰⁹ Erwägungsgrund 25 führt dazu aus: „Eine stillschweigende Einwilligung ohne Zutun der betroffenen Person stellt daher keine Einwilligung dar.“ Als Beispiel wird im gleichen Erwägungsgrund etwa das „Anklicken eines Kästchens beim Besuch einer Internetseite“ genannt. Von einem allgemeinen Schriftformerfordernis, wie es derzeit besteht, hat die Kommission abgesehen.³¹⁰

Es wird besonders zu erforschen sein, welche geeigneten Formen der Einwilligung für Big-Data-Anwendungen zulässig sind und inwiefern den einzelnen Möglichkeiten der Big-Data-Anwendung zugestimmt werden kann. Das strenge Schriftformerfordernis des BDSG war ein deutsches Kuriosum und einmalig in Europa.³¹¹ Vorgeschlagen wird auch die Begrenzung der Wirkungsdauer einer Einwilligung.³¹² Zukünftige Forschung wird überprüfen müssen, inwieweit hierdurch einerseits eine verbesserte Praktikabilität erreicht und andererseits effektiver Grundrechtsschutz sichergestellt werden kann.

Bereits jetzt werden die bereichsspezifischen Anforderungen an die Einwilligung und die damit einhergehende Uneinheitlichkeit kritisiert.³¹³ Eine Vereinheitlichung wird für „möglich, wenn nicht sogar notwendig“³¹⁴ erachtet. Künftige Big-Data-Forschung wird untersuchen müssen, inwieweit die bereichsspezifischen Regelungen in Einklang zu bringen sind und wie

³⁰⁴ Rogosch, S. 41.

³⁰⁵ Mayer-Schönberger/Cukier (2013), p. 100.

³⁰⁶ Rogosch, S. 42.

³⁰⁷ Buchner, DuD 2010, 39 (41).

³⁰⁸ Schneider, AnwBl 2011, 233 (239).

³⁰⁹ Breinlinger/Scheuing, RDV 2012, 64 (70).

³¹⁰ Hoeren/Giurgiu, NWB 2012, 1599 (1602).

³¹¹ Drewes/Sieger, RDV 2006, 139, (144); Rogosch, S. 196.

³¹² Rogosch, S. 192.

³¹³ Rogosch, S. 192.

³¹⁴ Rogosch, S. 192.

eine zukünftige generelle Regelung auszusehen hat. Insbesondere ist herauszuarbeiten, wie eine vereinheitlichte Regelung der Einwilligung in § 4a BDSG ausgestaltet sein kann.³¹⁵

Im Rahmen der Forschung zu Big-Data-Anwendungen kommt daher einer grundsätzlichen und kritischen Überprüfung der Einwilligung grundlegende Bedeutung zu. Es ist nicht nur zu ermitteln, in welchen Fällen sich ein Rückgriff auf gesetzliche Erlaubnistatbestände anbietet und wann eine Einwilligung einzuholen ist. Im Hinblick auf die Entwicklung des Datenschutzrechts „von einem Persönlichkeitsrecht zu einem Datenrecht“³¹⁶ wird ferner zu erforschen sein, inwieweit die Einwilligung überhaupt noch ein angemessenes Mittel zur Herstellung eines gerechten Interessensausgleiches zwischen Wirtschaft und Persönlichkeitsschutz ist. Denn der Begriff der Einwilligung gilt nicht nur als zu unspezifisch und wird für „gänzlich ungeeignet“ gehalten.³¹⁷ Der Funktionsfähigkeit der Einwilligung steht auch ein „dramatisches Vollzugsdefizit und die zu gering ausfallenden Sanktionen bei einem Verstoß“³¹⁸ entgegen, so dass die Regelungen des BDSG auch als „zahnlos“ bezeichnet werden.³¹⁹

**b) Erlaubnisnormen für öffentliche und nicht-öffentliche Stellen
(§§ 12 ff. BDSG und §§ 28 ff. BDSG)**

Die im BDSG enthaltenen Rechtsgrundlagen der Datenverarbeitung differenzieren zwischen der Datenverarbeitung der öffentlichen Stellen (§§ 12 ff. BDSG) und der Datenverarbeitung nicht-öffentlicher Stellen und öffentlich-rechtlicher Wettbewerbsunternehmen (§§ 27 ff. BDSG).

(1) Öffentliche Stellen

Öffentliche Stellen verarbeiten Daten grundsätzlich aufgabenbezogen und unterliegen dabei einem strengen Zweckbindungsgrundsatz.³²⁰ Insoweit kommt eine Big-Data-Auswertung durch öffentliche Stelle nur aufgrund gesetzlicher Spezialregelungen in Betracht.³²¹

(2) Nicht-öffentliche Stellen

Grundsätzlich ist gem. § 28 Abs. 1 BDSG das Erheben, Speichern, Verändern oder Übermitteln personenbezogener Daten oder ihre Nutzung als Mittel für die Erfüllung eigener Geschäftszwecke zulässig, wenn es für die Begründung, Durchführung oder Beendigung eines rechtsgeschäftlichen oder rechtsgeschäftsähnlichen Schuldverhältnisses mit dem Betroffenen erforderlich ist. Ferner ist der Umgang gestattet, soweit er zur Wahrung berechtigter Interessen der verantwortlichen Stelle erforderlich ist und kein Grund zu der Annahme besteht, dass das schutzwürdige Interesse des Betroffenen an dem Ausschluss der Verarbeitung oder Nutzung überwiegt. Schließlich kann ein Umgang erfolgen, wenn die

³¹⁵ Rogosch, S. 193.

³¹⁶ Hoeren, ZD 2011, 145 (145).

³¹⁷ Hoeren, ZD 2011, 145 (145).

³¹⁸ Rogosch, S. 194 f.

³¹⁹ Schneider, MMR 2009, VII, VIII.

³²⁰ Weichert, ZD 2013, 251 (254).

³²¹ Weichert, ZD 2013, 251 (254).

Daten allgemein zugänglich sind oder die verantwortliche Stelle sie veröffentlichen dürfte, es sei denn, dass das schutzwürdige Interesse des Betroffenen an dem Ausschluss der Verarbeitung oder Nutzung gegenüber dem berechtigten Interesse der verantwortlichen Stelle offensichtlich überwiegt.

Für Big-Data-Anwendungen/-Lösungen ist insbesondere die Kategorie der allgemein zugänglichen Daten gem. § 28 Abs. 1 Nr. 3 BDSG von erhöhtem Forschungsinteresse, denn die Art und Weise der Verwendung sowie deren Umfang und Zweck kann frei bestimmt werden.³²²

Derartige Daten liegen vor, wenn sie „sich sowohl ihrer Zielsetzung als auch ihrer Publikationsform nach dazu eignen, einem individuell nicht bestimmmbaren Personenkreis Informationen zu vermitteln.“³²³ Grundsätzlich zählen hierzu auch Angaben auf Internetseiten.³²⁴

Eine besondere Forschungsperspektive ergibt sich aus der Frage, wann Daten in sozialen Netzwerken als allgemein zugänglich, also öffentlich, zu beurteilen sind. Mittlerweile wird bereits davon ausgegangen, dass selbst die Erforderlichkeit einer Facebook-Mitgliedschaft dieser Annahme nicht entgegensteht³²⁵ und Daten aus sozialen Netzwerken jedenfalls dann als öffentlich einzustufen sind, wenn diese ohne „Authentisierung oder Autorisierung“ einsehbar sind.³²⁶ Hier wird dann weiter zu differenzieren sein, ob auch Informationen unter „Netzwerk-Freunden“ ggf. als öffentlich eingestuft werden können und ab welcher Anzahl von Freunden das der Fall ist³²⁷ oder ob diese nicht mehr als öffentlich anzusehen sind.³²⁸

c) Anonymisierung, Pseudonymisierung & De-Personalisierung

Das BDSG steht dem Big-Data-Management auch dann nicht im Wege, wenn die verarbeiteten Daten ihren Personenbezug verlieren.³²⁹ Anonymisierung und Pseudonymisierung von Daten stellen vielmehr eine Möglichkeit dar, die Auswertung von Big Data zu legitimieren. Anonymisieren ist das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmmbaren natürlichen Person zugeordnet werden können, § 3 Abs. 6 BDSG. Pseudonymisieren ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren, § 3 Abs. 6a BDSG. Sowohl bei Anonymisierung als auch im Rahmen der Pseudonymisierung muss unter den verschiedenen zur Verfügung stehenden Verfahren dasjenige gewählt werden, das die Möglichkeit ausschließt, anonymisierte Daten

³²² Weichert, ZD 2013, 251 (255).

³²³ Gola/Schomerus, § 28 Rn. 151.

³²⁴ Gola/Schomerus, § 28 Rn. 151; Plath, in: Plath, § 28 Rn. 76; Weichert, ZD 2013, 251 (257).

³²⁵ Ostendorf/Frahm/Doege, NStZ 2012, 529 (532).

³²⁶ Weichert, ZD 2013, 251 (257).

³²⁷ LG Aachen m. Anm. Uecker, MMR 2013, 269 (270).

³²⁸ Weichert, ZD 2013, 251 (257).

³²⁹ Weichert, ZD 2013, 251 (258).

wieder zu deanonymisieren.³³⁰ Die identifizierenden Daten müssen gelöscht oder zumindest derart verändert werden, dass eine (auch zufällige) Zuordnung nicht mehr möglich ist.³³¹ Auch wenn eine andere als die verarbeitende Stelle über das für eine De-Anonymisierung notwendige Zusatzwissen verfügt, liegt kein Fall der Anonymisierung mehr vor.³³²

Hier besteht insbesondere im Zusammenspiel mit der Informatik eine Perspektive zur Entwicklung solcher Big-Data-Verfahren, die eine Re-Identifikation sicher verhindern.

Bei Telemedien, ist gem. § 13 Abs. 6 TMG eine anonyme oder pseudonymisierte Nutzung zu ermöglichen, soweit dies technisch möglich und zumutbar ist. Soweit die Nutzung mittels Pseudonym oder anonym ermöglicht werden soll, beinhaltet dies jedoch keine Pflicht zur Begründung eines anonymisierten oder pseudonymisierten Vertragsverhältnisses.³³³ Vielmehr kann im Verhältnis Anbieter – Nutzer eine Abfrage der relevanten Daten erfolgen.³³⁴

Für Big-Data-Anwendungen werden insbesondere neue Möglichkeiten der De-Personalisierung durch Technik zu prüfen sein, um so die Möglichkeiten des Datenumgangs zu erhöhen. In Betracht kommt hier die Verschlüsselung von Daten, die normalerweise dem Bereich der Datensicherheit zugeordnet wird. Es stellt sich aber die Frage, ob bei der Nutzung entsprechender Verschlüsselungstechniken (Encryption) und der anschließenden Versendung möglicherweise überhaupt keine personenbezogenen Daten mehr transportiert werden, da § 3 Absatz 1 BDSG personenbezogene Daten als Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person definiert.³³⁵ Diese Informationen werden aber bei einem verschlüsselten Datensatz nicht transportiert. Zwar stellen die verschlüsselten Daten für die den „Schlüssel“ besitzende Stelle weiterhin personenbezogene Daten dar.³³⁶ Jedoch hat die Verschlüsselung zur Folge, dass der Personenbezug für alle Nicht-Schlüsselhaber ausgeschlossen wird, da die inhaltliche Kenntnisnahme der Daten nur mittels einer aufwendigen, unbefugten Entschlüsselung möglich ist. Die Daten sind also für Personen, denen der Code nicht zugänglich ist, keine personenbezogenen, sondern anonyme Daten. Ein Schutz gegen die bloße Kenntnisnahme der Daten erübrigts sich, da keine personenbezogenen Informationen vermittelt werden. Der Schlüsselcode hingegen muss für Dritte absolut unzugänglich sein. Wird der Code nachträglich doch preisgegeben, werden die vorher anonymen Daten zu personenbezogenen Daten.³³⁷ Entscheidend ist, dass die Verschlüsselung nach dem aktuellen Stand der Technik als sicher eingestuft werden kann und eine Gefahr der

³³⁰ Zu den Gefahren der Re-Identification vgl. <http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/> (zuletzt abgerufen am 6.9.2013).

³³¹ Weichert,, ZD 2013, 251 (258).

³³² Weichert,, ZD 2013, 251 (258).

³³³ Müller-Broich, § 13 Rn. 10.

³³⁴ OLG Hamburg, ZUM 2009, 417 (429); Müller-Broich, § 13 Rn. 10.

³³⁵ Vgl. Diskussion, Spies, MMR-Aktuell 2011, 313727.

³³⁶ Plath/Schreiber, in: Plath, § 3 Rn. 63 f.

³³⁷ Dammann, in: Simitis, § 3 Rn. 35.

Decodierung durch unbefugte Dritte nicht besteht.³³⁸ Nicht ausreichend ist Verschlüsselung mit einem bekannten Code, wie zum Beispiel dem Morse-Code oder dem ASCII-Code.³³⁹ Ob die notwendigen Sicherheitsstandards eingehalten wurden, ist im Zweifel von dem Verschlüsselnden nachzuweisen. Das BDSG ist somit grundsätzlich bei der Übermittlung von verschlüsselten Daten anwendbar, es sei denn der Datenexporteur weist nach, dass die Daten sicher verschlüsselt sind und dass nur der Betroffene oder ein nach dem BDSG Berechtigter die Daten mit dem Schlüssel einsehen kann.³⁴⁰

Es könnte sich hierdurch eine besondere Perspektive zur Erhöhung der Einsatzmöglichkeiten einmal gewonnener Daten im Rahmen von Big-Data-Anwendungen bieten.

6. Auftragsdatenverarbeitung und Cloud Computing

Die Auftragsdatenverarbeitung hat als „zentraler Anwendungsfall bei der Nutzung neuer Formen der Datenverarbeitung“³⁴¹ auch für das Big-Data-Management eine überragende Bedeutung. So stellt etwa das Cloud Computing einen typischen Fall der Auftragsdatenverarbeitung dar. Diese hat in § 11 BDSG eine detaillierte, allerdings als nicht mehr zeitgemäß kritisierte Ausgestaltung erfahren.³⁴²

Erfolgt die Erhebung, Verarbeitung oder Nutzung personenbezogener Daten im Auftrag durch andere Stellen (Auftragnehmer), ist der Auftraggeber für die Einhaltung der Vorschriften des BDSG und anderer Vorschriften über den Datenschutz verantwortlich. Er muss den Auftragnehmer sorgfältig und unter Berücksichtigung der Eignung der von ihm getroffenen technischen und organisatorischen Maßnahmen auswählen und ihm einen schriftlichen Auftrag erteilen. Zudem muss eine regelmäßige Kontrolle des Auftragnehmers durch den Auftraggeber stattfinden, die zu dokumentieren ist. Ein Umgang mit den jeweiligen Daten durch den Auftragnehmer darf nur im Rahmen der Weisungen des Auftraggebers erfolgen.

Abzugrenzen ist die Auftragsdatenverarbeitung von einer Funktionsübertragung, was in der Praxis oft schwer fällt. Grundsätzlich gilt zur Abgrenzung: Wird eine Funktion einem Dritten zur eigenverantwortlichen Wahrnehmung ohne Weisungsgebundenheit übertragen, wird dieser Dritte zur verantwortlichen Stelle im Sinne des BDSG und es ist von einer Funktionsübertragung auszugehen. Hingegen greift § 11 BDSG dann ein, wenn die Datenverarbeitung durch den Dritten weisungsgebunden erfolgt.

Im Zusammenhang mit Big Data stellen sich daher besondere Herausforderungen an die konkrete (vertragliche) Ausgestaltung: Die Herrschaft des Kunden über die beim Auftragsdatenverarbeiter befindlichen Daten muss sich in der technischen und rechtlichen

³³⁸ Schultze-Melling, in: Taeger/Gabel, § 9 Rn. 83.

³³⁹ Dammann, in: Simitis, § 3 Rn. 33.

³⁴⁰ Spies, MMR-Aktuell 2011, 313727.

³⁴¹ AG Rechtsrahmen des Cloud Computing, Trusted Cloud - Datenschutzrechtliche Lösungen für Cloud Computing, S. 4.

³⁴² AG Rechtsrahmen des Cloud Computing, Trusted Cloud - Datenschutzrechtliche Lösungen für Cloud Computing, S. 4; Plath, in: Plath, § 11 Rn. 48.

Gestaltung der Datenverarbeitung widerspiegeln.³⁴³ Die Ausgestaltung klarer und den Vorgaben der §§ 9, 11 BDSG entsprechender vertraglicher Vereinbarungen gestaltet sich dabei schwierig.³⁴⁴ Zunächst hat der Auftraggeber die Einhaltung geeigneter technischer und organisatorischer Maßnahmen zu überprüfen. Der Auftraggeber muss sich vor Beginn der Datenverarbeitung und sodann regelmäßig von der Einhaltung der beim Auftragnehmer getroffenen technischen und organisatorischen Maßnahmen überzeugen, § 11 Abs. 2 S. 4 BDSG. Dabei ist ungeklärt, ob und inwieweit dies durch Ortskontrollen geschehen muss.³⁴⁵ Nach einer Ansicht soll in der Regel vor Ort kontrolliert werden, denn nur so könne ein ausreichender Datenschutz gewährleistet werden.³⁴⁶ Nach anderer Ansicht seien dem Auftraggeber Ortskontrollen nicht zwingend vorgeschrieben.³⁴⁷ Aufgrund der unsicheren Rechtslage³⁴⁸ besteht großes Interesse für eine Herausarbeitung klarer Ansätze zur Bestimmung der Notwendigkeit von Ortskontrollen und somit auch Potential für einen weiteren Forschungsschwerpunkt in der Auftragsdatenverarbeitung.

Die angedeuteten Probleme, die sich im Rahmen des § 11 BDSG im Hinblick auf das Cloud Computing ergeben, wurden insbesondere von der „AG Rechtsrahmen des Cloud Computing“ detailliert herausgearbeitet. Darüber hinaus wurden Lösungsansätze für eine zeitgemäße Ausgestaltung des § 11 BDSG entwickelt. Dieser innovative Ansatz soll in der Folge dargestellt werden.³⁴⁹

Reformbedarf wird vor allem beim Schriftlichkeitserfordernis des § 11 Abs. 2 S. 2 BDSG gesehen. Eine hinreichende Begründung für ein solches Schriftlichkeitserfordernis ist hinsichtlich der alternativ vorstellbaren Möglichkeiten zur Dokumentation eines Vertrags nicht ersichtlich. Bei standardisierten Cloud-Computing-Diensten könnte ein vom Auftragnehmer vorformuliertes Formular, das den Auftraggeber zur Angabe seiner spezifischen Daten auffordert, dem Erfordernis der Vertragsdokumentation in Textform genügen. Die Möglichkeit der Vertragsschließung durch derartige „Webformulare“ müsste jedoch gesetzlich ermöglicht werden. Sollte es sich um komplexere Dienste handeln, bleibt ein Individualvertrag nach wie vor unausweichlich.

Zudem entspricht die in § 11 Abs. 2 S. 4 BDSG festgelegte regelmäßige Kontrolle des Auftragnehmers durch den Auftraggeber nicht mehr den modernen Formen der Datenverarbeitung. Der zu einer solchen Kontrolle notwendige Besuch zahlreicher Server- und Rechenzentren würde für die Auftraggeber einen enormen Kostenaufwand bedeuten. Ferner kann v.a. von Auftraggebern aus kleineren Unternehmen nicht das fachliche Verständnis erwartet werden, das zur angemessenen Beurteilung der vom Auftragnehmer

³⁴³ Nägele/Jacobs, ZUM 2010, 281 (290).

³⁴⁴ Grünwald/Döpkens, MMR 2011, 287 (287); Nägele/Jacobs, ZUM 2010, 281, (284).

³⁴⁵ Hallermann, RDV 2012, 226 (226).

³⁴⁶ Bergmann/Möhrle, in: Bergmann/Möhrle/Herb, § 11 Rn. 48a; Hallermann, RDV 2012, 226 (226 m.w.N).

³⁴⁷ Gola/Schomerus, § 11 Rn. 21; Hallermann, RDV 2012, 226 (226 m.w.N); Plath, in: Plath, § 11 Rn. 113.

³⁴⁸ Hallermann, RDV 2012, 226 (226).

³⁴⁹ Vgl. hierzu: AG Rechtsrahmen des Cloud Computing, Trusted Cloud - Datenschutzrechtliche Lösungen für Cloud Computing.

getroffenen Vorkehrungen notwendig wäre. Die gesetzlichen Vorschriften zur Kontrolle durch den Auftraggeber bedürfen deshalb der Reformierung.

Dies könnte dadurch erfolgen, dass die regelmäßige Kontrolle nicht mehr durch den Auftraggeber selbst, sondern durch einen unabhängigen Dritten stattfindet, der dann ein Testat über die Kontrolle im gesetzlich festgelegten Umfang verfasst. Gegenstand dieser Testate wäre die Überprüfung der Maßnahmen des Auftragnehmers in Bezug auf den Datenschutz. Hierfür wären jedoch genaue gesetzliche Vorschriften erforderlich. Insbesondere sind Bestimmungen notwendig, die Kriterien für die Ausarbeitung und Kontrolle dieser Testate formulieren, anhand derer die Diensteanbieter kontrolliert werden. In Anbetracht der Tatsache, dass mit der zunehmenden Entwicklung auf dem Gebiet der Datenverarbeitung voraussichtlich immer mehr Kontrollen für immer mehr Dienste durchgeführt werden müssen, stellt sich die Frage, welche Stellen diese Testate erheben dürfen. Hier wird angeregt, neben der Beauftragung staatlicher Stellen die gesetzliche Grundlage zu schaffen, auch Private mit der Erstellung dieser Testate beauftragen zu können. Diese privaten Stellen sollen ihre Eignung zur Erstellung der Testate im Rahmen einer Akkreditierung feststellen lassen. Die Akkreditierung dieser Stellen kann durch fachlich geeignete, unabhängige Stellen erfolgen und gemeinschaftsweit gelten. Zur Gewährleistung einer hohen Qualität der Testate könnte eine zivilrechtliche Haftung für fehlerhafte Testate ein geeignetes Mittel darstellen.

7. US Patriot Act

Eingriffe in das informationelle Selbstbestimmungsrecht (Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG) bzw. das Recht auf Schutz personenbezogener Daten (Art. 16 AEUV, Art. 8 GRC) drohen beim Cloud Computing auch durch den Datenzugriff staatlicher Stellen. Sowohl die USA als auch europäische Staaten kennen weitreichende Zugriffsbefugnisse, die meist im Bereich des Sicherheitsrechts und der Strafverfolgung verwurzelt sind.³⁵⁰ Bemerkenswert ist, dass amerikanische Zugriffsbefugnisse nicht zwingend an eine Datenverarbeitung in den USA anknüpfen, sondern nach US-amerikanischer Auffassung bereits dann eingreifen, wenn der infrage stehende Anbieter geschäftlich auch in den USA tätig ist.³⁵¹ In der Folge ergeben sich vielfältige Konflikte zwischen ausländischen (insbesondere amerikanischen) Zugriffsbefugnissen und europäischem bzw. nationalem Datenschutzrecht. Viele Anbieter entscheiden sich im Ergebnis für die Herausgabe der Daten, da sie sich rechtlich dazu verpflichtet fühlen. So hat z.B. der US-amerikanische Anbieter Microsoft bestätigt, dass Daten aus der Nutzung des Cloud-basierten Services Office 365 an US-Behörden

³⁵⁰ Hansen, DuD 2012, 407 (410).

³⁵¹ Hogan Lovells, A Global Reality: Governmental Access to Data in the Cloud. A comparative analysis of ten international jurisdictions, 23. Mai 2012, S. 5 f., [http://www.hldataprotection.com/uploads/file/Revised%20Government%20Access%20to%20Cloud%20Data%20Paper%20\(18%20July%202012\).pdf](http://www.hldataprotection.com/uploads/file/Revised%20Government%20Access%20to%20Cloud%20Data%20Paper%20(18%20July%202012).pdf) (zuletzt abgerufen am 6.9.2013).

weitergegeben werden und dabei indirekt zugestanden, dass nicht immer eine (auch nur nachträgliche) Information des Betroffenen stattfindet.³⁵²

Eine nahezu symbolhafte Stellung für Zugriffsbefugnisse der US-Behörden hat das Regelungswerk des USA Patriot Act³⁵³ aus dem Jahr 2001 erlangt. Als Reaktion auf die Terroranschläge vom 11. September 2001 wurde durch das Änderungsgesetz die Rechtslage derart modifiziert, dass US-Behörden weitreichende Zugriffsbefugnisse eingeräumt wurden, um deren Wehrhaftigkeit und Flexibilität im Kampf gegen den Terrorismus zu erhöhen. In der Folge wurden mehrfach Bestimmungen angepasst und in ihrer Laufzeit verlängert.

Relevant für den Zugriff auf Daten in der Cloud sind vor allem Änderungen, die im sog. Foreign Intelligence Surveillance Act (FISA) und dem Electronic Communications Privacy Act (ECPA) vorgenommen wurden.

a) *Foreign Intelligence Surveillance Act*

Auf der Rechtsgrundlage des FISA können nunmehr Unterlagen von jeder beliebigen Stelle herausverlangt werden, sofern sie nur in Verbindung mit einer Untersuchung von Terrorismus und Spionage stehen.³⁵⁴ Einschränkungen der Befugnisse werden insoweit durch den ersten Zusatzartikel der US-Verfassung (First Amendment) gezogen, der die Religions-, Meinungs-, Presse- und Versammlungsfreiheit schützt. Die Vorschriften erlauben auch die Herausgabe sehr sensibler Daten, da keine Einschränkung der Art der Daten stattfindet.³⁵⁵

b) *Stored Communications Act und Electronic Communications Privacy Act*

Zusammen mit dem Stored Communications Act (SCA) hält der ECPA Verfahren für das Abhören und die Datenerhebung sowie Datennutzung im Bereich der elektronischen Kommunikation bereit und setzt dieser gleichzeitig Grenzen. Die in der Cloud gespeicherten Daten werden dabei im Regelfall dem Schutzregime des SCA und ECPA unterfallen.³⁵⁶ In Grenzen wird es Polizei und Justizbehörden nichtsdestotrotz ermöglicht, auf gespeicherte Kommunikationsvorgänge zuzugreifen. Vorausgesetzt wird ein begründeter Verdacht, dass die angeforderten Daten für eine andauernde Ermittlung relevant und maßgeblich sind.³⁵⁷

³⁵² Schmerer/Whittaker, Cloud-Computing: US-Behörden dürfen auf Daten europäischer Server zugreifen, ZDNet vom 01.07.2011, <http://www.zdnet.de/41554621> (zuletzt abgerufen am 6.9.2013).

³⁵³ Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act.

³⁵⁴ Becker/Nikolaeva, CR 2012, 170 (171).

³⁵⁵ Becker/Nikolaeva, CR 2012, 170 (171); Pallasky, DuD 2002, 221 (225).

³⁵⁶ Institute of Information Law (University of Amsterdam), Cloud Computing in Higher Education and Research Institutions and the USA Patriot Act, S. 20, http://www.ivir.nl/publications/vanhoboken/Cloud_Computing_Patriot_Act_2012.pdf (zuletzt abgerufen am 6.9.2013).

³⁵⁷ Institute of Information Law (University of Amsterdam), Cloud Computing in Higher Education and Research Institutions and the USA Patriot Act, S. 30, http://www.ivir.nl/publications/vanhoboken/Cloud_Computing_Patriot_Act_2012.pdf (zuletzt abgerufen am 6.9.2013).

Bei der Auslegung der Zugriffsbefugnisse wäre auch die Schutzbestimmung des vierten Zusatzartikels der US-Verfassung (Fourth Amendment; Schutz der Privatsphäre) zu beachten. Jedoch gilt dieser nur für US-Bürger oder ihnen gleich gestellte Personen, so dass sich europäische Cloud-Nutzer regelmäßig nicht darauf berufen können.³⁵⁸

c) National Security Letters

Eine erhöhte Flexibilität wurde durch gelockerte Voraussetzungen für den Erlass von sog. National Security Letters erreicht. Diese erlauben den Justizbehörden, ohne Einschaltung eines Gerichts Anordnungen zu erlassen, mit denen u.a. von Providern bestimmte Arten von Informationen wie Name, Adresse, Dauer der Inanspruchnahme eines Dienstes und Abrechnungsdaten herausverlangt werden können.³⁵⁹ Voraussetzung ist lediglich, dass Daten für die Untersuchung zu Zwecken nationaler Sicherheit benötigt werden.³⁶⁰ Ein unmittelbarer Bezug zwischen dem Betroffenen und einer terroristischen Organisation oder einer fremden Staatsmacht muss in diesem Stadium nicht nachgewiesen werden.³⁶¹

d) Ausblick

Losgelöst von der reinen Möglichkeit der Kenntnisnahme ist zu fragen, wie oft und in welchem Umfang von diesen Möglichkeiten des Zugriffs tatsächlich Gebrauch gemacht wird. Verlässliche Aussagen sind dabei kaum zu treffen, da es in den USA keine Anzeigepflicht hinsichtlich Datenabfragen in Bezug auf Ausländer gibt.³⁶²

Zu klären ist auch das Verhältnis der europäischen bzw. nationalen Datenschutzbestimmungen zu den ausländischen Zugriffsbefugnissen. Hier bestünde – etwa im Rahmen des Erlasses der DS-GVO – die Möglichkeit, für mehr Rechtssicherheit zu sorgen.

Hinsichtlich der aufgezeigten Unwägbarkeiten wären die Zugriffsmöglichkeiten geeignet, das Vertrauen in die Nutzung von Cloud Services nachhaltig zu schwächen, sobald ein Bewusstsein für die Gefahrenlage in der Öffentlichkeit geschaffen ist. Auf die Risiken, die aus

³⁵⁸ *Institute of Information Law* (University of Amsterdam), Cloud Computing in Higher Education and Research Institutions and the USA Patriot Act, S. 21, http://www.ivir.nl/publications/vanhoboken/Cloud_Computing_Patriot_Act_2012.pdf (zuletzt abgerufen am 6.9.2013).

³⁵⁹ Becker/Nikolaeva, CR 2012, 170 (171); Doyle, National Security Letters in Foreign Intelligence Investigations: Legal Background and Recent Amendments, CRS Report for Congress, September 8, 2009, <http://www.fas.org/sgp/crs/intel/RL33320.pdf> (zuletzt abgerufen am 6.9.2013); Henning/Bazan/Doyle/Liu, Government Collection of Private Information: Background and Issues Related to the USA Patriot Act Reauthorization, CRS Report March 2, 2010, S. 10f., <http://fpc.state.gov/documents/organization/139232.pdf> (zuletzt abgerufen am 6.9.2013).

³⁶⁰ Becker/Nikolaeva, CR 2012, 170 (171).

³⁶¹ Becker/Nikolaeva, CR 2012, 170 (171); Henning/Bazan/Doyle/Liu, Government Collection of Private Information: Background and Issues Related to the USA Patriot Act Reauthorization, CRS Report March 2, 2010, S. 15, <http://fpc.state.gov/documents/organization/139232.pdf> (zuletzt abgerufen am 6.9.2013).

³⁶² *Institute of Information Law* (University of Amsterdam), Cloud Computing in Higher Education and Research Institutions and the USA Patriot Act, S. 30, http://www.ivir.nl/publications/vanhoboken/Cloud_Computing_Patriot_Act_2012.pdf (zuletzt abgerufen am 6.9.2013).

einer umfangreichen Überwachung der Kommunikation und persönlicher Daten entstehen können, hat das *BVerfG* bereits mehrfach hingewiesen.³⁶³

8. Datentransfer/Datenaustausch

a) Europäisches Inland

Soweit im Rahmen einer Big-Data-Anwendung die Übermittlung personenbezogener Daten an Stellen in anderen Mitgliedstaaten der Europäischen Union (§ 4b Abs. 1 Nr. 1 BDSG), in anderen Vertragsstaaten des Abkommens über den Europäischen Wirtschaftsraum (§ 4b Abs. 1 Nr. 2 BDSG) oder der Organe und Einrichtungen der Europäischen Gemeinschaften (§ 4b Abs. 1 Nr. 3 BDSG) erfolgt, gilt deutsches Datenschutzrecht. Die Zulässigkeit richtet sich somit nach dem BDSG und u.U. anderen anwendbaren bereichsspezifischen Vorschriften.³⁶⁴

b) Drittstaaten

Bei einer Datenverarbeitung außerhalb der EU bzw. des EWR im Rahmen einer Big-Data-Anwendung greift der Grundsatz des angemessenen Datenschutzniveaus des Empfängerlandes und die damit verbundene Regelung der §§ 4b und c BDSG.³⁶⁵ Hiernach muss zusätzlich sichergestellt werden, dass die EU-/EWR-weiten Datenschutzstandards eingehalten werden, zum Beispiel indem sich der Auftragnehmer Binding Corporate Rules im Sinne von Art. 25 Absatz 1, 2 bzw. Art 26 Absatz 2 RL 95/46/EG, Standard-Vertragsklauseln der Europäischen Kommission im Sinne von Art. 26 Absatz 2,4 RL 95/46/EG oder Safe-Harbor-Regelungen im Sinne von Art. 25 Absatz 1,2 RL 95/46/EG unterwirft.³⁶⁶ Die Übermittlung in Drittstaaten ist also zulässig, wenn dort ein angemessenes Datenschutzniveau besteht. Ob ein solches Niveau vorliegt, muss der Datenexporteur nach Maßgabe des BDSG selbst prüfen.³⁶⁷ Wird der von der EU abgesegnete Standardvertrag zur Auftragsdatenverarbeitung (Standard Contract Clauses for Data Processing) verwendet, führt dies zu einer weitgehenden Gleichstellung des Datentransfers in den Drittstaat mit einem Transfer innerhalb der EU bzw. des EWR. Soweit jedoch der Standardvertrag verwendet wird, fordern deutsche Datenschutzaufsichtsbehörden eine Ergänzung dieses Vertrages um die Vorgaben aus dem deutschen § 11 Absatz 2 BDSG. Diese Veränderung des Standardvertrages hat jedoch aus Sicht der EU den Verlust der Gleichstellungswirkung zur Folge. Im Ergebnis muss die Vertragsgestaltung derart erfolgen, dass die Vorgaben der deutschen Datenschutzaufsichtsbehörden zusätzlich in den Vertrag aufgenommen werden, ohne dabei die Regelungen des Standardvertrages zu beschränken.

Nach Art. 25 Abs. 6 EU-DatSchRL kann die Kommission im Rahmen des in Art. 31 Abs. 2 EU-DatSchRL geregelten Verfahrens für die Mitgliedstaaten verbindlich feststellen, dass ein

³⁶³ BVerfGE 120, 274 (335 ff.); 109, 279 (314).

³⁶⁴ Simitis, in: Simitis, § 4b Rn. 36; v.d. Busche, in: Plath, § 4b Rn. 13.

³⁶⁵ Söbbing, MMR 2008, Nr. 5, XII, XIV.

³⁶⁶ Nägele/Jacobs, ZUM 2010, 281 (290).

³⁶⁷ Tinnefeld, NJW 2001, 3078 (3082); v.d. Bussche, in: Plath, § 4b Rn. 25.

Drittland aufgrund seiner innerstaatlichen Rechtsvorschriften oder internationaler Verpflichtungen ein angemessenes Datenschutzniveau aufweist.³⁶⁸

Eine spezielle Entscheidung nach Art. 25 Abs. 6 EU-DatSchRL hat die Kommission bezüglich Datenübermittlungen in die USA getroffen. Danach gewährleistet das mit dem US-Handelsministerium ausgehandelte „Safe Harbor“-Paket einen angemessenen Datenschutz. Voraussetzung ist, dass der Empfänger in den USA sich durch Erklärung gegenüber der zuständigen US-Behörde zur Einhaltung bestimmter Datenschutzprinzipien verpflichtet.³⁶⁹ Bei einer Übermittlung von Daten in die USA ist allerdings auch zu berücksichtigen, dass auf Grundlage des U.S. Patriot Act die Möglichkeit besteht, bestimmten Belangen von nationalem Interesse Priorität gegenüber Datenschutzerwägungen einzuräumen. Ähnliche Bedenken bestehen auch bezüglich des Cybersecurity Act 2009, des US-Health Insurance Portability Act, des Accountability Act sowie des Sarbanes-Oxley Act.³⁷⁰

Ein besonderes Forschungsinteresse besteht bei Big-Data-Anwendungen also hinsichtlich der Frage, wann ein Datenumgang im Ausland ohne größeres Haftungsrisiko erfolgen kann. Für die USA besteht keine umfassende Angemessenheitsentscheidung und die Safe Harbor Principles reichen als bloß formale Zusicherung nicht mehr aus – vielmehr wird ein tatsächlicher Nachweis vom Datenexporteur verlangt.³⁷¹ Hinsichtlich der Daten mit steuerrechtlichem Bezug erscheint es bei Ländern ohne angemessenes Datenschutzniveau insbesondere fraglich, ob eine Genehmigung des Finanzamtes zur Verlagerung überhaupt erteilt wird.

9. Rechte der Betroffenen im nicht-öffentlichen Bereich

Die Betroffenen einer Big-Data-Anwendung haben nach dem BDSG auch gegenüber nicht-öffentlichen Stellen einige Rechte, die sich in den §§ 33 ff. finden. Hervorzuheben ist zunächst die Benachrichtigungspflicht gem. § 33 BDSG.

Sobald erstmals personenbezogene Daten ohne Kenntnis des Betroffenen im Rahmen von Big Data gespeichert werden, muss die verantwortliche Stelle von sich aus aktiv werden und den Betroffenen von der Speicherung, der Art der Daten, der Zweckbestimmung der Erhebung, Verarbeitung oder Nutzung und der Identität der verantwortlichen Stelle in Kenntnis setzen, § 33 Abs. 1 BDSG. Werden personenbezogene Daten geschäftsmäßig zum Zweck der Übermittlung ohne Kenntnis des Betroffenen gespeichert, ist der Betroffene von der erstmaligen Übermittlung und der Art der übermittelten Daten zu benachrichtigen.

³⁶⁸ *Gola/Schomerus*, § 4b Rn. 14; v. d. *Bussche*, in: *Plath*, § 4b Rn. 29.

³⁶⁹ *Gola/Schomerus*, § 4b Rn. 15; v. d. *Bussche*, in: *Plath*, § 4b Rn. 30.

³⁷⁰ *Nägele/Jacobs*, ZUM 2010, 281 (290); *Spies*, MMR 2009, Nr. 5, XI, XII.

³⁷¹ *Düsseldorfer Kreis*, Beschluss der obersten Aufsichtsbehörden für den Datenschutz im nicht-öffentlichen Bereich vom 28./29.4.2010; v. d. *Bussche*, in: *Plath*, § 4b Rn. 32.

Auch für nicht-öffentliche Stellen finden sich Ausnahmen von der Benachrichtigungspflicht, § 33 Abs. 2 BDSG. Die zahlreichen Ausnahmen lassen in der Praxis die Benachrichtigungspflicht nur selten zur Anwendung kommen.³⁷²

Für Big-Data-Anwendungen von Interesse ist insbesondere die Ausnahme gem. § 33 Abs. 2 Nr. 7 BDSG, nach der die Benachrichtigung unterbleiben kann, wenn sie die Geschäftszwecke der verantwortlichen Stelle erheblich gefährden würde, es sei denn, dass das Interesse an der Benachrichtigung die Gefährdung überwiegt. Erhebliche Gefährdung meint hier, dass eine Benachrichtigung zur Unmöglichkeit der generellen Art der Geschäfte der verantwortlichen Stelle führt.³⁷³ Es wird für Big-Data-Anwendungen genau zu untersuchen sein, inwieweit das Geschäftsmodell Big Data im Einzelnen durch eine solche Pflicht gefährdet würde.

Im Bereich der nicht-öffentlichen Stellen ist im Hinblick auf § 33 Abs. 2 Nr. 7 – 8 BDSG ebenso wie bei öffentlichen Stellen von besonderem Interesse, wann eine Unverhältnismäßigkeit aufgrund der Vielzahl der betroffenen Fälle vorliegt. Es ist gerade ein Kennzeichen von Big-Data-Anwendungen, dass eine große Zahl von Fällen betroffen ist – hier gilt es Maßstäbe dafür zu entwickeln, wann diese Zahl so groß ist, dass eine Benachrichtigung von Betroffenen unterbleiben kann.

Eng damit verbunden ist die Ausnahme für Fälle, in denen die Daten für eigene Zwecke gespeichert und aus allgemein zugänglichen Quellen entnommen wurden und eine Benachrichtigung wegen der Vielzahl der betroffenen Fälle unverhältnismäßig ist. Hier stellt sich ähnlich wie bei öffentlichen Stellen die Frage, wann bei Big-Data-Anwendungen ein Fall der Unverhältnismäßigkeit gegeben ist.

10. Rechtsfolgen und Haftung

Verstöße gegen die Bestimmungen des BDSG in Zusammenhang mit Big-Data-Management können für Unternehmen massive Konsequenzen zivil- und strafrechtlicher Natur haben.

In zivilrechtlicher Hinsicht ist insbesondere mit Ansprüchen auf Unterlassung, Löschung und Schadensersatz zu rechnen. Diese können sich aus dem BGB (§ 823 Abs. 1; § 823 Abs. 2 i.V.m. einem Schutzgesetz) oder unmittelbar aus dem BDSG selbst ergeben. § 7 BDSG sieht bei Verstößen gegen das BDSG eine verschuldensabhängige Schadensersatzpflicht nicht-öffentlicher Stellen vor, soweit diese die nach den Umständen des Falles gebotene Sorgfalt nicht beachtet haben. Öffentliche Stellen haften verschuldensunabhängig auf Schadensersatz und auf Schmerzensgeld, § 8 BDSG.

Darüber hinaus enthält das BDSG eigene Bußgeld- und Strafvorschriften. § 43 BDSG zählt mit Bußgeld bewährte Ordnungswidrigkeitstatbestände auf, die verschiedene Verstöße gegen Normen des BDSG sanktionieren. Wird eine in § 43 Abs. 2 BDSG aufgeführte vorsätzliche Handlung gegen Entgelt oder in der Absicht, sich oder einen anderen zu bereichern oder

³⁷² Gola/Schomerus, § 33 Rn. 27; Kamlah, in: Plath, § 33 Rn. 22.

³⁷³ Gola/Schomerus, § 33 Rn. 39; Kamlah, in Plath: § 33 Rn. 52.

einen anderen zu schädigen, begangen, so kann eine Freiheitsstrafe bis zu zwei Jahren oder eine Geldstrafe verhängt werden, § 44 BDSG. Daneben kommt – insbesondere in Bezug auf den Arbeitnehmerdatenschutz – stets eine strafrechtliche Verantwortlichkeit nach dem StGB in Betracht (§§ 202, 202a, 206, 303a StGB).

Diese möglichen Rechtsfolgen machen bereits deutlich, dass der bewusste Umgang eines Unternehmens mit personenbezogenen Daten unerlässlich ist, um eine Strategie in diesem Bereich dauerhaft, effektiv und somit auch gewinnbringend umzusetzen.

11. Produktbezogene informationelle Selbstbestimmung – Privacy by Design

Privacy by Design (PbD) wird definiert „als Methode zur Verringerung des für die informationelle Selbstbestimmung in einem Datenverarbeitungssystem immanent liegenden Gefahrenpotentials durch das Zusammenwirken proaktiver Technikgestaltung und organisatorischer Vorkehrungen.“³⁷⁴ Dabei sollen Datenschutzstandards bereits im Stadium der Entwicklung einer Anwendung bzw. eines Produkts berücksichtigt werden.³⁷⁵ Ein Beispiel dafür ist die Einstellung eines privaten Modus‘ als Standard bei Web-Browsern, sog. „privacy by default“.³⁷⁶ Der Datenschutz soll nach diesem Konzept also „proaktiv, nicht reaktiv, als Vorbeugung, nicht als Abhilfe“³⁷⁷ eingesetzt werden.

Bisher unklar ist, wer idealerweise Adressat eines „Privacy by Design“-Prinzips sein sollte: Hersteller und Entwickler von Anwendungen und Produkten oder, wie bisher, primär die für die Datenverarbeitung Verantwortlichen.³⁷⁸ In der nunmehr vorgeschlagenen Datenschutz-Grundverordnung werden die für die Datenverarbeitung Verantwortlichen verpflichtet, Privacy by Design zu berücksichtigen.³⁷⁹

Eine konsequente Umsetzung des Konzeptes Privacy by Design begegnet jedoch einigen Problemen.³⁸⁰ Würde man tatsächlich auch die Hersteller einer Anwendung zur Beachtung von Privacy by Design verpflichten, wäre eine solche Regelung möglicherweise eher als Produkthaftungs- denn als Datenschutzregelung zu strukturieren.³⁸¹ Von besonderer Bedeutung ist hierbei die Wirtschaftlichkeit der Entwicklung, denn ohne diesen Anreiz wird es kein Fortkommen auf diesem Gebiet geben.³⁸²

Mit Hilfe zukünftiger Forschung wird für Big Data herauszuarbeiten sein, in welcher Balance sich Datenschutzinteresse und Wirtschaftsinteresse begegnen: Bis zu welchem Punkt ist die Entwicklung einer neuen Anwendung / eines neuen Produkts noch hinnehmbar und ab wann

³⁷⁴ Schulz, CR 2012, 204 (204); vgl. auch Definition der EU Kommission: KOM (2010) 245 final/2, 17.

³⁷⁵ Schneider, AnwBI 2011, 233 (239); Weichert, ZD 2013, 251 (256).

³⁷⁶ ZD-Aktuell 2012, 03008.

³⁷⁷ Schulz, CR 2012, 204 (205).

³⁷⁸ Schulz, CR 2012, 204 (207).

³⁷⁹ Schulz, CR 2012, 204 (207).

³⁸⁰ Schulz, CR 2012, 204 (206).

³⁸¹ Schulz, CR 2012, 204 (207).

³⁸² Schulz, CR 2012, 204 (207).

ist eine Gefährdung der informationellen Selbstbestimmung möglicher Betroffener derart wahrscheinlich, dass von einer Entwicklung abgesehen werden muss?³⁸³ Ferner ist zu klären, ob die bloße Bereitstellung von Wahlmöglichkeiten zur Aktivierung bzw. Deaktivierung von Anwendungsmöglichkeiten ausreicht, oder ob von vornherein eine datenschutzfreundliche Programmierung vorzuschreiben ist.³⁸⁴

12. Profiling – Persönlichkeitsprofile

Big-Data-Anwendungen ermöglichen aufgrund großer Datenmengen und der Verknüpfung verschiedener Datenquellen besonders gut die Erstellung von Persönlichkeitsprofilen. Frei verfügbare Informationen bilden ein „verselbstständigtes Informationskapital“³⁸⁵. Hierin liegt die große Chance, aber auch die große Gefahr von Big Data.

Die Vielzahl von im Internet veröffentlichten Daten lässt sich oftmals ohne Wissen der Betroffenen verarbeiten und zu Profilen zusammenführen. Damit einher geht das Risiko falscher und irreführender Information sowie unerwarteter und ungewollter Nutzung.³⁸⁶ Dieses besondere Risiko wurde auch von der EU-Kommission erkannt. Deshalb ist Artikel 20 DS-GVO-E³⁸⁷ auf die auf Profiling basierenden Maßnahmen gerichtet.

Profiling findet sich dabei in verschiedensten Ausprägungen: Die Algorithmen der Suchmaschine Google arbeiten z.B. mit persönlichen Präferenzen, historischem Surf-Verhalten und anderen Informationen, um so eine personalisierte Suche zu ermöglichen und folglich bei verschiedenen Nutzern völlig unterschiedliche Suchergebnisse hervorbringen zu können.³⁸⁸ Diese Algorithmen arbeiten dabei unbemerkt und für den Nutzer nicht transparent.³⁸⁹ Das automatische Profiling dient den Suchmaschinen zur Serviceverbesserung und als Unterscheidungsmerkmal; sichtbar sind diese Filterprozesse jedoch nicht.³⁹⁰ Daneben bleiben jedoch weitere Fragen offen. So ist noch immer nicht geklärt, was rechtlich unter einem Profil zu verstehen ist und wann es gebildet werden darf.³⁹¹

Werden aufgrund einer Big-Data-Anwendung prinzipiell gleichen Bürgern bei einer identischen Anfrage bzw. identischem Arbeitsauftrag völlig unterschiedliche Ergebnisse, basierend auf dem jeweiligen Persönlichkeitsprofil, ausgegeben, ist dies zumindest fragwürdig. Die Maschine greift auf diese Weise der menschlichen Entscheidung für oder gegen ein bestimmtes Suchergebnis unerkannt vorweg.³⁹²

³⁸³ Schulz, CR 2012, 204 (207).

³⁸⁴ Schulz, CR 2012, 204 (207).

³⁸⁵ Moos, MMR 2006, 718.

³⁸⁶ Leutheusser-Schnarrenberger, MMR 2012, 709 (710).

³⁸⁷ KOM (2012) 11 endg.

³⁸⁸ Holler, FAZ v. 16.04.2013, S. 29.

³⁸⁹ Holler, FAZ v. 16.04.2013, S. 29.

³⁹⁰ Holler, FAZ v. 16.04.2013, S. 29.

³⁹¹ Leutheusser-Schnarrenberger, MMR 2012, 709 (710).

³⁹² Vgl. auch Kapitel „II. J.“ zum Roboterrecht.

Ein wichtiges Kriterium im Zusammenhang mit Profiling ist daher Transparenz. Vorgeschlagen wird etwa eine Kennzeichnungspflicht für Algorithmen, so dass eine technisch gesteuerte Auswahl von Ergebnissen kenntlich wird.³⁹³

Zukünftige Forschung wird klären müssen, was genau ein Profil ist und unter welchen Voraussetzungen das Profiling stattfinden darf. Gleichzeitig sollten die genauen Bedingungen der Information und Aufklärung der Betroffenen geregelt werden,³⁹⁴ um Transparenz herzustellen, gleichen Zugang zu Informationen zu gewährleisten und demokratische Teilhabe zu garantieren.

13. Ausblick auf die Datenschutz-GrundVO

Das derzeit geltende Datenschutzrecht basiert im Wesentlichen auf der Datenschutzrichtlinie 95/46/EG und ihrer Umsetzung im BDSG. Bereits im Jahre 2010 hat die EU-Kommission aufgrund der höchst unterschiedlichen Umsetzung dieser Richtlinie in den einzelnen Mitgliedsstaaten angekündigt, den Datenschutz in der Europäischen Union auf eine neue Stufe stellen zu wollen.³⁹⁵ Eine möglichst weitgehende Europäisierung des Datenschutzrechts sei im Hinblick auf den Fortschritt der Informationsgesellschaft unumgänglich.³⁹⁶ Der Umgang mit Big Data wird zum Teil als ein aussichtsloser Kampf mit Werkzeugen des analogen Zeitalters empfunden.³⁹⁷ Insbesondere wurden die Beherrschung neuer Technologien,³⁹⁸ der Umgang mit der Globalisierung und Verbesserung internationalen Datentransfers³⁹⁹ und eine kohärentere Regelung für den Datenschutz⁴⁰⁰ als Herausforderungen für den Datenschutz benannt. Dementsprechend sollen u.a. sowohl die Rechte des Einzelnen gestärkt⁴⁰¹ als auch die globale Dimension des Datenschutzes⁴⁰² stärker in den Fokus gerückt werden. Die Europäische Kommission brachte einen ersten Vorschlag für eine Datenschutz-Grundverordnung⁴⁰³ ein. Die Verordnung soll in allen Mitgliedsländern unmittelbar gelten und so bestehende Divergenzen bei der Umsetzung beseitigen.⁴⁰⁴ Der Vorschlag für die Datenschutz-Grundverordnung hat eine breite Diskussion ausgelöst, die sich vor allem mit den Fragen beschäftigt, ob eine hinreichende Modernisierung des Datenschutzes erreicht wurde und ob die Europäische Union über hinreichende Kompetenzen zum Erlass einer solchen Verordnung verfügt.⁴⁰⁵ Die Datenschutz-Grundverordnung kodifiziert dabei zahlreiche neue Ansätze im Datenschutzrecht der letzten

³⁹³ Holler, FAZ v. 16.04.2013, S. 29.

³⁹⁴ Leutheusser-Schnarrenberger, MMR 2012, 709 (710).

³⁹⁵ KOM (2010) 609 endg.

³⁹⁶ So schon Tettenborn, MMR 1998, 18 (22).

³⁹⁷ Mayer-Schönberger/Cukier (2013), p. 156; Wagner, ZRP 2013, 33 (33).

³⁹⁸ KOM (2010) 609, 3.

³⁹⁹ KOM (2010) 609, 4.

⁴⁰⁰ KOM (2010) 609, 4.

⁴⁰¹ KOM (2010) 609, 5 ff.

⁴⁰² KOM (2010) 609, 17 ff.

⁴⁰³ KOM (2012) 11.

⁴⁰⁴ Gola, EuZW 2012, 332 (332 f.).

⁴⁰⁵ Gola, EuZW 2012, 332 (333).

Jahre, unter anderem „die Instrumente ‚privacy by default‘, ‚privacy by design‘⁴⁰⁶, ‚Recht auf Vergessen werden‘ und Datenportabilität“⁴⁰⁷. Die Beratungen über die Datenschutz-Grundverordnung sind derzeit noch nicht abgeschlossen. Eine Einigung über einen Kompromiss zur Datenschutz-Grundverordnung erfolgte jedenfalls im zuständigen Ausschuss des EU-Parlaments im Oktober 2013.

E. Vertragsrecht

Die Vertragsgestaltung im Big-Data-Business ist aufgrund der vielen technischen Fragen sowie der Berührungs punkte zum Datenschutz- und Urheberrecht kompliziert. Die Vielzahl denkbarer Big-Data-Anwendungen und ihrer vertraglichen Umsetzungsmöglichkeiten führt dazu, dass sich pauschalisierte Aussagen zur Vertragsgestaltung schwerlich treffen lassen. Die Verteilung der Rechte und Pflichten aus dem jeweiligen Vertragsverhältnis liegt nach dem Grundsatz der Vertragsfreiheit in den Händen der beteiligten Parteien. Die rechtliche Einordnung eines Vertrages und damit einhergehend vertragsrechtliche Fragen müssen stets im Einzelfall beurteilt werden.

Vertragliche Regelungen können etwa bei im Internet weitgehend frei zugänglichen Daten (z.B. in sozialen Netzwerken) eine Rolle spielen. Denn bei der Nutzung fremder Daten bzw. fremder Inhalte eines Webseitenbetreibers können vertragliche Nutzungsbeschränkungen die rechtliche Zulässigkeit oder den Umfang einer zulässigen Nutzung einschränken.⁴⁰⁸

Fragen der Berechtigung sowie weitere grundlegende Probleme der Vertragsgestaltung im Geschäftsbereich von Big-Data-Unternehmen lassen sich beispielhaft auch anhand des Cloud-Computing veranschaulichen, bei dem der Cloud-Nutzer zur Verwaltung seiner Datenbestände auf Ressourcen des Cloud-Anbieters zurückgreift.

Vertragsrechtliche Fragen im Zusammenhang mit dem Cloud-Computing, einem Hauptanwendungsfall des Big-Data-Managements, werden in der Folge exemplarisch betrachtet.

1. Hauptleistungspflichten

Geht man von der mietrechtlichen Beurteilung des Cloud-Computing Vertrages aus,⁴⁰⁹ so schuldet der Cloud-Provider Überlassung und Erhaltung der Mietsache.⁴¹⁰ Konkrete Anhaltspunkte, was das für die Hauptleistungspflichten bedeutet, ergeben sich daraus jedoch noch nicht. Deshalb müssen die Hauptleistungspflichten in solchen Verträgen technisch genau beschrieben sein. Quantitative und qualitative Vereinbarungen erfolgen im Rahmen von Service Level Agreements (SLA).⁴¹¹ Unter anderem sind die Sicherheitsvorkehrungen und das Sicherheits-Management (Firewall, Verschlüsselung, Authentifikation) in der Cloud zu bestimmen. Es bedarf auch eines Notfall-Plans für

⁴⁰⁶ Schulz, CR 2012, 204 (204).

⁴⁰⁷ Gola, EuZW 2012, 332 (334).

⁴⁰⁸ Ausführlich dazu Zieger/Smirra, MMR 2013, 418 (421).

⁴⁰⁹ Dazu Wicker, MMR 2012, 783 (788).

⁴¹⁰ Emmerich, in: Staudinger, § 535 Rn. 21.

⁴¹¹ Schumacher, MMR 2006, 12 (13).

unerwarteten Webtraffic. Die Frage des Daten-Backups sowie Eckdaten des User Supports (24h-Fähigkeit) sind ebenfalls zu regeln. Zudem sollten die SLA Sanktionen für den Fall festlegen, dass der Anbieter die geschuldeten Service Levels nicht erfüllt. Denn den für Schadensersatzansprüche zumeist notwendigen Verschuldensnachweis kann der Nutzer nur schwer erbringen.

a) Rückgriff auf das gesetzliche Leitbild

Was aber passiert, wenn einer dieser Aspekte nicht geregelt wurde, ist völlig offen. Ob die vertraglichen Lücken durch Rückgriff auf das gesetzliche Leitbild des Mietrechts sachgerecht geschlossen werden können, erscheint fraglich. Dies zeigt sich beispielsweise bei der Frage der Verfügbarkeit von Übertragungskapazitäten. Fällt die Telekommunikationsverbindung zwischen dem Kunden und seinem Cloud-Provider aus oder treten sonstige technische Störungen auf, so kann das Big-Data-Unternehmen, das auf den Datenzugriff angewiesen ist, seiner Tätigkeit nicht nachgehen. Hier drohen enorme wirtschaftliche Verluste, die dazu führen können, dass Regressansprüche gegen den Cloud-Provider geltend gemacht werden. Dabei ist entscheidend, ob Überlassung und Erhaltung der Mietsache „rund um die Uhr“ geschuldet sind, und wen das Risiko trifft, dass technische Voraussetzungen für die Vertragsdurchführung gestört sind. Konkrete Forschungsergebnisse könnten hier die notwendige Klarheit schaffen.

b) AGB-rechtliche Regelung der Hauptleistungspflichten

Um gar nicht erst Gefahr zu laufen, möglicherweise existenzgefährdenden Regressansprüchen ausgesetzt zu sein, regeln in der Praxis viele Unternehmen Haftungsausschlüsse in allgemeinen Geschäftsbedingungen. Diese sind im Gegensatz zu den privatautonom ausgehandelten Leistungspflichten, in die der Richter nicht korrigierend eingreifen darf, der vollen Inhaltskontrolle nach § 307 BGB zugänglich.⁴¹² Hier zeigen sich jedoch erhebliche rechtliche Unsicherheiten, die Unternehmen vom Einstieg in das Big-Data-Geschäft abhalten könnten. Es kann im Einzelfall vorkommen, dass die Rechtsprechung Klauseln, die in erster Linie nicht der Haftungsfreistellung dienen, aufgrund ihrer vergleichbaren Wirkung dennoch als solche bewertet und als unzulässig einstuft.⁴¹³

Ein noch größeres Problem besteht hinsichtlich der Frage, wie eine wirksame Klausel konkret ausgestaltet sein muss, um den strengen Anforderungen der Rechtsprechung gerecht zu werden. Der *BGH* hat eine wichtige Einschränkung hinsichtlich der Begrenzung von Schadensersatzansprüchen aufgestellt. So soll ein Ausschluss der Haftung im Falle leichter Fahrlässigkeit dann nach § 307 Abs. 2 BGB unzulässig sein, wenn es um die Erfüllung sog. Kardinalpflichten geht.⁴¹⁴ Dabei handelt es sich um die Erfüllung solcher Pflichten, die für die Leistung des AGB-Verwenders gerade typisch sind. Wenn ein Big-Data-Unternehmen aber eine vertragliche Beziehung zum Cloud-Provider eingeht, um Daten auf dessen Servern auszulagern, dann liegt die Kardinalpflicht des Vertragspartners in der ständigen

⁴¹² Spindler, in: Benlian/Hess/Buxmann, S. 34.

⁴¹³ *BGH*, NJW 2001, 751 (751).

⁴¹⁴ *BGH*, NJW 1985, 3016.

Bereithaltung und Zugriffsmöglichkeit. Ein Haftungsausschluss, der fahrlässige Pflichtverletzungen regelt, ist diesbezüglich nicht möglich. Vor diesem Hintergrund besteht ein nicht zu unterschätzendes Konfliktpotenzial, das durch die Erarbeitung rechtswirksamer Haftungsklauseln entschärft werden kann. Langwierige und unternehmensschädliche Rechtsstreitigkeiten könnten auf diesem Wege reduziert werden, so dass sich in diesem Bereich des Big-Data-Managements eine lohnende Forschungsperspektive eröffnet.

Gleiches gilt für den Fall der Haftungsbegrenzung durch AGB. Wenn schon ein vollständiger Haftungsausschluss Probleme bereitet, dann könnten Cloud-Provider zumindest bestrebt sein, die Haftung zu begrenzen. Denkbar sind Klauseln, die die Haftungshöhe auf vertragstypisch vorhersehbare Schäden begrenzen oder die Haftung für unvorhersehbare Schäden ausschließen.⁴¹⁵ Daneben käme die Möglichkeit in Betracht, dass die Haftung auf die Wiederherstellung ordnungsgemäß gesicherter Daten oder auf den Schaden begrenzt wird, der bei ordnungsgemäßer Datensicherung entstanden wäre. Für den Bereich der Softwareerstellungsverträge wird dies grundsätzlich als zulässig erachtet.⁴¹⁶ Ob diese Alternativen jedoch auch auf Big-Data-Geschäftsmodelle übertragen werden können, erscheint unklar. Denn der Cloud-Nutzer geht ja gerade die vertragliche Beziehung zum Provider ein, um seine großen Datenmengen überhaupt verwalten zu können. Es wäre widersprüchlich, ihm im Falle des Datenverlusts gleichsam daran ein Mitverschulden anzulasten, dass er die Daten nicht gesichert hat. Zudem stellt sich auch hier wieder die Frage, ob nicht eine Kardinalpflicht betroffen ist, bei der selbst im Falle von leichter Fahrlässigkeit eine Haftungsfreizeichnung unwirksam ist. Es bleibt also dabei, dass die Bewertung von Klauseln, die Haftungsfragen regeln, erhebliche rechtliche Schwierigkeiten aufwirft. Wissenschaftliche Erkenntnisse in diesem Bereich könnten die tatsächlich bestehenden Hemmnisse für Unternehmen, die in diesen Markt eintreten wollen, beseitigen.

c) Entgelt

Der Cloud-Nutzer schuldet die mietvertraglich vorgesehene Vergütung, typischerweise monatlich. Der BGH hat für Vermietlösungen CPU-Erhöhungsklauseln als wirksam angesehen.⁴¹⁷ Insofern kann der Cloud-Provider bei Veränderungen der technischen Infrastruktur ein erhöhtes Nutzungsentgelt verlangen. Im Gegensatz dazu besteht auch die Möglichkeit, dass sich der Mietzins verringert. Wenn Cloud-Geschäfte nämlich als Mietvertrag einzuordnen sind, obliegt dem Cloud-Provider die Pflicht zur Instandhaltung und Instandsetzung (§ 535 Abs. 1 Satz 2). Jede Einschränkung der Nutzbarkeit des Cloud-Systems führt dann grundsätzlich automatisch zu einer Minderung der Mietgebühren.

2. Nebenleistungspflichten

Durch § 241 Abs. 2 BGB sind die Vertragsparteien verpflichtet, auf die Rechtsgüter und Interessen des jeweiligen anderen Teils Rücksicht zu nehmen. Angenommen wird, dass dies

⁴¹⁵ BGH, NJW 2001, 292 (302).

⁴¹⁶ Redeker, Rd. 470.

⁴¹⁷ BGH, MMR 2003, 393.

den Cloud-Anbieter jedenfalls dazu verpflichtet, den Zugriff Dritter auf die Daten zu verhindern und die Verfügbarkeit der Daten für den Nutzer sicherzustellen.⁴¹⁸ Darüber hinaus ist vieles unklar. So ist streitig, ob der Cloud-Provider ein Verbot der Nutzung von Anlagen anderer Hersteller aussprechen kann. Zudem ist es denkbar, dass den Cloud-Nutzer die Nebenpflicht trifft, für eine bestimmte Qualität der Daten zu sorgen. Denn durch veraltete Datenformate oder virenbefallene Daten könnten die Speicherkapazitäten des Anbieters beschädigt oder zerstört werden. Insofern müsste der Cloud-Nutzer vor der Übermittlung der Daten möglicherweise eine Überprüfung seiner Daten vornehmen. Eine wissenschaftliche Klärung der Frage über die Reichweite der Nebenpflichten würde die wirtschaftliche Risikoprognose für am Markteintritt interessierte Big-Data-Unternehmen erleichtern und somit einen Anreiz für mehr Marktzutritte schaffen.

3. Nachvertragliche Pflichten

Aus Sicht des Cloud-Nutzers können nachvertragliche Pflichten eine entscheidende Bedeutung haben. Fraglich ist, wie dieser seine Daten vollständig zurückhält. Kommt es zur Beendigung des Vertrages, gewährt § 539 Abs. 2 BGB dem Mieter ein Wegnahmerecht. In konsequenter Fortführung der Anwendung des Mietrechts müsste somit dem Cloud-Nutzer bezüglich seiner eigenen Daten ebenfalls dieses Recht zustehen. Deshalb ist der Cloud-Provider zur Herausgabe sämtlicher Daten inklusive angefertigter Sicherungskopien verpflichtet. Zu beachten ist, dass dieser Wegnahmeanspruch nur einer kurzen Verjährung von sechs Monaten ab Beendigung des Vertrages (§ 548 Abs. 2 BGB) unterliegt. Inwiefern hinsichtlich dieser Regelungen klauselmäßige Modifikationen möglich sind, müsste in diesem Zusammenhang untersucht werden. Ferner stehen dem Kunden datenschutzrechtliche Löschungsansprüche zu (§ 35 Abs. 2 Nr. 3 BDSG).

4. Schadensermittlung bei Datenverlust

Kommt es zum schuldhaften Datenverlust zum Nachteil eines Big-Data-Unternehmens, stehen diesem je nach Konstellation unterschiedliche Schadensersatzansprüche zur Verfügung. Diese können sich nicht nur aus dem Vertrag wegen Verletzung einer Haupt- oder Nebenleistungspflicht ergeben, sondern unter Umständen auch aus dem Deliktsrecht oder Spezialgesetzen wie § 7 BDSG. Allerdings stellt sich die Frage, wie im konkreten Fall die Höhe des Schadens zu ermitteln ist. Das *OLG Oldenburg* hat zur Schadensberechnung die Kosten zugrunde gelegt, die für die Wiederherstellung des Datensatzes erforderlich waren.⁴¹⁹ Denn im Rahmen des Schadensersatzes ist in erster Linie Naturalrestitution geschuldet. Möglich ist dies aber nur, soweit die Dateien auf Grund einer in anderer Form noch vorhandenen Vorlage technisch reproduzierbar sind.⁴²⁰ Vor diesem Hintergrund zeigt sich die Wichtigkeit von Datensicherungsmaßnahmen. Liegen dagegen im Falle des Datenverlustes keine reproduzierbaren Daten vor, so scheidet nicht nur der Schadensersatz im Wege der Naturalrestitution aus. Auch eine Forderung der Wiederherstellungskosten gem. § 249 Abs. 2 S. 1 BGB durch den Geschädigten kommt dann nicht in Frage. Denn es

⁴¹⁸ Wicker, MMR 2012, 783 (785).

⁴¹⁹ OLG Oldenburg, B. v. 24.11.2011 - 2 U 98/11; MMR-Aktuell 2011, 326091.

⁴²⁰ BGH, MMR 2009, 250 (251).

dürfte sich in einem solchen Fall regelmäßig um qualifizierte geistige oder schöpferische Leistungen handeln, deren Neuanschaffung keine Wiederherstellung im Rechtssinne darstellen würde.⁴²¹ Demnach träfe den Schädiger die Wertersatzpflicht aus § 251 Abs. 1 BGB. Für diesen Fall hat der *BGH* festgestellt, dass die Vermögenseinbuße vom Tatrichter nach freiem Ermessen zu schätzen ist.⁴²² Dabei sollen nicht nur die konkreten Kosten, die der Geschädigte seit dem Schadensereignis für die Rekonstruktion von verlorenen Daten aufgewendet hat, als Bemessungsgrundlage heranzuziehen sein. Vielmehr ist auch von Bedeutung, inwieweit durch ihr Fehlen Betriebsabläufe gestört und erschwert werden.⁴²³ Gegenüber diesem Anspruch kann sich der Schädiger nicht auf Unverhältnismäßigkeit gem. § 251 Abs. 2 BGB berufen.⁴²⁴ Dementsprechend droht im Big-Data-Business eine hohe Haftung für die schuldhafte Herbeiführung eines Datenverlustes, so dass rechtssichere Instrumente zur Haftungsbegrenzung gefunden werden müssen.

F. Informationshaftung – Informationsqualität

Die aus den Datenmengen abgeleiteten Informationen sind Grundlage für Entscheidungen in jeglichen Lebensbereichen. Fehlerhafte Informationen können gravierende Auswirkungen haben. So lösten beispielsweise fehlerhafte Daten von Geschwindigkeitssonden eine Fehler-Kette aus, die letztlich zum Absturz eines Flugzeugs führte, bei dem 228 Insassen ihr Leben verloren.⁴²⁵ Angesichts der immer stärkeren Berücksichtigung von Daten in Entscheidungsprozessen sowie in Anbetracht der daraus entstehenden Gefahren stellt sich die Frage nach einem Haftungsregime für fehlerhafte Informationen. Zu klären ist, unter welchen Voraussetzungen eine vertragliche oder deliktische Haftung in solchen Fällen möglich ist. Dabei ist auch auf etwaige Beweiserleichterungen einzugehen. Abschließend soll untersucht werden, inwiefern eine Freizeichnung von der Haftung möglich ist.

1. Informationsqualität

Damit die Haftung eines Big-Data-Unternehmens in Betracht kommt, muss auf Seite des Unternehmens ein Fehlverhalten vorliegen. Auf den ersten Blick scheint dies unproblematisch, da ein solches durch die Übermittlung einer fehlerhaften Information gegeben sein kann. Bei genauerer Betrachtung stellt sich allerdings die Frage, wann eine Fehlinformation vorliegt. Denn es ist unklar, ob es qualitative Mindeststandards für Informationen gibt und welche das ggf. sind.

Betrachtet man nationale und internationale Rechtsquellen, so ergibt sich, dass die Normgeber bei informationsrelevanten Tätigkeiten stets die Einhaltung eines qualitativen Levels vorschreiben oder zumindest von dessen Einhaltung ausgehen. Art. 5 lit. d) des

⁴²¹ *BGH*, MMR 2009, 250 (251).

⁴²² *BGH*, MMR 2009, 250.

⁴²³ *BGH*, MMR 2009, 250 (252).

⁴²⁴ *BGH*, MMR 2009, 250 (251).

⁴²⁵ *Lehnartz*, Welt vom 05.07.2012, abrufbar unter

<http://www.welt.de/vermisches/weltgeschehen/article107913815/Wie-das-Airbus-Drama-in-elf-Kilometern-Hoehe-abrief.html>.

Datenschutzübereinkommens (DatSchÜbK) von 1985⁴²⁶ fordert beispielsweise für personenbezogene Daten, dass diese sachlich richtig sind und notfalls auf den neuesten Stand gebracht werden müssen. Auch aus § 35 Abs. 1 BDSG ergibt sich unmittelbar, dass personenbezogene Daten richtig sein müssen. Neben der Richtigkeit verweist Art. 4 lit. d) der EG-Datenschutzverordnung⁴²⁷ auf die Vollständigkeit von personenbezogenen Daten. Auch Pressevertreter als typische Verbreiter von Informationen sind verpflichtet, ihre Nachrichten sorgfältig auf Inhalt, Wahrheit und Herkunft zu überprüfen.⁴²⁸ Konsens besteht somit dahingehend, dass Daten bzw. die daraus gewonnenen Informationen korrekt im Sinne von nachweislich richtig sein müssen. Dieser allgemeine Maßstab dürfte auch praktisch keine Probleme bereiten, da notfalls durch Sachverständigengutachten eine juristische Überprüfbarkeit gewährleistet werden kann.

Schwieriger sind über diesen Mindeststandard hinausgehende weitere qualitative Kriterien, anhand derer sich die Qualität einer Information bestimmen lassen könnte. Der bereits angesprochene Maßstab der Vollständigkeit benötigt, um effektiv zu sein, eine Bezugsgröße. Die EG-Datenschutzverordnung erreicht dies, indem die Vollständigkeit auf den mit der Datenverarbeitung angestrebten Zweck bezogen wird.⁴²⁹ Im Rahmen von vertraglichen Beziehungen kann eine solche Bezugsgröße abgesprochen bzw. durch Auslegung ermittelt werden. Verspricht beispielsweise ein Big-Data-Unternehmen seinem Geschäftspartner, Informationen über das Kaufverhalten von Kunden in einem Einkaufszentrum auszuwerten und zu übermitteln, so kommt trotz Korrektheit der Informationen eine Haftung mangels Vollständigkeit in Frage, sofern nicht berücksichtigt wurde, dass der Geschäftspartner etwa ein Brautmodengeschäft eröffnen will und dementsprechend eine Differenzierung der Kundeninformationen nach Geschlecht und Alter notwendig wäre. Losgelöst von vorgegebenen Bezugsgrößen kann die Vollständigkeit dagegen kaum ein wirksames qualitatives Kriterium sein. Gleiches gilt für die Nützlichkeit als Bewertungsmaßstab der Informationsqualität. Denn auch hier wirkt es sich aus, dass bei der Frage nach der Fehlerhaftigkeit von Informationen nicht nur die Information als solche, sondern auch der Sender, der Empfänger und der Kontext, in dem die Information verbreitet wird, von Bedeutung sind.

Besonders intensiv hat sich der Gesetzgeber in den USA mit dem Thema Informationsqualität befasst. Dort wurden ein „Federal Data Quality Act“ sowie diesen konkretisierende Richtlinien erlassen.⁴³⁰ Diese beziehen sich zwar ausschließlich auf die Informationsverbreitung durch Behörden. Nichtsdestotrotz können die darin aufgestellten Kriterien womöglich teilweise auf die Frage der Qualität großer Datenmengen übertragen werden. Ausdrücklich wird in den Richtlinien neben der Nützlichkeit auf das Kriterium der

⁴²⁶ Abrufbar unter <http://conventions.coe.int/Treaty/GER/Treaties/Html/108.htm>.

⁴²⁷ VO (EG) 45/2001.

⁴²⁸ Vgl. § 6 LPG NRW; § 3 Abs. 2 BayPrG; § 3 Abs. 2 LPG Berlin.

⁴²⁹ Vgl. Art. 4 lit. d) VO (EG) 45/2001.

⁴³⁰ Vgl. <http://www.whitehouse.gov/sites/default/files/omb/fedreg/reproducible2.pdf>.

„Objektivität“ abgestellt.⁴³¹ Bemerkenswert im Hinblick auf die Objektivität der Informationen ist, dass diese vermutet wird, sofern unabhängige Sachverständige sie nach förmlicher Untersuchung bestätigt haben.⁴³² Offensichtlich hat der Normgeber erkannt, dass dieses eigentlich inhaltliche Kriterium unscharf ist und deshalb mögliche Konflikte über dessen Vorliegen durch die Eröffnung eines prozeduralen Weges entschärft. Auch dieses Beispiel zeigt, dass die Aufstellung allgemeiner Kriterien zur Bestimmung der Qualität einer Information schwierig ist.

Die Gefahr einer Lähmung des Informationsaustauschs durch zu hohe und gleichzeitig zu vage formulierte Qualitätskriterien zeigt sich am Beispiel der „Nachvollziehbarkeit“. In Bezug auf bestimmte Informationen fordern US-Richtlinien, dass die Methoden der Informationsgewinnung transparent dargestellt werden, um die Reproduzierbarkeit und damit die Nachvollziehbarkeit der Informationen durch unabhängige Dritte zu ermöglichen.⁴³³ Dieses Kriterium nahm eine landwirtschaftliche Lobbygruppe zur Hilfe, um gegen eine behördliche Warnung vor Gesundheitsgefahren durch ein Produkt vorzugehen. Die Argumentation stützte sich im Wesentlichen darauf, dass die Forschungsmethode, mit der das Ergebnis erreicht wurde, nicht unumstritten sei und die gezogenen Schlüsse deshalb nicht nachvollziehbar seien.⁴³⁴ Auch wenn ein solches Problem im Geschäftsbereich zweier Unternehmen untereinander wohl eher unwahrscheinlich ist, deutet sich doch eine gewisse Gefahr an, dass Qualitätsstandards genutzt werden könnten, um unliebsame Informationen zu bekämpfen oder einen Informationsvorsprung eines Konkurrenten zu verringern.

Alles in allem gibt es im Bereich der Informationsqualität noch viele offene Fragen. Ob eine Information oder ein Datum „gut“ sind und somit eine Haftung mangels Fehlinformation ausgeschlossen ist, lässt sich abstrakt nur mit dem Maßstab der sachlichen Richtigkeit bestimmen. Darüber hinaus können sich im vertraglichen Bereich aus den Umständen höhere Anforderungen an die Qualität der Information ergeben. Dabei spielen Kriterien wie Vollständigkeit, Nachvollziehbarkeit, Objektivität oder auch Nützlichkeit eine Rolle.

2. Vertragliche Haftung

Lässt sich im Rahmen einer vertraglichen Beziehung nach den oben dargestellten Grundsätzen die Übermittlung einer fehlerhaften Information zwischen einem Big-Data-Unternehmen und seinem Kunden feststellen, so kommt ein Schadensersatzanspruch wegen Pflichtverletzung des Schuldverhältnisses gem. § 280 Abs. 1 BGB in Betracht. Dabei kann zum einen die Verletzung einer Hauptleistungspflicht, zum anderen die Verletzung einer Nebenleistungspflicht aus § 241 Abs. 2 BGB vorliegen.

Während es im Big-Data-Bereich offensichtlich noch keine höchstrichterlichen Entscheidungen bezüglich einer vertraglichen Haftung für fehlerhafte Informationen gibt, zeigen sich in verwandten Geschäftsfeldern vergleichbare Probleme. Das bereits

⁴³¹ Gasser, S. 11 f.

⁴³² Gasser, S. 17.

⁴³³ Gasser, S. 18.

⁴³⁴ Gasser, S. 21.

angesprochene Kriterium der Vollständigkeit war Hauptstreitpunkt zwischen einem Anlageberater und einem Kunden, der wegen fehlerhafter Aufklärung und somit Verletzung einer Hauptleistungspflicht Schadensersatz verlangte.⁴³⁵ Unter Berücksichtigung aller Umstände entschied der *BGH* entgegen der Auffassung der Vorinstanzen, dass die erteilten Informationen vollständig waren. Die gegenläufige Argumentation der mit den Streitigkeiten befassten Gerichte zeigen jedoch, wie schwer im Einzelfall die Feststellung einer den Schadensersatz auslösenden Pflichtverletzung sein kann. Ähnliche Schwierigkeiten könnten auch im Big-Data-Geschäftsbereich drohen.

Klarer war die Entscheidung im Fall einer Schadensersatzklage gegen einen Steuerberater.⁴³⁶ Dort stellte der *BGH* das Vorliegen einer sachlich unrichtigen Information fest und sah deshalb eine die Haftung auslösende Pflichtverletzung als gegeben an. Für die Annahme der haftungsausfüllenden Kausalität argumentierte das Gericht mit dem Beweis des ersten Anscheins. Inwiefern dies für Big-Data-Unternehmen relevant ist, wird an späterer Stelle untersucht werden.⁴³⁷

Auch im Rahmen von vertraglichen Nebenpflichten können fehlerhafte Informationen für eine Haftung relevant sein. So entschied der *BGH*, dass ein Makler, der im Exposé eine objektiv falsche Angabe machte, seine Sorgfaltspflichten gegenüber dem Kunden verletze.⁴³⁸ Gleiches gilt für einen Krankenhausmitarbeiter, der sich dazu bereit erklärte, einem Patienten bei der Erstellung seines Testaments zu helfen.⁴³⁹ Eine fehlerhafte Information sorgte in diesem Fall für die Unwirksamkeit des Testaments. Das *OLG Hamm* entschied, dass der Krankenhausträger dem nicht wirksam zum Erben eingesetzten Kläger dessen Schaden ersetzen muss. An der Grenze zwischen objektiver Richtigkeit und Vollständigkeit der Information befand sich ein weiterer Fall, in dem es erneut um die Haftung eines Maklers ging.⁴⁴⁰ Trotz festgestellter objektiver Fehlerhaftigkeit der Auskunft ging der *BGH* nicht von einer Schadensersatzpflicht aus, da sich aus den Umständen ergab, dass der Makler sorgfältig hinsichtlich der Information gehandelt hatte und auf deren Richtigkeit vertrauen durfte. Insofern hatte er alles Notwendige getan, um vollständig zu informieren. Dieses Beispiel zeigt, dass trotz Unterschreitens der qualitativen Mindestanforderungen, eine vertragliche Haftung aufgrund der Gesamtumstände – insbesondere der Erfüllung anderer qualitativer Kriterien – ausscheiden kann. Ob dieses Ergebnis auch auf Big-Data-Unternehmen übertragbar ist, scheint allerdings zweifelhaft. Immerhin ist es in der Regel deren Hauptpflicht, Informationen zu übermitteln, so dass im Falle der reinen Weiterleitung erhaltenener Informationen die Berufung auf ein Vertrauen in deren Richtigkeit nicht unproblematisch ist. In diesem Bereich bestehen noch einige Unklarheiten, was weitergehende Forschung notwendig macht.

⁴³⁵ *BGH*, WM 2013, 836.

⁴³⁶ *BGH*, WM 2009, 369.

⁴³⁷ Vgl. Abschnitt 4.

⁴³⁸ *BGH*, NJW 2000, 3642.

⁴³⁹ *OLG Hamm*, Urteil vom 13.04.2010 – I-21 U 94/09.

⁴⁴⁰ *BGH*, NJW-RR 2007, 711.

3. Deliktische Haftung

Führt eine fehlerhafte Information zu einer Verletzung eines Rechtsgutes im Sinne des § 823 Abs. 1 BGB, kommt für den Geschädigten unabhängig von einer vertraglichen Beziehung zum Schädiger ein Schadensersatzanspruch aus dem Deliktsrecht in Betracht. Einen solchen hat der BGH im Falle einer fehlerhaften Angabe in einem Nachschlagewerk für Mediziner geprüft.⁴⁴¹ Aufgrund eines in dem Werk enthaltenen Druckfehlers hatte ein Arzt einen Behandlungsfehler verursacht. Im Wege eines Prozessvergleichs zahlte die Haftpflichtversicherung des Arztes sowie der Krankenhausträger eine finanzielle Entschädigung an den Patienten. Den gezahlten Betrag wollte sich die Versicherung wiederum vom Verleger des Buches ausgleichen lassen. Einen solchen Ausgleichsanspruch lehnte der BGH allerdings ab. In diesem Zusammenhang stellte er einige Grundsätze auf, die für die Informationshaftung generell von Bedeutung sein könnten.

a) Aussagen des BGH im Fall „Kochsalzlösung“

Nach Ansicht des BGH kann nicht davon ausgegangen werden, dass jeder Fehler eines in den Verkehr gebrachten Erzeugnisses eine Haftung desjenigen begründet, der nach den Grundsätzen der Herstellerhaftung an sich verantwortlich ist. Zudem erlaube das Auftreten eines einzelnen (Druck-)Fehlers nicht bereits den Schluss, dass der für die Korrektur Verantwortliche die im Verkehr erforderliche Sorgfalt vernachlässigt habe.⁴⁴² Diese für den Verlag positiven Feststellungen wurden dahingehend ergänzt, dass der Verleger im Regelfall ohne Verstoß gegen eine Verkehrssicherungspflicht die Korrektur dem Verfasser übertragen darf. Nur unter besonderen Umständen – etwa bei mathematischen und technischen Tabellen, baustatischen Arbeitsanleitungen und Ähnlichem – erfordere die Verkehrssicherungspflicht vom Verleger Maßnahmen, die Fehler mit Sicherheit vermeiden.⁴⁴³

b) Übertragung der Leitlinien auf Big-Data-Unternehmen

Überträgt man diese Maßstäbe, ergibt sich, dass ein Anbieter, dessen Produkt aufgrund fehlerhafter Informationen Schäden verursacht, grundsätzlich nicht deliktisch haftet. Ausnahmsweise kommt jedoch eine Haftung wegen Verletzung der Verkehrssicherungspflichten in Betracht, wenn keine Maßnahmen getroffen wurden, die vor Fehlern schützen. Gerade der in diesem Zusammenhang ergehende Hinweis auf mathematische und technische Tabellen, sowie baustatische Arbeitsanleitungen und Ähnliches lässt es möglich erscheinen, dass Big-Data-Unternehmen, die ihrerseits Daten erstellen oder beziehen und die daraus abzuleitenden Informationen weiterverbreiten, für deren Richtigkeit verkehrssicherungspflichtig sind. Eine Überprüfungs- und Korrekturpflicht hinsichtlich der ausgesandten Informationen ist demnach nicht ausgeschlossen. Ob das in Bezug auf sämtliche Geschäftsbereiche gilt, oder ob die Verkehrssicherungspflicht lediglich bei Informationen mit unmittelbarem Bezug zu Leib und Leben von Menschen besteht, lässt sich allerdings aus der Entscheidung nicht ableiten. Eine vollumfängliche

⁴⁴¹ BGH, NJW 1970, 1963 - Kochsalzlösung.

⁴⁴² BGH, NJW 1970, 1963 (1963).

⁴⁴³ BGH, NJW 1970, 1963 (1964).

Verkehrssicherungspflicht für Big-Data-Unternehmen hätte jedoch zur Folge, dass Kontrollmechanismen in den Unternehmen eingeführt werden müssten, die zum einen nicht unbedeutliche Kosten nach sich ziehen würden, und zum anderen die Geschwindigkeit des Informationsflusses herabsetzen könnten. Angesichts dieser negativen Konsequenzen, sowie der Tatsache, dass Verkehrssicherungspflichten nach Ansicht des *BGH* die Ausnahme sein sollen, spricht vieles dafür, eine generelle Verkehrssicherungspflicht für Big-Data Unternehmen abzulehnen.⁴⁴⁴

c) Einfluss der Grundsätze zur Produzentenhaftung

Eine tiefergehende Diskussion dieser Problematik erübrigt sich jedoch, falls sich die vor über 40 Jahren aufgestellten Grundsätze des *BGH* mittlerweile überholt haben. Der Fall würde heutzutage womöglich anders entschieden werden.⁴⁴⁵ Zum einen liegt dies an der Einführung des Produkthaftungsgesetzes, wonach eine verschuldensunabhängige Haftung eingreift, die nicht an das schuldhafte Verhalten des Produzenten, sondern an die Gefährlichkeit des Produkts an sich anknüpft.⁴⁴⁶

Zum anderen könnten auch die Grundzüge der Produzentenhaftung gem. § 823 Abs. 1 BGB zu einer Haftung des Schädigers führen. Hierbei handelt es sich um eine Verschuldenshaftung mit Beweiserleichterungen für den Kläger.⁴⁴⁷ Der Grundgedanke dieser Rechtsprechung besteht ebenfalls in der Annahme einer Verkehrssicherungspflicht. So hat der Hersteller eines Produkts dafür zu sorgen, dass die aus diesem Produkt für andere entstehenden Gefahren möglichst gering sind.⁴⁴⁸ Wird bei bestimmungsgemäßer Benutzung eines Erzeugnisses sodann eine Person oder eine Sache dadurch geschädigt, dass das Produkt fehlerhaft hergestellt wurde, muss der Hersteller beweisen, dass ihn hinsichtlich des Fehlers kein Verschulden trifft.

Die Übertragung dieser Rechtsprechung auf die Haftung von Big-Data-Unternehmen wirft Probleme auf, denn der ursprüngliche Anwendungsbereich der deliktsrechtlichen Produzentenhaftung lag bei Industrieprodukten.⁴⁴⁹ Davon ausgehend könnte eine Verkörperung der Information in einem Medium erforderlich sein, um überhaupt die deliktische Haftung auszulösen. Gerade bei der Übertragung von Daten über das Internet besteht diese Verkörperung normalerweise nicht. Deshalb fragt sich, ob etwaige Fehlinformationen unter die anerkannten Fallgruppen der Produzentenhaftung (Konstruktionsfehler, Fabrikationsfehler, Instruktionsfehler oder Produktbeobachtungspflichtversäumnisse) subsumiert werden können.

⁴⁴⁴ Ähnlich *Reese*, DStR 1994, 1121 (1122 f.).

⁴⁴⁵ *Meyer*, ZUM 1997, 26 (26).

⁴⁴⁶ *Oechsler*, in: *Staudinger*, Einl. zum ProdHaftG Rn. 27.

⁴⁴⁷ *Lehmann*, NJW 1992, 1721 (1721).

⁴⁴⁸ *Meyer*, ZUM 1997, 26 (30).

⁴⁴⁹ *Meyer*, ZUM 1997, 26 (30).

In Betracht kommt ein Vergleich mit der Einordnung von Software-Fehlern.⁴⁵⁰ Dort ist nach überwiegender Ansicht der Literatur die Produzentenhaftung einschlägig.⁴⁵¹ Da insbesondere bei Software regelmäßig nicht das körperliche Medium den Fehler aufweist, sondern ein darauf enthaltenes Datum, ist eine Übertragung dieser Grundsätze auf das Big-Data-Management möglich. Dies muss erst recht vor dem Hintergrund gelten, dass auch Software immer häufiger in unkörperlicher Form verbreitet wird. Dementsprechend erscheint die Anwendung der Produzentenhaftung auch ohne Verkörperung möglich.

4. Anscheinsbeweis

Große Relevanz besitzt in der haftungsrechtlichen Praxis der Anscheinsbeweis. Dieser ermöglicht es dem Richter, im Rahmen der freien Beweiswürdigung aus feststehenden Tatsachen unter Berücksichtigung der Lebenserfahrung Schlüsse auf das Vorliegen streitiger Tatsachenbehauptungen zu ziehen.⁴⁵² So ist es möglich mit Hilfe des Anscheinsbeweises, aus einem aufgetretenen Schaden aufgrund der Typizität des Geschehensablaufs auf das Verschulden einer Partei zu schließen. Problematisch ist jedoch, inwieweit dieses Instrument bei Haftungsfragen im Big-Data-Geschäftsbereich Anwendung finden kann.

a) Vertragliche Beziehungen

Zunächst ist zu untersuchen, ob es Anhaltspunkte für das Eingreifen des Anscheinsbeweises im Rahmen vertraglicher Haftung gibt. Bezieht der Geschädigte auf vertraglicher Grundlage Informationen von einem Big-Data-Unternehmen, so kann es vorkommen, dass er aufgrund falscher oder unvollständiger Informationen eine Entscheidung trifft, die zu einem Vermögensschaden führt. In einem solchen Fall wäre es für den Geschädigten von Vorteil, wenn die haftungsausfüllende Kausalität bejaht wird, weil er sich auf den Beweis des ersten Anscheins berufen kann. Eine gerichtliche Entscheidung zu dieser Frage liegt im Big-Data-Business noch nicht vor. Möglicherweise lässt sich jedoch eine Tendenz aus der Rechtsprechung zu Beratungsfehlern ablesen. In diesem Bereich wird ebenfalls einseitig eine Information übermittelt, die Grundlage für bedeutende wirtschaftliche Entscheidungen des Empfängers ist, so dass insofern eine strukturelle Vergleichbarkeit zu den hier im Fokus stehenden Beziehungen gegeben ist.

Im Rahmen von Verträgen mit Rechts- oder Steuerberatern wurde die „Vermutung beratungsgerechten Verhaltens“ entwickelt. Sie gilt allerdings nur dann, wenn im Hinblick auf die Interessenlage oder andere objektive Umstände ein bestimmter Entschluss des zutreffend unterrichteten Mandanten mit Wahrscheinlichkeit zu erwarten gewesen wäre.⁴⁵³ Ohne tatsächliche Feststellungen, die im Falle sachgerechter Aufklärung durch den Berater aus der Sicht eines vernünftig urteilenden Mandanten eindeutig eine bestimmte tatsächliche Reaktion nahe gelegt hätten, wirkt die Beweiserleichterung für den Mandanten nicht.⁴⁵⁴ Kommen also unter wirtschaftlichen Gesichtspunkten verschiedene Verhaltensweisen des

⁴⁵⁰ Vgl. dazu *Lehmann*, NJW 1922, 1721 (1723).

⁴⁵¹ *Reese*, DStr, 1994, 1121 (1122).

⁴⁵² *Prütting*, in: MüKO, § 286 ZPO Rn. 48.

⁴⁵³ BGH, NJW 2009, 1591 (1592).

⁴⁵⁴ BGH, NJW-RR 2006, 1645 (1646); NJW 1993, 3259 (3259).

Informationsempfängers ernsthaft in Betracht und besteht die Aufgabe des Informanten lediglich darin, dem Vertragspartner durch die erforderlichen fachlichen Informationen eine sachgerechte Entscheidung zu ermöglichen, ist der Anscheinsbeweis nicht anwendbar.⁴⁵⁵

Mit Hilfe dieser Leitlinien könnte sich die Haftung eines Big-Data-Unternehmens bestimmen lassen. Ähnlich wie bei Verträgen mit Rechtsanwälten und Steuerberatern kommt dann ein Schluss auf das Verschulden des Big-Data-Unternehmens in Betracht, wenn nach der Lebenserfahrung bei vertragsgemäßer Leistung für den Kunden lediglich eine einzige sinnvolle Entschlussmöglichkeit bestand, durch die der Schaden vermieden worden wäre.⁴⁵⁶ Allerdings ist in diesem Zusammenhang zu beachten, dass die „Vermutung beratungsgerechten Verhaltens“ eine Ausnahme zum Grundsatz ist, dass es keinen Anscheinsbeweis für individuelle Verhaltensweisen von Menschen in bestimmten Lebenslagen gibt. Dementsprechend tendiert die Rechtsprechung dazu, diese Ausnahme nur mit Vorsicht bei klaren und eindeutigen Ausgangslagen anzuwenden.⁴⁵⁷ Läge ein solcher Fall aber vor, so würde es dem Big-Data-Unternehmen obliegen, die bestehende richterliche Überzeugung durch einen Gegenbeweis zu erschüttern.⁴⁵⁸ Ob die Übertragung dieser Rechtsprechungsgrundsätze auf das Big-Data-Business tatsächlich sachgerechte Ergebnisse liefern kann oder ob es alternative Lösungswege gibt, muss im Rahmen weiterer Forschungsmaßnahmen geklärt werden.

b) Deliktsrecht

Während die Annahme des Anscheinsbeweises im Vertragsrecht maßgeblich vom Eingreifen der Vermutung beratungsgerechten Verhaltens und somit von den Umständen im Einzelfall abhängen könnte, erscheint eine andere Beurteilung im Hinblick auf die deliktische Haftung möglich. Wie bereits dargelegt, beruhen sowohl die vom *BGH* in der Kochsalzlösung-Entscheidung aufgestellten Grundsätze für die deliktische Haftung als auch die Produzentenhaftung auf der Annahme von Verkehrssicherungspflichten für denjenigen, der Informationen verbreitet. Bei der Verletzung von Verkehrssicherungspflichten ist der Beweis des ersten Anscheins laut *BGH* aber gerade dann geboten, wenn sich im Schadensfall diejenige Gefahr verwirklicht, die durch die Auferlegung bestimmter Verhaltenspflichten vermieden werden soll.⁴⁵⁹ Geht man also davon aus, dass dem Big-Data-Unternehmen eine Verkehrssicherungspflicht für die Korrektheit der Informationen – zumindest bei solchen mit unmittelbaren Bezug zu Leib und Leben⁴⁶⁰ – obliegt, so könnte bei Schadenseintritt trotz ordnungsgemäßer Verwendung der Information der Anscheinsbeweis angenommen werden. Allerdings kommt der Anscheinsbeweis nicht in Betracht, wenn die Verletzung der Verkehrssicherungspflicht selbst noch nicht erwiesen ist.⁴⁶¹ Insgesamt gibt es also auch in diesem Bereich einige offene Fragen. Klarheit könnten insoweit die Erforschung von Art und

⁴⁵⁵ OLG Düsseldorf, NJOZ 2012, 1356 (1356).

⁴⁵⁶ OLG Köln, BeckRS 2012, 10350.

⁴⁵⁷ BGH, NJW 2009, 1591 (1592).

⁴⁵⁸ Prütting, in: MüKo, § 286 ZPO Rn. 65.

⁴⁵⁹ BGH, NJW 1994, 945 (945).

⁴⁶⁰ Vgl. dazu Abschnitt 3, Unterabschnitt b.

⁴⁶¹ Prütting, in: MüKo, § 286 ZPO Rn. 75.

Umfang der dem Big-Data-Unternehmen obliegenden Verkehrssicherungspflichten sowie ihre Auswirkungen auf den Beweis des ersten Anscheins bringen.

5. Haftungsausschluss

Angesichts der erheblichen Haftungsrisiken stellt sich letztlich die Frage, ob ein Big-Data-Unternehmen seine Haftung vertraglich beschränken kann. Wie ein Umkehrschluss zu § 276 Abs. 3 BGB zeigt, ist eine vertragliche Haftungsmilderung grundsätzlich möglich. Nach dieser Vorschrift kann lediglich die Haftung für Vorsatz nicht im Voraus – auch nicht teilweise, z.B. durch eine summenmäßige Haftungsbegrenzung – ausgeschlossen werden.⁴⁶² Für das fahrlässige sowie vorsätzliche Verhalten eines Erfüllungsgehilfen ist dagegen nach § 278 S. 2 BGB eine vollständige Haftungsfreizeichnung in der Regel zulässig. Etwas anderes gilt jedoch im Rahmen von allgemeinen Geschäftsbedingungen. So ist eine klauselmäßige Freizeichnung von grober Fahrlässigkeit – und zwar auch der des Erfüllungsgehilfen – wegen § 309 Nr. 7 lit. a und b BGB unwirksam. Dies gilt unabhängig davon, ob der Vertragsgegner des Verwenders Verbraucher oder Unternehmer ist.⁴⁶³ Hinsichtlich der Verletzung von Leben, Körper und Gesundheit ist über § 309 Nr. 7 lit. b BGB ein Haftungsausschluss allgemein untersagt.

Dagegen müssen Ausschlüsse bezüglich des Schadensersatzes bei Schlechterfüllung und Verzug nach § 307 BGB bewertet werden. In diesen Fällen gibt die Rechtsprechung zwei Grundlinien vor: Die vollständige oder weitgehende Freizeichnung von wichtigen Pflichten, nicht nur Hauptpflichten, ist unwirksam. Die (nur bei anderen als Körperschäden zulässige) summenmäßige Begrenzung von Ansprüchen ist so festzusetzen, dass in der Summe der Durchschnittsschaden erreicht wird.⁴⁶⁴ Die Möglichkeit einer Haftungsfreizeichnung hängt also von den vertraglichen Pflichten im konkreten Einzelfall ab.

G. Aufbewahrungs- und Löschungspflichten

Der Umgang mit großen Datenmengen erfordert die Beachtung zahlreicher gesetzlich geregelter Datenaufbewahrungs- und Datenlöschungspflichten.

Die Aufbewahrungspflichten für Daten werden u.a. im Zusammenhang mit dem Auskunftsrecht des Betroffenen gem. § 34 BDSG relevant. Die verantwortliche Stelle hat dem Betroffenen auf Verlangen Auskunft zu erteilen über die zu seiner Person gespeicherten Daten, auch soweit sie sich auf die Herkunft dieser Daten beziehen (§ 34 Abs. 1 Nr. 1 BDSG), den Empfänger oder die Kategorien von Empfängern, an die Daten weitergegeben werden (§ 34 Abs. 1 Nr. 2 BDSG), und den Zweck der Speicherung (§ 34 Abs. 1 Nr. 3 BDSG). Dabei treffen Stellen, die Daten geschäftsmäßig zur Übermittlung speichern, besondere Pflichten.⁴⁶⁵ Der Auskunftspflicht sind selbst solche Daten unterworfen, die zwar

⁴⁶² Grüneberg, in: Palandt, § 276 BGB Rn. 35.

⁴⁶³ BGH, NJW 2007, 3774; NJW 1991, 2630 (2362); Grundmann, in: MüKo, § 276 BGB Rn. 184.

⁴⁶⁴ Grundmann, in: MüKo, § 276 BGB Rn. 185.

⁴⁶⁵ Gola/Schomerus, § 34 Rn. 8a.

gegenwärtig keinen Personenbezug haben, bei denen dieser aber durch die Auskunftserteilung hergestellt wird.⁴⁶⁶

Daten müssen aber so lange aufbewahrt werden, dass die betroffene Person die Möglichkeit hat, sich von der einwandfreien, d.h. fehlerfreien Verarbeitung ihrer Daten zu vergewissern.⁴⁶⁷ Dabei muss das Auskunftsrecht auch in die Vergangenheit reichen. Andernfalls, so der *EuGH*, „wäre die betreffende Person weder in der Lage, wirksam ihr Recht auf Veranlassung der Berichtigung, Löschung oder Sperrung von Daten wahrzunehmen, [...] noch einen gerichtlichen Rechtsbehelf einzulegen und Schadensersatz zu erlangen.“⁴⁶⁸

Bislang ungeklärt ist, wie weit genau dieses Recht zurückreicht.⁴⁶⁹ Jede Frist muss aber einen „gerechten Ausgleich [...] zwischen dem Interesse der betroffenen Person am Schutz ihres Privatlebens [...] und der Belastung, die die Pflicht zur Aufbewahrung der betreffenden Information für den mit der Verarbeitung Verantwortlichen“⁴⁷⁰ mit sich bringt, darstellen. Nach Ansicht des *EuGH* „erhöhe sich mit der Menge der Daten und der Anzahl der Empfänger bzw. der Frequenz der Übertragungen“⁴⁷¹ dieser Aufwand, was insbesondere Big-Data-Anwendungen betrifft.

Forschungsbedarf im Rahmen von Big Data ergibt sich folglich aus der Herausarbeitung und Abwägung dieser widerstreitenden Interessen und deren Abstimmung mit den sonstigen gesetzlichen Aufbewahrungspflichten, unter anderem aus der Abgabenordnung (AO) und dem Handelsgesetzbuch (HGB). Diese sehen Aufbewahrungsfristen von bis zu zehn Jahren vor, §§ 147 AO bzw. 257 HGB.

Zukünftige Forschung wird betroffenen Personen wie Datenverarbeiter genauere Richtlinien zur Aufbewahrung(-sfrist) und zum Umgang mit aufbewahrten Daten geben müssen, um die Rechtssicherheit gerade für Big-Data-Anwendungen zu erhöhen. Insbesondere die Frage, wann eine Datenmenge so groß ist, dass eine Speicherung einen zu großen Aufwand darstellt, bedarf angesichts immer neuer Speicherkapazitäten der Klärung.

Spiegelbildlich zur Datenaufbewahrungspflicht sehen zahlreiche Gesetze Datenlöschungspflichten vor. Diese sind vor allem zur Sicherung des Grundrechts auf informationelle Selbstbestimmung von zentraler Bedeutung und bei der Entwicklung einer Big-Data-Anwendung zu berücksichtigen. Eine erschöpfende Aufzählung der jeweiligen Vorschriften ist kaum möglich. Aus datenschutzrechtlicher Sicht sind personenbezogene Daten insbesondere dann zu löschen, wenn ihre Kenntnis für die Erfüllung des Zwecks der Speicherung nicht mehr erforderlich ist, § 35 Abs. 2 Nr. 3 BDSG. Ferner ist eine Löschung von Daten angezeigt, wenn sie geschäftsmäßig zum Zweck der Übermittlung verarbeitet werden

⁴⁶⁶ *Gola/Schomerus*, § 34 Rn. 8b.

⁴⁶⁷ *EuGH*, EuZW 2009, 546 (548).

⁴⁶⁸ *EuGH*, EuZW 2009, 546 (548).

⁴⁶⁹ *EuGH*, EuZW 2009, 546 (548).

⁴⁷⁰ *EuGH*, EuZW 2009, 546 (549).

⁴⁷¹ *EuGH*, EuZW 2009, 546 (546).

und eine Prüfung jeweils am Ende des vierten, soweit es sich um Daten über erledigte Sachverhalte handelt und der Betroffene der Löschung nicht widerspricht, am Ende des dritten Kalenderjahres ergibt, dass eine länger währende Speicherung nicht erforderlich ist, § 35 Abs. 2 Nr. 4 BDSG. Videoaufzeichnungen von öffentlich zugänglichen Räumen sind unverzüglich zu löschen, wenn sie zur Erreichung des Zwecks nicht mehr erforderlich sind oder schutzwürdige Interessen der Betroffenen einer weiteren Speicherung entgegenstehen, § 6b Abs. 5 BDSG.

Daneben gelten verschiedene Datenlöschungspflichten im Gesundheit- und Sozialwesen. § 84 SGB X sieht die Pflicht zur Berichtigung, Löschung und Sperrung von Sozialdaten vor. Hierauf verweist etwa auch § 304 Abs. 1 S. 1 SGB V hinsichtlich der Löschung von Daten, die bei Krankenkassen gespeichert sind. Eine vergleichbare Regelung trifft § 284 SGB V.

Datenlöschungspflichten berühren darüber hinaus auch das Strafrecht. Dies gilt etwa für die Löschung personenbezogener Daten von Gefangenen (§ 184 Abs. 1 S. 1 StVollzG), die Löschung von Daten aus dem Waffenregister (§ 18 Abs. 1, 2 NWRG) oder die Löschung personenbezogener Daten, die in unzulässiger Weise vom Bundeskriminalamt gespeichert worden sind, § 32 Abs. 2 S. 1 BKAG.

Schließlich können für eine Big-Data-Anwendung vor allem die im TKG statuierten Löschungspflichten im Hinblick auf gespeicherte Verkehrsdaten relevant sein.⁴⁷² Gleichermaßen gilt je nach Ausgestaltung und Einsatzfeld der Big-Data-Anwendung auch für einzelne Löschungspflichten, die in verschiedenen Rechtsgebieten verstreut geregelt sind. Hierzu zählen beispielsweise § 6 Abs. 1 S. 2 Artikel 10-Gesetz (Löschung personenbezogener Daten, die durch den Bundesnachrichtendienst erhoben wurden), § 18 Abs. 4 S. 3 PassG (Löschung biometrischer Daten, die durch Beförderungsunternehmen erhoben wurden), § 14 Abs. 1 S. 1 BMG (Löschung von Daten, die durch die Meldebehörde erhoben wurden) und § 44 Abs. 1 StVG (Löschung von Daten in den Fahrzeugregistern).

H. Steuerrecht

Big Data ist mittlerweile auch zu einer Frage des Steuerrechts geworden. Gem. § 147 Abs. 1 AO sind Unterlagen aufzubewahren, soweit sie für die Besteuerung von Bedeutung sind. Dies beinhaltet unter anderem Bücher und Aufzeichnungen, Inventare, Jahresabschlüsse, Lageberichte, die Eröffnungsbilanz, die empfangenen Handels- oder Geschäftsbriebe und Buchungsbelege. Nach § 147 Abs. 3 S.1 AO sind Bücher und Aufzeichnungen, Inventare, Jahresabschlüsse, Lageberichte, die Eröffnungsbilanz sowie die zu deren Verständnis erforderlichen Arbeitsanweisungen und Organisationsanweisungen zehn Jahre, alle anderen Unterlagen 6 Jahre aufzubewahren.

Die allgemeinen Anforderungen an die Aufbewahrung ergeben sich aus § 146 Abs. 1 AO, wonach „Buchungen und sonst erforderliche Aufzeichnungen vollständig, richtig, zeitgerecht und geordnet vorzunehmen“ sind. Diese Mitwirkungspflichten können – und müssen

⁴⁷² Vgl. dazu Breyer, ZD-Aktuell 2012, 03218.

teilweise sogar – in elektronischer Weise und unter Einsatz von Computersystemen erfüllt werden.⁴⁷³ Bis auf Jahresabschlüsse und die Eröffnungsbilanz, welche im Original vorliegen müssen, können alle erforderlichen Unterlagen auch auf digitalen Datenträgern aufbewahrt werden, wenn sich nicht aus anderen Gesetzen die Verpflichtung zur Speicherung der Originale ergibt.⁴⁷⁴ Dabei muss nach § 147 Abs. 2 AO gewährleistet werden, dass die digitale Wiedergabe bezüglich der Buchungsbelege sowie der Handels- und Geschäftsbriefe bildlich und bezüglich der sonstigen Unterlagen inhaltlich mit den Originalen übereinstimmen und diese während des Aufbewahrungszeitraums jederzeit verfügbar sowie abrufbar sind und maschinell ausgewertet werden können.

Im Zusammenhang mit Big Data werden diese Aufbewahrungspflichten vor allem relevant, wenn die gespeicherten Daten ausgelagert werden. Gem. § 146 Abs. 2 Satz 1 AO sind die Bücher und die sonst erforderlichen Aufzeichnungen grundsätzlich „im Geltungsbereich dieses Gesetzes zu führen“ und aufzubewahren. Insofern ist ein Outsourcing der elektronischen Daten an einen Auftragnehmer in Deutschland mit Blick auf die AO durchaus möglich, Probleme können sich aber im Zusammenhang mit Outsourcing auf ausländische Auftragnehmer ergeben.

Nach § 146 Abs. 2a AO kann die zuständige Finanzbehörde auf schriftlichen Antrag des Steuerpflichtigen bewilligen, dass elektronische Bücher und sonstige erforderliche elektronische Aufzeichnungen oder Teile davon außerhalb des Geltungsbereichs der AO in einem EU-Mitgliedsstaat oder einem EWR-Mitgliedsstaat geführt und aufbewahrt werden können. Voraussetzung für eine Genehmigung ist, dass der Steuerpflichtige der zuständigen Finanzbehörde den Standort des Datenverarbeitungssystems und bei Beauftragung eines Dritten dessen Namen und Anschrift mitteilt, der Steuerpflichtige seinen sich aus den §§ 90, 93, 97, 140 bis 147 AO und 200 Abs. 1 und 2 AO ergebenden Pflichten ordnungsgemäß nachgekommen ist, der Datenzugriff der deutschen Finanzbehörde nach § 147 Abs. 6 AO in vollem Umfang möglich ist und die Besteuerung hierdurch nicht beeinträchtigt wird. Zudem besteht die Pflicht einer umfassenden Verfahrensdokumentation. Liegen diese Voraussetzungen bei einem Antrag auf Verlagerung der elektronischen Buchführung in das EU-/EWR-Ausland vor, wird ein Anspruch des Steuerpflichtigen auf Bewilligung durch das Finanzamt zu bejahen sein. Sind hingegen die Voraussetzungen der Nr. 1 oder Nr. 2 nicht erfüllt, darf nach § 146 Abs. 2a Satz 5 AO das Finanzamt die Bewilligung nur erteilen, wenn die Besteuerung hierdurch nicht beeinträchtigt wird. Wie diese Anforderungen bei Speicherung extrem großer Datensätze erfüllt werden können, ist herauszuarbeiten.

Auch eine Verlagerung der elektronischen Buchführung in das übrige Ausland ist, wie sich aus § 146 Abs. 2 Satz 5 AO ergibt, grundsätzlich möglich. Jedoch wird eine Datenverlagerung in das nicht europäische Ausland regelmäßig steuerrechtlich problematisch sein. Nach § 146 Abs. 2a Satz 5 AO kann die zuständige Finanzbehörde, wenn die Voraussetzungen der Sätze 1 und 2 oder Satz 3 Nr. 1 oder Nr. 2 nicht vorliegen, die Führung und Aufbewahrung

⁴⁷³ v. Freedon, in: Kilian/Heussen, Teil 1, Abschnitt 9 Rn. 1.

⁴⁷⁴ v. Freedon, in: Kilian/Heussen, Teil 1, Abschnitt 9 Rn. 30.

elektronischer Bücher und sonstiger erforderlicher elektronischer Aufzeichnungen außerhalb des Geltungsbereichs dieses Gesetzes nur bewilligen, wenn die Besteuerung hierdurch nicht beeinträchtigt wird. Wann eine Besteuerung nicht beeinträchtigt wird, ergibt sich weder unmittelbar aus Absatz 2a noch aus den Gesetzesmaterialien oder aus § 148 AO, in dem ebenfalls dieses Tatbestandsmerkmal verwendet wird. Entscheidend dürfte sein, dass dem deutschen Finanzamt eine elektronische Außenprüfung nach § 147 Abs. 6 AO möglich ist und der Steuerpflichtige bisher seine steuerlichen Pflichten nicht verletzt hat.⁴⁷⁵ Die Bewilligung ist eine Ermessensentscheidung und stellt einen Verwaltungsakt dar. Widerruft die ausländische Stelle ihre Zustimmung für einen Zugriff durch die deutsche Finanzbehörde, ist die Bewilligung der Verlagerung der elektronischen Buchführung zu widerrufen, Rückverlagerung der EDV-Buchführung in das Inland zu verlangen und vom Steuerpflichtigen nachzuweisen. Kommt der Steuerpflichtige dieser Aufforderung nicht nach, kann ein sogenanntes Verzögerungsgeld (in Höhe von 2500 € bis 250 000 €) gegen ihn festgesetzt werden.

Die Einhaltung dieser Vorgaben ist auch beim Umgang mit extrem großen Datensätzen von besonderer Bedeutung. Fehlerhafte Buchführung kann für die Verantwortlichen erhebliche Konsequenzen haben. Eine nicht ordnungsmäßige Buchführung kann zu einer Schätzung nach § 162 AO, zu Zwangsmitteln nach § 328 AO oder zu einer Ahndung nach § 379 Abs. 1 AO führen. Bei Zahlungseinstellung oder Insolvenz kann es auch zu einer Bestrafung nach § 283 StGB (Bankrott) oder § 283 b StGB (Verletzung der Buchführungspflicht) kommen.

Besondere Relevanz haben die Grundsätze zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen (GDPdU). Gem. § 147 Abs. 6 der Abgabenordnung (AO) hat die Finanzbehörde im Rahmen einer Außenprüfung das Recht, Einsicht in die gespeicherten Daten zu nehmen und das Datenverarbeitungssystem zur Prüfung dieser Unterlagen zu nutzen, soweit Daten und Aufzeichnungen elektronisch geführt wurden. Sie kann im Rahmen einer Außenprüfung auch verlangen, dass die Daten nach ihren Vorgaben maschinell ausgewertet oder ihr die gespeicherten Unterlagen und Aufzeichnungen auf einem maschinell verwertbaren Datenträger zur Verfügung gestellt werden. Streitig ist dabei die Reichweite der Befugnisse der Finanzbehörden, insbesondere im Hinblick auf den Zugriff auf freiwillig erstellte elektronische Unterlagen.⁴⁷⁶ Wie der Bundesfinanzhof (BFH) klargestellt hat, sind nur die für das Verständnis und die Überprüfung gesetzlich geforderter Aufzeichnungen erforderlichen Unterlagen gem. § 147 Abs. 1 AO aufzubewahren.⁴⁷⁷ Insbesondere ein Zugriff auf Daten, die private aufzeichnungspflichtige Vorgänge betreffen, ist nicht zulässig.⁴⁷⁸

Es bedarf im Rahmen weiterer Forschung genauer Klärung, welche bei Big-Data-Anwendungen anfallenden Daten auch zum Verständnis und zur Überprüfung gesetzlich

⁴⁷⁵ Klein, Abgabenordnung, § 146 Rn. 5a.

⁴⁷⁶ BFH, BC 2009, 438 (438) m. Anm. d. Red.

⁴⁷⁷ BFH, U. v. 24.06.2009 – VIII R 80/06 = BC 2009, 438.

⁴⁷⁸ BFH, BC 2009, 438 (439) m. Anm. d. Red.

geforderter Aufzeichnungen erforderlich sind und welche eindeutig in den Bereich der privaten bzw. nicht-relevanten Vorgänge fallen.

I. Strafrecht

Das Strafgesetzbuch sanktioniert in verschiedenen Straftatbeständen den rechtswidrigen Umgang mit Daten. Hierzu zählen vor allem die §§ 202a, 202b, 202c, 203, 204, 303a und 303b StGB. Diese Normen sind auch im Hinblick auf das Big-Data-Management von besonderer Relevanz, denn die strafrechtlichen Dimensionen des Datenumgangs haben sich in den letzten Jahren erheblich verändert. So wird die digitale Gesellschaft mittlerweile von einer beinahe unüberschaubaren Masse an Daten bestimmt. Sogar der Staat selbst kauft etwa Daten-CDs, um Steuersünder zu überführen.

Die Frage, ob der strafrechtliche Rahmen dieser Entwicklung in ausreichender Weise Rechnung trägt, liegt damit auf der Hand. Somit gilt es, gegebenenfalls Vorschläge zur Anpassung des strafrechtlichen Rahmens zu entwickeln. Die Notwendigkeit einer Anpassung beschäftigt jedenfalls bereits den Gesetzgeber. Jüngst hat das Land Hessen im Bundesrat einen Gesetzentwurf eingebracht, nach dem im StGB der Tatbestand der Datenhehlerei in einem neuen § 202d StGB eingeführt werden soll. Der Ankauf von Steuerdaten durch den Staat soll allerdings weiterhin möglich bleiben: § 202d Abs. 5 StGB-E stellt in diesem Zusammenhang klar, dass „Handlungen, die ausschließlich der Erfüllung gesetzlicher Pflichten durch Amtsträger dienen“, davon ausgenommen sind. Neben der Einführung des neuen Straftatbestandes soll auch das Strafmaß zu den bestehenden § 202a StGB (Ausspähen von Daten) und § 202b StGB (Abfangen von Daten) erhöht werden, wenn die Taten in Bereicherungs- bzw. Schädigungsabsicht oder gewerbsmäßig/bandenmäßig erfolgen.

Darüber hinaus werfen auch die bereits bestehenden Normen des StGB im Hinblick auf das Big-Data-Management verschiedene Fragen auf. Die Ungewissheit über mögliche strafrechtliche Konsequenzen des Datenumgangs hemmt die volle Ausschöpfung des Potenzials, das die Auswertung großer Datenmengen in sich trägt. Es erscheint dringend geboten, diese Rechtsunsicherheit zu beseitigen – sei es durch gesetzgeberisches Handeln oder durch die Rechtsprechung.

Anhand des § 203 StGB werden in der Folge auftretende Probleme exemplarisch aufgezeigt. Hierbei wird insbesondere die Einordnung von Outsourcing-Dienstleistern problematisiert. Diese Fragestellung betrifft vor allem das Cloud Computing als typische Erscheinungsform des Big-Data-Managements.

Die unbefugte Weitergabe oder Offenbarung eines fremden Geheimnisses, namentlich eines zum persönlichen Lebensbereich gehörenden Geheimnisses oder eines Betriebs- oder Geschäftsgeheimnisses, das jemandem als Angehörigen eines Unternehmens der privaten Kranken-, Unfall- oder Lebensversicherung oder einer privatärztlichen, steuerberaterlichen

oder anwaltlichen Verrechnungsstelle anvertraut wurde, ist nach § 203 Abs. 1 Nr. 6 StGB strafbar. Die Aufzählung des § 203 Abs. 1 Nr. 6 StGB ist abschließend.⁴⁷⁹ Unter einem fremden Geheimnis im Sinne des § 203 StGB versteht man solche Tatsachen, die sich auf einen anderen Betroffenen beziehen und nur einem begrenzten Personenkreis bekannt sind.⁴⁸⁰ Der Geheimnisträger muss unter Berücksichtigung seiner persönlichen Situation ein sachlich begründetes Interesse an der Geheimhaltung haben. Bezuglich gesundheitlicher und finanzieller Verhältnisse ist dies regelmäßig der Fall.⁴⁸¹ Allein die Tatsache des Bestehens eines Versicherungsverhältnisses zwischen einer bestimmten Person und einem Unternehmen stellt bereits ein Geheimnis im Sinne des § 203 StGB dar.⁴⁸² Unter der Offenbarung eines Geheimnisses ist jegliche Mitteilung über die geheim zu haltende Tatsache an einen Dritten zu verstehen.⁴⁸³ Bei § 203 StGB handelt es sich um ein echtes Sonderdelikt, so dass Täter nur jemand sein kann, der Angehöriger eines der in § 203 Abs. 1 Nr. 6 StGB genannten Unternehmen ist. Zu Angehörigen im Sinne des § 203 Abs. 1 Nr. 6 StGB zählen Inhaber, Leiter, Organe bzw. deren Mitglieder sowie darüber hinaus sämtliche Bedienstete, die durch ihre Unternehmensfunktion bestimmungsgemäß mit Geheimnissen in Berührung kommen.⁴⁸⁴ Andere Personen können lediglich Teilnehmer gem. §§ 26, 27 StGB sein. Die Verletzung von Privatgeheimnissen im Sinne von § 203 StGB wird mit Freiheitsstrafe bis zu einem Jahr oder mit Geldstrafe geahndet.

Inwieweit diese Regelung die Zulässigkeit der Auslagerung von Daten an einen Auftragnehmer im Sinne von § 11 BDSG betrifft, ist in strafrechtlicher Hinsicht noch nicht abschließend geklärt. Zentraler Ansatzpunkt zur Beantwortung dieser Frage ist die Auslegung des Gehilfenbegriffs in § 203 Abs. 3 S. 2 StGB. An ihre berufsmäßig tätigen Gehilfen können Angehörige eines Unternehmens fremde Geheimnisse ohne die Zustimmung des Betroffenen weitergeben. Gehilfe ist jemand, der für einen zur Verschwiegenheit Verpflichteten eine unterstützende Tätigkeit ausübt, welche die Kenntnis fremder Geheimnisse mit sich bringt oder diese ohne Überwindung besonderer Hindernisse ermöglicht.⁴⁸⁵ Die Weitergabe von personenbezogenen Daten an einen Gehilfen ist dann eine interne Weitergabe, welche nicht mit Strafe bedroht ist.⁴⁸⁶ Bislang ist jedoch weder durch die Rechtsprechung noch durch das Schrifttum abschließend geklärt worden, welche Voraussetzungen erfüllt sein müssen, um einen Outsourcing-Dienstleister als Gehilfen im Sinne des § 203 Abs. 3 StGB anzusehen.⁴⁸⁷ Die wohl herrschende Meinung geht davon aus, dass bei Bestehen eines Dienst- oder Arbeitsverhältnisses jedenfalls ein Indiz für die Begründung einer berufsmäßigen Gehilfenstellung vorliegt.⁴⁸⁸ In der Literatur lassen sich

⁴⁷⁹ Weidemann, in: Beck'scher Onlinekommentar StGB, § 203 Rn. 14.

⁴⁸⁰ OLG Dresden, NJW 2007, 3509 (3509); Kühl, in: Lackner/Kühl, § 203 Rn. 14.

⁴⁸¹ Lenckner/Eisele, in: Schönke/Schröder, § 203 Rn. 7.

⁴⁸² BVerfG, NJW 2002, 2164 (2164 f.); Cierniak/Pohlit, in: Müko, § 203 StGB Rn. 24; Köpke, S. 32.

⁴⁸³ BGH, NJW 1992, 737; Kühl, in: Lackner/Kühl, § 203 Rn. 17.

⁴⁸⁴ Fischer, § 203 Rn. 21.

⁴⁸⁵ Lensdorf/Mayer-Wegelin/Mantz, CR 2009, 62 (64).

⁴⁸⁶ BGH, NJW 1995, 2915.

⁴⁸⁷ Hegmanns/Niehaus, NStZ 2008, 57 (58).

⁴⁸⁸ Hegmanns/Niehaus, NStZ 2008, 57 (59); Hoenike/Hülsdunk, MMR 2004, 788 (791).

einige Kriterien für diese finden. Der Gehilfe muss an der berufsmäßigen Schweigepflicht unterstützend teilnehmen, dem primär Schweigepflichtigen bewusst zuarbeiten und somit eine Hilfestellung leisten, die sich unmittelbar auf dessen Berufstätigkeit bezieht. Der Gehilfe muss ferner durch ein Direktionsrecht des primär zur Verschwiegenheit Verpflichteten direkt in die Arbeitsorganisation eingegliedert sein.⁴⁸⁹

Vor diesem Hintergrund ist der Outsourcing-Dienstleister beispielsweise wohl als „berufsmäßig tätiger Gehilfe“ eines Versicherungsunternehmens anzusehen. Die Verwaltung von Versichertendaten gehört zu dem Hauptaufgabengebiet eines Versicherungsunternehmens. Wer einer Versicherung bei der Verwaltung von Versichertendaten zuarbeitet, unterstützt sie somit unmittelbar in einem Kernbereich ihrer schweigepflichtbegründenden Tätigkeit.⁴⁹⁰

Um das Risiko einer Strafbarkeit zu verringern, sind zwei mögliche Vorgehensweisen bei der Auslagerung von Daten denkbar. Zum einen könnte der Auftraggeber die Daten in verschlüsselter Form auslagern. Wie oben beschrieben, ist ein Geheimnis dann offenbart, wenn es in irgendeiner Weise an einen Dritten gelangt. Dies ist jedoch bei einem verschlüsselten Datensatz nicht der Fall. Der „nicht zum Wissen Berufene“ kann zwar auf den Datensatz zugreifen, nicht jedoch auf dessen Inhalt – also das „eigentliche Geheimnis“. Problematisch erscheint bei der Weitergabe eines verschlüsselten Datensatzes jedoch der Fall, dass sich ein Unberechtigter durch Umgehen der Verschlüsselung Zugang zu den Daten verschafft. Bei verschlüsselten Daten handelt es sich um gegen unberechtigten Zugang besonders gesicherte Daten. Die unbefugte Zugänglichmachung dieser Daten ist nach § 202a StGB strafbar. Wird eine Verschlüsselung rechtswidrig umgangen, liegt seitens des Auslagernden kein Verstoß gegen § 203 StGB vor. Handeln Dritte im Rahmen des Beschaffens der Daten selbst rechtswidrig, scheidet ein Verstoß gegen die Verschwiegenheitspflicht aus.⁴⁹¹ Unterlässt das Unternehmen hingegen bedingt vorsätzlich Schutzmaßnahmen, könnte darin ein Verstoß gegen § 203 StGB zu sehen sein.⁴⁹² Welche Schutzmaßnahmen allerdings seitens der Daten auslagernden Stelle ausreichend sind, um im Fall einer rechtswidrigen Informationsverschaffung durch Dritte eine Strafbarkeit des Daten entsendenden Unternehmens zu vermeiden, lässt sich nicht vollumfänglich und abschließend beurteilen. Je aufwändiger die Verschlüsselung der Daten ist, desto unwahrscheinlicher erscheint in diesem Zusammenhang ein Verstoß gegen § 203 StGB. Im Ergebnis wird man hier im Einzelfall beurteilen müssen, ob der Grad der Verschlüsselung eine ausreichende Sicherung der Daten darstellt.

Zum anderen besteht die Möglichkeit einer Verletzung der Vorschrift des § 203 StGB dadurch vorzubeugen, dass ein doppeltes Arbeitsverhältnis begründet wird. Doppelarbeitsverhältnisse für solche Mitarbeiter, die in der Projektarbeit eingesetzt werden

⁴⁸⁹ Hilgendorf, Informationsstrafrecht und Rechtsinformatik, 2004, S. 92.

⁴⁹⁰ Hilgendorf, Informationsstrafrecht und Rechtsinformatik, 2004, S. 93.

⁴⁹¹ Sassenberg/Bamberg, DStR 2006, 2052.

⁴⁹² Hilgendorf, Informationsstrafrecht und Rechtsinformatik 2004, S. 101.

und gegebenenfalls die Daten einsehen, sollten genügen, sofern der Mitarbeiter tatsächlich organisatorisch in das Unternehmen eingebunden wird und diesem weisungsunterworfen ist.⁴⁹³ Freie Mitarbeiter sind hingegen als Externe anzusehen, es sei denn, sie sind hinsichtlich der organisatorischen Einbindung und Weisungsgebundenheit einem festangestellten Mitarbeiter gleichgestellt.⁴⁹⁴

J. Daten und Produkthaftungsrecht

Ein weiteres Problemfeld eröffnet sich bei der Frage, ob eine Haftung für fehlerhafte Daten nach dem ProdHaftG in Betracht kommt. Für alle im Bereich des Big-Data-Managements tätigen Personen, insbesondere jedoch für Unternehmen, deren Geschäftsmodell auf der Herstellung von Software beruht, ist es von herausragender Bedeutung, ob sie sich dem Risiko einer verschuldensunabhängigen Haftung gem. § 1 Abs. 1 ProdHaftG aussetzen oder nicht.

Zunächst ist festzustellen, dass die wissenschaftliche Diskussion weitestgehend auf dem Stand der Neunziger Jahre stehen geblieben ist. Trotz der raschen Entwicklung im IT-Bereich finden sich zu der Frage, ob auch Daten Produkte i.S.d. § 2 ProdHaftG sein können, keine Darstellungen aus den vergangenen Jahren. Umso notwendiger erscheint daher eine aktuelle Betrachtung der Problematik. Schließlich haben sich nicht nur die technischen Möglichkeiten massiv weiterentwickelt. Auch das Verhalten und die Einstellung der Internetnutzer haben sich verändert.

Hinzu kommt, dass die gesamte Thematik im bereits vorhandenen Schrifttum äußerst umstritten ist. Daraus ergeben sich rechtliche Unsicherheiten, die Risiken für Unternehmen unkalkulierbar machen und im schlimmsten Fall von einer unternehmerischen Betätigung abhalten. Im Folgenden wird zur Verdeutlichung der Problematik eine kurze Übersicht über die zentralen Streitpunkte gegeben. Dafür wird zunächst zwischen der Haftung für mangelhafte Software und der Haftung für Daten unterschieden. Die Beurteilung von Software wiederum unterscheidet sich zum Teil danach, ob es sich um Standard- oder Individualsoftware handelt und ob die Software auf Datenträgern geliefert oder nur heruntergeladen werden kann.

Zwar ist auch bei Standardsoftware umstritten,⁴⁹⁵ ob die Anwendung der Produkthaftung im Falle ihrer Fehlerhaftigkeit denkbar ist, jedoch scheint die Mehrheit der Literaturstimmen eine solche Haftung zu befürworten.⁴⁹⁶ Die Befürworter begründen ihre Ansicht u. a. damit,

⁴⁹³ Hilgendorf, Informationsstrafrecht und Rechtsinformatik 2004, S. 101.

⁴⁹⁴ Hilgendorf, Informationsstrafrecht und Rechtsinformatik 2004, S. 101.

⁴⁹⁵ Dagegen: *BFH*, CR 1987, 576; *Honsell*, JuS 1995, 211 (212); *Kort*, DB 1994, 1505 (1506); *Müller-Hengstenberg*, NJW 1994, 3128 (3131); *Taschner*, in: *Taschner/Frietsch*, Art. 6 Rn. 28; v. *Westphalen*, NJW 1990, 83 (87); aus der englischen Lit. *Tettenborn*, in: *Clerk & Lindsell on Torts*, No. 11–50.

⁴⁹⁶ Stellungnahme der Kommission v. 8. 5. 1989, ABI. EG Nr. C 114 S. 42; *Beckmann/Müller*, MMR 1999, 14 (15); *Cahn*, NJW 1996, 2899 (2904); *Hohmann*, NJW 1999, 521 (524); *Lehmann*, NJW 1992, 1721 (1724); *Marly*, BB 1991, 432 (433 ff.); *Schiemann*, in: *Erman* Rn. 2; *Spindler*, NJW 1999, 3737 (3742); *Sprau*, in: *Palandt* Rn. 1; *Taeger*, CR 1996, 257 (261 ff.); *Taschner*, in: *Taschner/Frietsch* Rn. 22 f.; v. *Westphalen*, in: *Produkthaftungshandbuch*, § 73 Rn. 36 ff.; *Wagner*, in: *MüKo*, § 2 ProdHaftG Rn. 15.

dass bei Standardsoftware die Informationen auf Datenträgern verkörpert seien und bei wertender Betrachtung der Waren- und nicht der Dienstleistungscharakter im Vordergrund stehe.⁴⁹⁷ Zum Teil wird dies auch dann vertreten, wenn die Daten ohne Beschaffung eines Datenträgers online auf bereits vorhandene Speichermedien übertragen werden.⁴⁹⁸ Allerdings ist diese Ansicht selbst unter den Befürwortern der Anwendung des Produkthaftungsgesetzes auf Software stark umstritten.⁴⁹⁹

Noch unklarer ist die Beurteilung der Rechtslage im Hinblick auf Individualsoftware. Zwar tendiert auch hier wohl die Mehrheit der Literatur zur Anwendbarkeit des Produkthaftungsgesetzes,⁵⁰⁰ allerdings scheint dies noch fragwürdiger als im Hinblick auf Standardsoftware, da hier eher der Dienstleistungscharakter im Vordergrund steht.⁵⁰¹ Dafür spricht, dass der Gesetzgeber gerade eine Haftung für Waren und nicht für Dienstleistungen, die Informationen regelmäßig aber darstellen, schaffen wollte.⁵⁰² Dieser Ansicht wird wiederum entgegengehalten, dass es gerade nicht darauf ankomme, ob ein Produkt individuell oder massenhaft gefertigt wird.⁵⁰³

Schließlich ist auch die Produkthaftung im Hinblick auf sonstige Daten äußerst unklar. Selbst in klassischen Fällen, in denen es um die Haftung für falsche Informationen in Druckerzeugnissen geht, herrscht in Literatur und Rechtsprechung keine Einigkeit.⁵⁰⁴ Gestützt auf den Dienstleistungscharakter von Informationen wird eine teleologische Reduktion des § 2 ProdHaftG in Erwägung gezogen.⁵⁰⁵ Problematisch ist in diesem Kontext, dass danach nur für gedruckte bzw. anderweitig verkörperte Informationen gehaftet werden müsste, nicht aber für unkörperliche Informationen.⁵⁰⁶

Fraglich ist auch, ob nicht die hinter dem ProdHaftG stehenden Überlegungen grundlegend überdacht werden müssen. Wie bereits erwähnt, wird anhand des Wortlauts des § 2 ProdHaftG argumentiert, dass eine Haftung nur für ein verkörpertes Produkt bestehen kann. Begründet wird dies damit, dass das Vertrauen in eine verkörperte Sache typischerweise höher sei als in eine nicht verkörperte Sache.⁵⁰⁷ Ob diese Annahme in Zeiten, in denen die meisten Programme, bis hin zu ganzen Betriebssystemen, auch oder nur noch online bezogen werden können, noch tragbar ist, erscheint zumindest zweifelhaft. Der

⁴⁹⁷ Wagner, in: MüKo, § 2 ProdHaftG Rn. 15.

⁴⁹⁸ Cahn, NJW 1996, 2899 (2904); Spindler, MMR 1998, 119 (121); Taeger, CR 1996, 257 (261 f.); v. Westphalen, in: Produkthaftungshandbuch, § 73 Rn. 40; Wagner, in: MüKo, § 2 ProdHaftG Rn. 16.

⁴⁹⁹ Beckmann/Müller, MMR 1999, 14 (17f.); Meyer, Instruktionshaftung, S. 86; Taschner, in: Taschner/Frietsch Rn. 22.

⁵⁰⁰ Deutsch, NJW 1992, 73 (76); Hoeren, CR 1992, 606; Hollmann, DB 1985, 2389 (2390); Lehmann, NJW 1992, 1721 (1724); Taeger, CR 1996, 257 (262f); v. Westphalen, in: Produkthaftungshandbuch, Bd. 2, § 61 Rn. 40.

⁵⁰¹ Engel, CR 1986, 702 (705f.); Kort, CR 1990, 171 (175); Wagner, in: MüKo, § 2 ProdHaftG Rn. 15; vgl. auch Mincke, jurPC 1990, 406.

⁵⁰² Cahn, NJW 1996, 2899 (2901 ff.); Meyer, ZUM 1997, 26 (28); Schiemann, in: Erman Rn. 2; v. Westphalen, in: Produkthaftungshandbuch, § 73 Rn. 14; Wagner, in: MüKo, § 2 ProdHaftG Rn. 13.

⁵⁰³ Oechsler, in: Staudinger, § 2 ProdHaftG Rn. 69.

⁵⁰⁴ Für eine Darstellung des Streitstandes, siehe Oechsler, in: Staudinger, § 2 ProdHaftG Rn. 74 ff.

⁵⁰⁵ Vgl. Wagner, in: MüKo, § 2 ProdHaftG Rn. 13.

⁵⁰⁶ Oechsler, in: Staudinger, § 2 ProdHaftG Rn. 78 ff.

⁵⁰⁷ Müller-Terpitz, in: Kröger/Gimmy, Handbuch zum Internetrecht.

Benutzer vertraut schließlich in gleichem Maße auf die Fehlerfreiheit des Programms, wenn er es auf dem Wege der Online-Übermittlung bezieht.⁵⁰⁸

Schließlich muss überlegt werden, ob durch die Übermittlung von fehlerhaften Daten nicht auch schützenswertes Vertrauen des Benutzers verletzt wird. Bis jetzt scheiterte eine Haftung gem. § 1 Abs. 1 ProdHaftG daran, dass übermittelte Informationen dem Benutzer nicht dauerhaft zur Verfügung stehen und somit kein Produkt i.S.d. ProdHaftG darstellen.⁵⁰⁹ Vor dem Hintergrund der immens gestiegenen Bedeutung des Datenaustausches seit der Einführung des Produkthaftungsgesetzes und aufgrund der Überlegung, dass Daten heutzutage rund um die Uhr zur Verfügung stehen können, muss zumindest in Erwägung gezogen werden, ob das Vertrauen des Benutzers auf die Fehlerfreiheit der übermittelten Informationen unter gewissen Voraussetzungen zu schützen ist.

Insgesamt ist also festzustellen, dass auf diesem Gebiet eine Vielzahl offener und umstrittener Fragen besteht, die der Klärung bedürfen. Die technologischen Entwicklungen der letzten Jahre müssen dabei ebenso berücksichtigt werden wie das geänderte Benutzerverhalten im Internet.

K. Roboterrecht

Der Mensch ist fehlbar, die Maschine (meist) nicht.⁵¹⁰ Es gibt kaum einen Lebensraum, den moderne Maschinenwesen nicht erobert haben.⁵¹¹ Roboter sind mittlerweile ein permanenter Bestandteil des menschlichen Alltags und können das menschliche Leben in jeder Hinsicht erheblich beeinflussen.⁵¹² Dennoch ist die Robotik ein in der Rechtswissenschaft bislang wenig beleuchtetes Feld. Dieses rechtliche Schattendasein verwundert, da der Einsatz von Robotern zahlreiche – bislang ungeklärte – rechtliche und ethische Fragen aufwirft. Ihre Beantwortung ist zweifellos eine große Aufgabe und wird die zukünftige Entwicklung der Robotik prägen. Diese Zukunftsperspektive wurde auf nationaler Ebene bislang vor allem von Prof. Hilgendorf mit seiner *Forschungsstelle RobotRecht* an der Universität Würzburg erkannt, die wertvolle Vorarbeiten zu Rechtsfragen der Robotik liefert. Das weltweite Forschungsbedürfnis im Bereich der Robotik zeigt sich schon daran, dass sich mittlerweile auch die *Stanford Law School* mit dem Projekt „Robotics And The Law“ der Klärung von roboterrechtlichen Fragen widmet. Auch die *Internationale Vereinigung für Rechts- und Sozialphilosophie* veranstaltete unter dem Titel „Recht, Wissenschaft und Technik“ eigens einen Kongress zu Fragen der Robotik.

Die theoretischen und praktischen juristischen Probleme, die sich aus dem Einsatz von Robotern ergeben, sind erheblich und betreffen verschiedenste Rechtsgebiete.⁵¹³ Im Kern

⁵⁰⁸ Spindler, MMR 1998, 119 (121).

⁵⁰⁹ Spindler, MMR 1998, 119 (121).

⁵¹⁰ Ernst, CRAktuell 3/2013, 27.

⁵¹¹ Lindinger, FAZ v. 2.1.2013, S. N 2.

⁵¹² Ernst, CRAktuell 3/2013, 27 (28).

⁵¹³ Vgl. auch zu den folgenden Beispielen: Beck, Roboter, Cyborgs und das Recht – von der Fiktion zur Realität, in: Spranger (Hrsg.), Aktuelle Herausforderungen der Life Sciences, S. 102 ff.; Beck, JR 2009, 225 (226 ff.).

gilt es, die strafrechtliche und zivilrechtliche Dimension des Einsatzes von Robotern sowie Kausalitätsfragen zu durchdringen.⁵¹⁴ Mangels spezieller Regelungen für die Herstellung und Verwendung von Robotern im deutschen Recht stellt sich die Frage, inwiefern bestehende Normen auf derartige Sachverhalte anwendbar sind.⁵¹⁵ Damit einher geht die Frage, ob der Gesetzgeber den neuen rechtlichen Herausforderungen durch Schaffung gesetzlicher Regelungen zu begegnen hat.

In zivilrechtlicher Hinsicht kommt vor allem dem Produkthaftungsrecht eine große Bedeutung zu. Daneben stellt sich die Frage, inwiefern sich beim Einsatz von Robotern Schadensersatzansprüche aus den §§ 280, 823 BGB ergeben können. Die Beweiserbringung für Kausalität und Verschulden wird in diesen Fällen regelmäßig schwer fallen. Auch ist noch vollkommen ungeklärt, wie Roboter zu versichern sind. In öffentlich-rechtlicher Hinsicht berührt der Einsatz von Robotern verschiedenste Rechtsgebiete, wie etwa das Polizei- und Ordnungsrecht oder das Umweltrecht. In Bezug auf autonome Roboter ist sogar zu überlegen, ob diese grundrechtsberechtigt und -verpflichtet sind. Die Analyse der strafrechtlichen Verantwortlichkeit von Hersteller und Verwender eines Roboters wird sich zunächst auf die §§ 222, 229 StGB beschränken. Wie im Zivilrecht wird der Einsatz von Robotern auch im Strafrecht zu einem erschwerten Kausalitätsnachweis führen. Zu diskutieren ist ferner, ob den Verwender eines Roboters eine Garantenpflicht gegenüber anderen Beteiligten trifft oder ob ein Roboter als Fahrzeug im Sinne der §§ 315 ff. StGB klassifiziert werden kann. Am Beispiel von Hirnschrittmachern und künstlichen Gelenken zeigt sich, dass Maschine und Mensch immer mehr verschmelzen. Auch dies eröffnet weitere Forschungsperspektiven. Es stellt sich die Frage, ob der Träger eines Hirnschrittmachers, der das Gehirn künstlich beeinflusst, noch in vollem Umfang für sein Handeln verantwortlich gemacht werden kann. Die Beantwortung dieser Frage ist Voraussetzung zur Beurteilung der Schuldfähigkeit des Trägers, § 20 f. StGB.

Die exemplarisch aufgezeigten Fragestellungen verdeutlichen, dass die Rechtswissenschaft durch die Robotik vor neue Herausforderungen gestellt wird. Die sich daraus ergebenden Rechtsfragen erscheinen beinahe grenzenlos und können keinesfalls mehr als reine Zukunftsmusik bezeichnet werden. Ihre Beantwortung ist von enormer Bedeutung – gerade auch, weil die fortlaufende technische Entwicklung in Zukunft sicher neue Herausforderungen für Gesetzgeber und Gerichte bereithalten wird.⁵¹⁶

L. Datenmarktplätze

Die kommerzielle Nutzung und Verwertung von Daten als verselbständigttem Informationskapital ist ein zentrales Merkmal der Informationsgesellschaft.⁵¹⁷ Ein für die Beteiligten gewinnbringender Informationsaustausch findet inzwischen auf sog.

⁵¹⁴ Ernst, CRaktuell 3/2013, 27 (27f.).

⁵¹⁵ Beck, Roboter, Cyborgs und das Recht – von der Fiktion zur Realität, in: Spranger (Hrsg.), Aktuelle Herausforderungen der Life Sciences, S. 102; Beck, JR 2009, 225 (226).

⁵¹⁶ Ernst, CRaktuell 3/2013, 27 (28).

⁵¹⁷ Moos, MMR 2006, 718.

Datenmarktplätzen statt. Hierbei handelt es sich um Handelsplattformen, auf denen Anbieter Daten aus verschiedenen Quellen, insbesondere sozialen Netzwerken, zusammentragen und anschließend an Unternehmen veräußern. Datenmarktplätze erfordern folglich ein besonderes Big-Data-Management.

Dies führt zu der Frage auf welche Weise Datenmarktplätze rechtskonform ausgestaltet werden können. Wie bereits allgemein für das Big-Data-Management beschrieben, ergeben sich vor allem datenschutz- und urheberrechtliche Probleme. Insofern kann zunächst auf obige Ausführungen verwiesen werden, die sich auf Datenmarktplätze im Speziellen übertragen lassen. Die Frage nach der rechtlichen Zulässigkeit von Datenmarktplätzen soll jedoch zum Anlass genommen werden einige Sonderfragen, die zuvor noch nicht ausführlich behandelt wurden, aufzugreifen. Die rechtliche Beurteilung erfordert eine Differenzierung zwischen der Datenerhebung und -speicherung einerseits und der Weitergabe der Daten andererseits.

Eine Frage, die nicht unmittelbar die rechtliche Zulässigkeit eines Datenmarktplatzes betrifft, ergibt sich im Hinblick auf die Qualität der erfassten und weitergegebenen Daten. Der Erfolg eines Datenmarktplatzes wird wesentlich dadurch beeinflusst, ob die an Dritte weitergegebenen Daten auch den gewünschten Anforderungen entsprechen. Erwirbt ein Unternehmen auf einem Datenmarktplatz etwa zielgruppenorientiert Daten, so müssen die jeweiligen Daten auch tatsächlich aus der jeweiligen Zielgruppe stammen. Es stellt sich also die Frage der Datenqualität und einer etwaigen (Informations-)Haftung. Für dieses Folgeproblem wird auf obige Ausführungen verwiesen.⁵¹⁸

1. Datenschutzrecht

Für Datenmarktplätze von entscheidender Bedeutung ist die Ermächtigungsgrundlage für den Datenumgang. Hierfür kommen sowohl die Einwilligung des Betroffenen in Betracht als auch gesetzliche Ermächtigungen. Grundsätzlich wird aus Praktikabilitätserwägungen eher auf die Ermächtigungsgrundlagen des BDSG als auf die Einwilligung des Betroffenen zurückgegriffen.⁵¹⁹ Charakteristisch für Datenmarktplätze ist, dass der Datenumgang nicht zu eigenen Zwecken erfolgt, weshalb § 28 BDSG ausscheidet. Entscheidend ist vielmehr § 29 BDSG⁵²⁰, der sich vornehmlich mit der Weitervermittlung von Daten befasst.⁵²¹

Gemäß § 29 Abs. 1 Nr. 2 BDSG kann der Umgang mit Daten aus allgemein zugänglichen Quellen unter erleichterten Bedingungen erfolgen.⁵²² Von besonderer Bedeutung ist hierbei vor allem die Frage nach der allgemeinen Zugänglichkeit der Daten. Um in Übereinstimmung mit dem Datenschutzrecht zu handeln, müssen sämtliche Daten aus derartigen Quellen stammen.⁵²³ Grundsätzlich werden darunter alle Quellen verstanden, „die sich sowohl ihrer

⁵¹⁸ Siehe dazu bereits „F.“.

⁵¹⁹ Moos, MMR 2006, 718 (718).

⁵²⁰ Moos, MMR 2006, 718 (718 f.).

⁵²¹ Simitis, in: Simitis, § 29 Rn. 196.

⁵²² Gola/Schomerus, § 29 Rn. 19.

⁵²³ Moos, MMR 2006, 718 (719).

technischen Ausgestaltung als auch ihrer Zielsetzung nach dazu eignen, einem individuell nicht bestimmmbaren Personenkreis Informationen zu vermitteln“⁵²⁴. Hierunter fallen neben Rundfunk und Presse weitere allgemein zugängliche Veröffentlichungen.⁵²⁵ Hierzu zählen auch öffentliche Register, soweit nicht weitere Voraussetzungen, wie z.B. ein besonderes Interesse, für die Einsichtnahme notwendig sind.⁵²⁶ Internetseiten sind öffentlich zugängliche Quellen in diesem Sinne.⁵²⁷ Zur schwierigen Frage der Einstufung von Daten aus sozialen Netzwerken wie z.B. Facebook wird auf die obigen Ausführungen verwiesen.

Die Daten können auch nur mittelbar aus öffentlich zugänglichen Quellen stammen. Solange die Daten an sich generell allgemein zugänglich sind, ist auch eine direkte Gewinnung aus nicht allgemein zugänglichen Quellen gestattet.⁵²⁸ Dies betrifft folglich auch Daten von Datenmarktplätzen, soweit diese Daten aus derartigen Quellen weitergeben werden.

Der Datenumgang ist nach § 29 Abs. 1 Nr. 2 BDSG jedoch ausgeschlossen, wenn das schutzwürdige Interesse des Betroffenen an dem Ausschluss des Datenumgangs offensichtlich überwiegt. Dazu hat eine Abwägung der Interessen der verantwortlichen Stelle mit den Interessen des Betroffenen stattzufinden. Der Datenumgang ist ausschließlich dann untersagt, wenn das entgegenstehende Interesse des Betroffenen offensichtlich überwiegt. Offensichtlich heißt in diesem Fall, es muss „eindeutig erkennbar sein, dass der Betroffene ein solches Gegeninteresse hat und zum anderen, dass es gegenüber dem Interesse der verantwortlichen Stelle überwiegt.“⁵²⁹

Insbesondere der unantastbare Kernbereich privater Lebensgestaltung darf nicht berührt werden. Das ist jedenfalls dort nicht der Fall, wo an Verhalten in der Öffentlichkeit angeknüpft wird.⁵³⁰ Im Ergebnis ist der Datenumgang solange zulässig, wie er nicht Daten aus dem Kernbereich privater Lebensgestaltung erfasst oder es zur Bildung eines Persönlichkeitsprofils infolge der „Aggregation personenbezogener Daten und deren Verknüpfung“⁵³¹ kommt. Eine konkrete Abgrenzung zwischen zulässiger Datensammlung und unzulässigem Persönlichkeitsprofil hat sich in der Rechtsprechung allerdings noch nicht herausgebildet.⁵³²

Gemäß § 29 Abs. 1 Nr. 1 BDSG ist der Datenumgang zulässig, wenn kein eventuell entgegenstehendes schutzwürdiges Interesse des Betroffenen besteht, das einen Datenumgang ausschließt. Hierbei ist vor allem auf Art, Inhalt und Aussagekraft der jeweiligen Daten abzustellen.⁵³³ Sofern die verantwortliche Stelle konkrete Anhaltspunkte

⁵²⁴ Simitis, in: Simitis, § 28 Rn. 151.

⁵²⁵ Gola/Schomerus, § 28 Rn. 32.

⁵²⁶ Gola/Schomerus, § 28 Rn. 32.

⁵²⁷ Weichert, ZD 2013, 251 (257).

⁵²⁸ Moos, MMR 2006, 718 (719).

⁵²⁹ Gola/Schomerus, § 29 Rn. 19.

⁵³⁰ Moos, MMR 2006, 718 (720).

⁵³¹ Moos, MMR 2006, 718 (721).

⁵³² Moos, MMR 2006, 718 (721).

⁵³³ BGH, NJW 1984, 1889 (1890).

für ein entgegenstehendes Interesse hat, so ist der Datenumgang ausgeschlossen. Im Gegensatz zu § 29 Abs. 1 Nr. 2 BDSG müssen die Interessen des Betroffenen nicht offensichtlich überwiegen, sondern nur bestehen.⁵³⁴

Die Übermittlung der gewonnenen Daten ist im Rahmen der Zwecke nach Absatz 1 zulässig, wenn der Dritte, dem die Daten übermittelt werden, ein berechtigtes Interesse an ihrer Kenntnis glaubhaft dargelegt hat und kein Grund zur Annahme besteht, dass der Betroffene ein schutzwürdiges Interesse am Ausschluss der Übermittlung hat, § 29 Abs. 2 BDSG.

Als berechtigtes Interesse an der Kenntnis in diesem Zusammenhang kommt jedes tatsächliche, wirtschaftliche oder ideelle Interesse in Betracht.⁵³⁵ Der freie Vertrieb von Daten ist aus dem gleichen Grunde jedoch nicht gestattet, denn die Darlegung des berechtigten Interesses muss für jeden Fall einzeln glaubhaft dargelegt werden.⁵³⁶

Ferner darf kein schutzwürdiges Interesse des Betroffenen an dem Ausschluss der Übermittlung bestehen. Soweit die Daten aus öffentlich zugänglichen Quellen stammen, wird man regelmäßig vom Fehlen eines solchen Interesses ausgehen können.⁵³⁷

2. Urheberrecht

Im Hinblick auf urheberrechtliche Fragestellungen in Zusammenhang mit Datenmarktplätzen kann im Wesentlichen auf die allgemeinen Ausführungen zum Urheberrecht verwiesen werden. Rechtlich relevant ist auch hier vor allem das Recht des Datenbankherstellers, § 87a ff. UrhG. Für das Web 2.0, aus dem Daten für Datenmarktplätze gewonnen werden, spielen Datenbanken eine besondere Rolle.⁵³⁸

Ein Datenmarktplatz selbst wird regelmäßig durch das Leistungsschutzrecht des Datenbankherstellers geschützt. Die erfassten Daten werden inhaltlich nach logischen Gesichtspunkten geordnet, um sie zielgruppengerecht weitergeben zu können. Ein Datenmarktplatz ist auf Vollständigkeit ausgerichtet. Die gesammelten Daten sind einzeln mit Hilfe elektronischer Mittel oder auf andere Weise zugänglich.

Für den Betreiber eines Datenmarktplatzes ist allerdings die Frage relevant, ob er durch die Datenbeschaffung in fremde Urheberrechte eingreift. Einzelne wissenschaftliche Daten sind regelmäßig nicht urheberrechtlich geschützt.⁵³⁹ Nutzergenerierten Inhalten (UGC) in sozialen Netzwerken kann unter Umständen urheberrechtlicher Schutz, zum Beispiel als Sprachwerk (§ 2 Abs. 1 Nr. 1 UrhG) oder Lichtbildwerk (§§ 2 Abs. 1 Nr. 5, 72 UrhG), zukommen.⁵⁴⁰ In den Fällen, in denen einzelne Beiträge urheberrechtlich geschützt sind, stellt deren Kopie und Speicherung in einem Datenmarktplatz einen unzulässigen Eingriff in die Verwertungsrechte

⁵³⁴ Moos, MMR 2006, 718 (722).

⁵³⁵ Gola/Schomerus, § 29 Rn. 19; Moos, MMR 2006, 718 (722).

⁵³⁶ Moos, MMR 2006, 718 (722).

⁵³⁷ Moos, MMR 2006, 718 (723).

⁵³⁸ Herrmann/Dehijelles, K&R 2009, 23.

⁵³⁹ Schulze, in: Dreier/Schulze, § 2 Rn. 130; Spindler/Hillegeist, in: Handbuch Forschungsdatenmanagement, S. 63 f.

⁵⁴⁰ Vgl. hierzu Reinemann/Remmertz, ZUM 2012, 216; Solmecke/Wahlers, ZD 2012, 550 (553 f.).

des Urhebers dar. Für die urheberrechtliche Beurteilung kommt es mithin auf den konkreten Einzelfall an. In der Regel werden Datenmarktplätze aber vor allem soziodemografische Daten, Interessen und Kaufabsichten der Verbraucher sammeln, die selbst nicht urheberrechtlich geschützt sind.

Soziale Netzwerke im Internet, aus denen die Daten in den meisten Fällen gewonnen werden, stellen allerdings Datenbanken im Sinne einer Sammlung von systematisch bzw. methodisch angeordneten Daten dar.⁵⁴¹ Möglicherweise verletzt die Auswertung der Plattformen deshalb die Rechte der jeweiligen Plattformbetreiber. Urheber eines solchen Datenbankwerks (§ 4 Abs. 2 UrhG) bzw. Hersteller der Datenbank (§ 87 a Abs. 2 UrhG) ist der Anbieter eines solchen Dienstes, der die Gestaltung und Verwaltung der Datenbank organisiert.⁵⁴² Dies gilt, obwohl der Großteil des Inhalts in sozialen Netzwerken von Dritten beigesteuert wird, die diese Inhalte beliebig hinzufügen und ändern können, wenn nur die Datenbankstruktur erhalten bleibt.⁵⁴³

Ein Schutz als Datenbankwerk wird jedoch in der Regel für Kommunikationsplattformen wie Facebook, Twitter und Youtube, die für Datenmarktplätze ausgewertet werden, abzulehnen sein. Dieser setzt nach § 4 Abs. 2 UrhG voraus, dass die Datenbank hinsichtlich der Auswahl und Anordnung eine gewisse Schöpfungshöhe aufzeigt. Gerade bei wirtschaftlich geprägten Datenbanken, wie Bewertungsplattformen, aber auch bei sozialen Netzwerken wird die erforderliche Schöpfungshöhe nicht gegeben sein.⁵⁴⁴ Soweit die Plattformbetreiber über die Aufbereitung und Gliederung der von den Nutzern eingespeisten Daten selbst entscheiden, kann jedoch sehr wohl eine persönliche schöpferische Leistung vorliegen, die zu einem Schutz nach § 4 Abs. 2 UrhG führt.⁵⁴⁵ Es muss daher im Einzelfall entschieden werden, ob die entsprechende Plattform die erforderliche Schöpfungshöhe aufweist oder nicht.

Scheidet ein Schutz als Datenbankwerk gem. § 4 Abs. 2 UrhG aus, steht dem Betreiber einer Plattform jedoch das Leistungsschutzrecht des Datenbankherstellers zu.⁵⁴⁶ Die Erstellung einer Datenbank, z.B. in Form eines sozialen Netzwerks, ist mit hohem Aufwand und erheblichen Investitionen verbunden.⁵⁴⁷ Nach § 87 b UrhG hat der Hersteller der Datenbank das ausschließliche Recht, die Datenbank insgesamt oder zu einem wesentlichen Teil nach Art und Umfang zu vervielfältigen, zu verbreiten und öffentlich wiederzugeben. Die Entnahme und Verwertung von Daten aus einer Datenbank durch die Betreiber eines Datenmarktplatzes könnten daher gegen die Rechte des Datenbankherstellers verstossen.

Voraussetzung dafür ist, dass es sich bei der vom Datenmarktplatzbetreiber entnommenen Datenmenge um einen wesentlichen Teil der Datenbank handelt. Wann von einer

⁵⁴¹ Koch, in: Loewenheim, Handbuch des Urheberrechts, § 77 Rn. 5a; Herrmann/Dehjßelles, K&R 2009, 23.

⁵⁴² Koch, in: Loewenheim, Handbuch des Urheberrechts, § 77 Rn. 5a.

⁵⁴³ Koch, in: Loewenheim, Handbuch des Urheberrechts, § 77 Rn. 5a.

⁵⁴⁴ Herrmann/Dehjßelles, K&R 2009, 23; Reinemann/Remmertz, ZUM 2012, 216 (220).

⁵⁴⁵ Reinemann/Remmertz, ZUM 2012, 216 (220).

⁵⁴⁶ So für Ebay *LG Berlin*, Urt. v. 22.12.2005 - 16 O 743/05 = ZUM 2006, 343 (344 f.); Herrmann/Dehjßelles, K&R 2009, 23; Reinemann/Remmertz, ZUM 2012, 216 (220).

⁵⁴⁷ Herrmann/Dehjßelles, K&R 2009, 23 f.; Zieger/Smirra, MMR 2013, 418 (420).

Übernahme wesentlicher Teile auszugehen ist, hängt von den Umständen des Einzelfalles ab. Hierbei muss das Ziel des Leistungsschutzrechts, einen effektiven Investitionsschutz für den Datenbankhersteller zu gewähren, Berücksichtigung finden.⁵⁴⁸ Die Wesentlichkeit muss sowohl nach quantitativen als auch nach qualitativen Gesichtspunkten bewertet werden. Eine quantitative Wesentlichkeit lässt sich durch das Verhältnis des entnommenen und/oder weiterverwendeten Datenvolumens im Vergleich zum Gesamtvolumen der geschützten Datenbank bestimmen.⁵⁴⁹ Wesentlichkeit wäre jedenfalls bei einer Datenentnahme von mehr als 50 % des gesamten Datenbankinhalts anzunehmen.⁵⁵⁰ Werden nur einzelne Teile einer Datenbank abgefragt und liegt dieser Anteil bei unter 10 %, wird keine Wesentlichkeit gegeben sein.⁵⁵¹ Ein rein quantitativer Maßstab könnte allerdings Datenbankhersteller mit einem hohen Datenvolumen benachteiligen. Über die Frage, wann ein wesentlicher Teil einer Datenbank betroffen ist, muss auch eine Bewertung der Daten aus qualitativer Sicht entscheiden. Wesentlichkeit soll danach vorliegen, wenn sich in dem übernommenen Teil der Datenbank selbst eine wesentliche Investition niedergeschlagen hat.⁵⁵² Ob es sich bei den von den Betreibern der Datenmarktplätze entnommenen und genutzten Daten um einen wesentlichen Teil im Sinne des § 87 b UrhG handelt, muss also im Einzelfall entschieden werden. Werten die Betreiber von Datenmarktplätzen allerdings soziale Netzwerke mit hohen Nutzerzahlen und enormen Content wie etwa Facebook aus, werden die gesammelten Daten in aller Regel keinen wesentlichen Teil der Datenbank ausmachen.

Verneint man jedoch die Übernahme eines wesentlichen Teils der Datenbank, greift möglicherweise der Auffangtatbestand des Art. 87 b Abs. 1 S. 2 UrhG. Der Entnahme eines wesentlichen Teils der Datenbank steht die wiederholte und systematische Entnahme von unwesentlichen Teilen der Datenbank gleich, sofern diese Handlungen einer normalen Auswertung der Datenbank zuwiderlaufen oder die berechtigten Interessen des Datenbankherstellers unzumutbar beeinträchtigen. Datenmarktplätze gewinnen ihre Daten zweifellos durch die mehrfache und wiederholte Abfrage bestimmter Internetplattformen. Auch bei dieser systematischen und wiederholten Übernahme von Datenbankteilen muss allerdings in der Summe die Wesentlichkeitsschwelle des Art. 87 Abs. 1 S. 1 UrhG erreicht werden. Dies wurde vom *OLG Köln*⁵⁵³ bestätigt, mit der Konsequenz, dass der Auswertung von Web 2.0-Datenbanken keine wesentlichen Bedenken gegenüberstehen. Folge ist jedoch auch, dass der Umgehungsschutz – eigentlich Sinn und Zweck des Art. 87 Abs. 1 S. 2 UrhG –

⁵⁴⁸ Siehe dazu bereits „C. 2. b)“; *Herrmann/Dehjelles*, K&R 2009, 23 (25).

⁵⁴⁹ *Thum*, in: *Wandtke/Bullinger*, § 87b Rn.12 ff.

⁵⁵⁰ *Thum*, in: *Wandtke/Bullinger*, § 87b Rn.15.

⁵⁵¹ *BGH*, Urt. v. 22.6.2011 - I ZR 159/10 = MMR 2012, 544; *BGH*, Urt. v. 1.12.2010 - I ZR 196/08 = MMR 2011, 676.

⁵⁵² *EuGH*, Urt. v. 9. 11. 2004 - C-203/02 The British Horseracing Board Ltd u.a./William Hill Organization Ltd = GRUR 2005, 244, *Herrmann/Dehjelles*, K&R 2009, 23 (25).

⁵⁵³ *OLG Köln*, Urt. v. 14.11.2008 - 6 U 57/08 = MMR 2009, 191 (192) mit Verweis auf *EuGH*, Urt. v. 9. 11. 2004 - C-203/02 The British Horseracing Board Ltd u.a./William Hill Organization Ltd = GRUR 2005, 244 (251); so auch *Dreier/Schulze*, § 87 b Rn. 11; a.A. noch *LG Berlin*, Urt. v. 22.12.2005 - 16 O 743/05 = ZUM 2006, 343 (345).

leerläuft, so dass der Vorschrift kein tatsächlicher eigener Anwendungsbereich neben Art. 87 Abs. 1 S. 1 UrhG verbleibt.⁵⁵⁴

M. Medizindaten

Vom Einsatz von Big Data-Technologien im Gesundheitswesen werden in Zukunft besonders spürbare Effekte erwartet. Daten fallen in der Medizin etwa bei einem MRT, EKG oder EEG in großen Mengen an. Die konkreten Anwendungsbeispiele werden in der Folge dargestellt und damit verbundene Rechtsfragen problematisiert.

1. Konkrete Anwendungsbeispiele

Die Auswertung großer Datenmengen ermöglicht eine personalisierte und somit effektivere Patientenbehandlung. Die Sammlung verschiedener Krankheiten und Patienteninformationen in vernetzten Datenbanken sind zum Beispiel in der Krebsforschung gewinnbringend. Deren Auswertung ermöglicht eine schnelle Krankheitsanalyse und Diagnose, sowie eine individuelle, auf den jeweiligen Patienten zugeschnittene Therapie-Entscheidung. Gespeicherte Daten gestatten so dem Arzt einen jederzeitigen Zugriff auf die Krankheitshistorie eines Patienten, die Medikamentenwirksamkeit und Therapieerfolge.

Auch die Auswertung von Handydaten oder Daten in sozialen Netzwerken kann für medizinische Zwecke genutzt werden. Sie ermöglichen die Ausbreitung von Krankheiten nachzuvollziehen und Krankheitswellen frühzeitig zu erkennen. Auf diese Weise können sich Ärzte und Pharmakonzerne frühzeitig auf bestimmte Krankheiten einstellen und entsprechend, etwa durch die Aufstellung bestimmter Impfpläne, reagieren.

Durch die Analyse von Herzschlägen eines Babys können Forscher feststellen, wie hoch die Wahrscheinlichkeit ist, dass ein Kind später erkrankt. Eine solche Diagnose lässt sich nicht durch eine normale Arztuntersuchung, sondern allein durch datengestützte Software stellen. Diese basiert auf Langzeitstudien, in denen Muster erkannt wurden, die einen Herzinfarkt auslösen können. Die frühzeitige Erkennung dieser Risiken ermöglicht rechtzeitige vorbeugende Maßnahmen. Hierzu können beispielsweise auch Sensoren am oder im Körper eines Patienten eingesetzt werden, die Signale senden, sobald sich der körperliche Zustand eines Patienten ändert.

2. Rechtliche Rahmenbedingungen

Der Einsatz von Big Data im Medizinsektor muss den geltenden Rechtsrahmen wahren. Die durch die Forschungsfreiheit (Art. 5 Abs. 3 GG) geschützte medizinische Forschung steht dabei in einem Zielkonflikt mit dem Recht auf informationelle Selbstbestimmung (Art. 2 Abs. 1, Art. 1 Abs. 1 GG).⁵⁵⁵

a) Allgemeiner Rechtsrahmen

Insbesondere berührt der Umgang mit Big Data auch im Gesundheitswesen die Aspekte von Datenschutz und Datensicherheit. Es ergeben sich Konflikte mit den Grundsätzen der

⁵⁵⁴ Herrmann/Dehijelles, K&R 2009, 23 (25).

⁵⁵⁵ Dickmann/Rienhoff, Langzeitarchivierung von Forschungsdaten, Kapitel 12 Medizin, S. 247.

Datensparsamkeit und Zweckbindung. Daten erhalten durch die Weiterentwicklung medizinischer Forschung eine über den ursprünglichen Erhebungszweck hinausreichende Bedeutung.⁵⁵⁶ Gerade medizinische Datenbanken profitieren davon, dass zu verschiedensten Zwecken erhobene Daten miteinander vernetzt werden, um so neue Erkenntnisse zu gewinnen. Die Zulässigkeit des Datenumgangs könnte durch eine Einwilligung der Patienten begründet werden. Eine solche legitimierende Einwilligung wird allerdings regelmäßig nicht vorliegen. Dies gilt etwa in den Fällen, in denen Nachrichten in sozialen Netzwerken, wie z.B. Twitter, ausgewertet werden.

Um den Datenumgang zu legitimieren, erscheinen Anonymisierung und Pseudonymisierung der Daten als geeigneter Weg. Als oberstes Gebot muss jedoch gelten, dass eine Re-Identifikation, etwa durch die Kombination verschiedener Merkmale, nicht möglich sein darf. In der Vergangenheit ist es entgegen aller Vorhersagen immer wieder gelungen, Daten zu entschlüsseln. Dies könnte möglicherweise durch eine zweistufige Pseudonymisierung verhindert werden. Ein weiterer zu beachtender Aspekt ist die ärztliche Schweigepflicht. Verstößt ein Arzt gegen diese Pflicht und gibt sensible Patientendaten weiter, so verwirklicht er gegebenenfalls den Straftatbestand des § 203 StGB.

In diesem Zusammenhang ist die Handhabung der durch privatärztliche Verrechnungsstellen verarbeiteten Daten umstritten.⁵⁵⁷ Ärzte unterliegen einer besonderen Schweigepflicht, die bereits durch die Abtretung einer Honorarforderung an eine gewerbliche Verrechnungsstelle verletzt und somit der Straftatbestand des § 203 StGB erfüllt sein kann, wie der *BGH* bereits 1991 entschieden hat.⁵⁵⁸ Der Hinweis auf die Schweigepflicht der Verrechnungsstellen hilft dabei nicht weiter. Die ärztliche Schweigepflicht gilt grundsätzlich auch gegenüber anderen Schweigepflichtigen.⁵⁵⁹ Die Honorarabrechnung enthält oftmals besondere Arten personenbezogener Daten i.S.d. § 3 Abs. 9 BDSG und ist somit besonders zu schützen.⁵⁶⁰ Der Umgang mit Patientendaten muss in einem für den Patienten vorhersehbaren und überschaubaren Bereich erfolgen.⁵⁶¹ Das war bisher jedenfalls in aller Regel die Praxis des Arztes mitsamt dessen Mitarbeitern.⁵⁶² Es bedarf nunmehr weiterer Forschung, ob an diesem Grundsatz festzuhalten ist. In diesem Zusammenhang ist auch differenziert über ein „Selbstbestimmungsrecht“⁵⁶³ für Patientendaten nachzudenken.

Bei der Anwendung von Big Data-Tools für das Gesundheitswesen sind die obigen allgemeinen rechtlichen Rahmenbedingungen zu beachten. Es gilt, wirksame Anonymisierungs- und Pseudonymisierungsmethoden zu entwickeln. Außerdem muss das Big Data-Tool einen unbefugten Datenzugriff Dritter verhindern. Im Gesundheitswesen

⁵⁵⁶ Dickmann/Rienhoff, Langzeitarchivierung von Forschungsdaten, Kapitel 12 Medizin, S. 229.

⁵⁵⁷ Giesen, NStZ 2012, 122 (122).

⁵⁵⁸ BGH, NJW 1991, 2955 (2955); Gramberg-Danielsen/Kern, NJW 1998, 2708 (2708).

⁵⁵⁹ Gramberg-Danielsen/Kern, NJW 1998, 2708 (2709).

⁵⁶⁰ BGH, NJW 1991, 2955 (2957); Giesen, NStZ 2012, 122 (122).

⁵⁶¹ BGH, NJW 1991, 2955 (2957).

⁵⁶² Giesen, NStZ 2012, 122 (122).

⁵⁶³ Giesen, NStZ 2012, 122 (123).

könnten sich aus Verbrauchersicht dazu Datenbriefe, die einen Überblick über jeweils gespeicherte personenbezogene Daten ermöglichen, als hilfreich erweisen.

b) Big Data und Medizinrecht

Neben den allgemeinen Rechtsvorschriften sind im Bereich des Gesundheitswesens auch gesetzliche Spezialvorschriften zu beachten, die sich insbesondere auf Speicher- und Aufbewahrungspflichten beziehen. An den dadurch entstehenden großen Datenmengen besteht für die medizinische Forschung ein besonderes Interesse.

Grundsätzlich ergibt sich aus dem Haftungsregime der §§ 195, 199 und 852 BGB eine Aufbewahrungspflicht „von 30 Jahren für stationäre und zehn Jahren für ambulante Behandlungen.“⁵⁶⁴ Sofern und soweit es für eine weitere Behandlung vonnöten ist, kommt auch eine über die 30-Jahres-Grenze hinausgehende Aufbewahrung in Betracht, etwa im Falle chronischer Erkrankungen.⁵⁶⁵

Daneben finden sich Konkretisierungen dieser Aufbewahrungspflichten in verschiedenen Gesetzen. Gemäß § 28 Abs. 3 der Verordnung über den Schutz vor Schäden durch Röntgenstrahlen (Röntgenverordnung - RÖV) sind Aufzeichnungen über Röntgenbehandlungen 30 Jahre lang nach der letzten Behandlung aufzubewahren. Röntgenbilder und die Aufzeichnungen über Röntgenuntersuchungen sind zehn Jahre lang nach der letzten Untersuchung aufzubewahren. Röntgenbilder und die Aufzeichnungen von Röntgenuntersuchungen einer Person, die das 18. Lebensjahr noch nicht vollendet hat, sind bis zur Vollendung des 28. Lebensjahres dieser Person aufzubewahren.

Ähnliche Sondervorschriften gibt es auch für den Bereich der Wehrverwaltung. § 5 der Verordnung über die Führung der Personalakten der Soldaten und ehemaligen Soldaten (Personalverordnung Soldaten - SPersAV) bestimmt, dass Gesundheitsunterlagen bis zum Ablauf des 90. Lebensjahres des Soldaten aufzubewahren und danach zu vernichten sind.

Auf europäischer Ebene finden sich relevante Bestimmungen unter anderem in den Richtlinien 2002/98/EG und 2005/61/EG, die im Zusammenhang mit menschlichem Blut und Blutbestandteilen eine Mindestaufbewahrungspflicht von 15 bis 30 Jahren für Krankenhausblutdepots und Blutpendeeinrichtungen vorsehen. Die Richtlinien 2004/23/EG und 2006/86/EG sehen im Zusammenhang mit menschlichem Gewebe und Zellen ähnliche Dokumentations- und Aufbewahrungspflichten für einen Zeitraum von zehn bis 30 Jahren vor.

⁵⁶⁴ Dickmann/Rienhoff, Langzeitarchivierung von Forschungsdaten, Kapitel 12 Medizin, S. 231.

⁵⁶⁵ Dickmann/Rienhoff, Langzeitarchivierung von Forschungsdaten, Kapitel 12 Medizin, S. 231.