

Zusammenfassung

Die zunehmende Vernetzung im Internet der Dinge und Dienste sowie die Verwendung von fortgeschrittener 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 erheblich Wettbewerbsvorteile zu erwarten. 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.

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: Datenschutz, Urheber- und Vertragsrecht

und die **Ausbildung von Fachkräften.**

Die Entwicklung konvergenter IKT im vorwettbewerblichen Bereich für das „Big Data Management“ ist ein folgerichtiger und besonders förderungswürdiger Schwerpunkt des Bundesministeriums für Wirtschaft und Technologie (BMWi).

Die vorliegende Kurzfassung ist eine Zusammenfassung einer ausführlichen Studie, die dieses Zusammenspiel bewertet, um durch das Verwalten und Analysieren von „Big Data“ zu „Smart Data“ zu gelangen. Die vollständige Studie steht unter der folgenden Adresse http://www.dima.tu-berlin.de/menue/research/big_data_management_report/ zum Download bereit.

Einleitung: 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 Analyse, 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,
- wobei gleichzeitig die Entscheidungszeitfenster, in denen Analyseergebnisse bereitgestellt werden müssen, immer kürzer werden,
- eine Vielzahl von unterschiedlichen Datenquellen (z.B. Zeitreihen, Tabellen, Textdokumente, Bilder, Audio- und Videodatenströme) in die Datenanalyse einbezogen werden,
- und aufgrund der Unschärfe von einigen Datenquellen (z.B. Sensoren mit fixer Genauigkeit) bzw. von Informationsextraktionsverfahren und Integrationsverfahren müssen Systeme und Analysten mit wahrscheinlichkeitsbasierten Modellen und Konfidenzen im Rahmen der Analyse von Big Data umgehen.

Zudem werden neue deklarative Sprachen für Spezifikation und automatische Optimierung und Parallelisierung von komplexen Datenanalyseprogrammen (inkl. neuer statistischer und mathematischer Algorithmen) benötigt um das Datenvolumen, die Verarbeitungsgeschwindigkeit, die unterschiedlichen Datenformate und die Vertrauenswürdigkeit der Daten in den Griff zu bekommen. Big Data besitzt ausserdem eine neue Analysekomplexität, die sich daran zeigt, 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.

Diese Anforderungen werden zu einem Paradigmenwechsel bei Datenanalyse-sprachen, Datenanalyse-systemen und auch Datenanalysealgorithmen führen und völlig neuartige Anwendungen ermöglichen.

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

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 Forschungsgemeinschaft 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, Hyper der TU München, die Forschungen zu Hadoop++ und HAIL an der Universität Saarbrücken). 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.

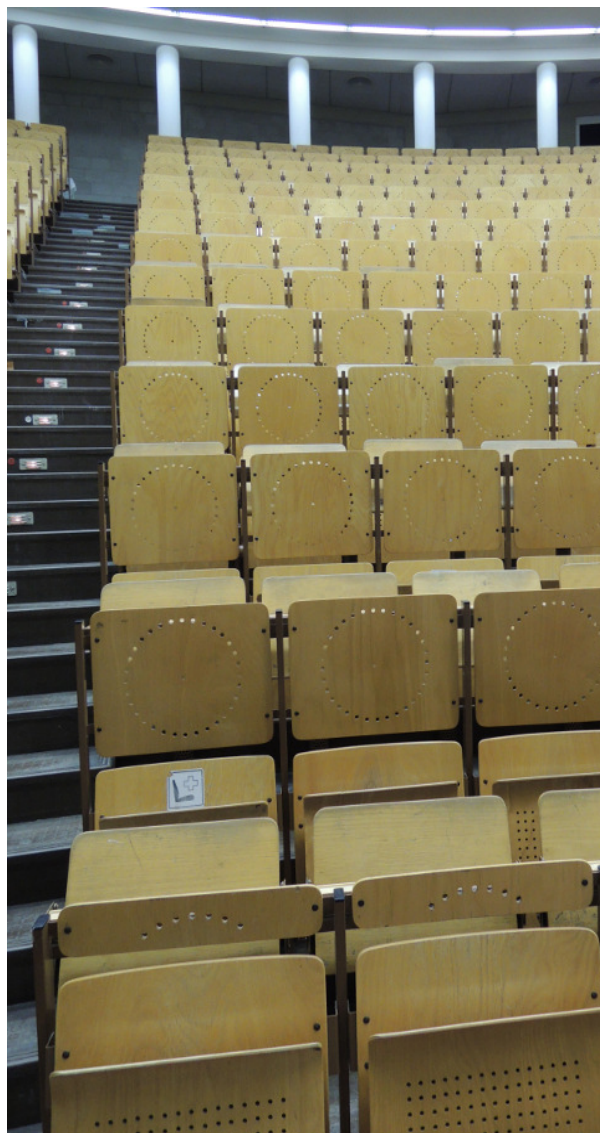
Unter den Herausforderern, die sich Chancen und Marktanteile in dem gerade entstehenden Big-Data Markt versprechen, finden sich dabei auch eine Vielzahl an deutschen Technologieunternehmen, neben Großunternehmen wie z.B. SAP mit HANA und der Software AG mit Terracotta auch viele High-Tech-Startups wie z.B. ParStream, Exasol. Um diese Unternehmen in ihrer Produkt- und Markteroberungsstrategie zu unterstützen, ist es wichtig, ein Technologietransfer- und Innovationsklima zu schaffen, das es deutschen Unternehmen, insbesondere 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.

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, Lehre und 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. Das neuartige Berufsbild **Data Scientist** 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 Anwendungsdomänen flankiert werden.



Rechtliche Aspekte von Big Data

Die technische Entwicklung der letzten Jahre hat die verfügbaren Datenmengen erheblich vergrößert. Der Einsatz der 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. Der Datenumgang wirft allerdings vielfältige rechtliche Fragen auf, die vor allem die Rechtsdisziplinen des Datenschutz-, Urheber- und des Vertragsrechts betreffen. Daneben hat sich in der Rechtswissenschaft eine Diskussion entwickelt, deren Ergebnis gravierende Folgen für die Big-Data-Branche hat. Es geht um die Frage, ob Daten eigentumsfähig sind und, falls dies so ist, wem das Eigentum daran zusteht.

Eigentum an Daten

Ein Eigentumsrecht an Daten mag mit Blick auf die bestehenden Differenzierungen hinsichtlich Sacheigentum, Urheberrecht, Schutz des Datenbankherstellers und Datenschutz zunächst einmal als überflüssiges theoretisches Konstrukt erscheinen. Allerdings haben sich Daten zu einem erheblichen Wirtschaftsfaktor entwickelt. Sie stellen ein Handelsgut dar und haben einen eigenständigen fühlbaren Wert. Die Entwicklung eines absoluten Schutzregimes erscheint vor diesem Hintergrund als angemessen.

Dogmatisch bereitet das Konzept vom Dateneigentum jedoch Schwierigkeiten. Eigentumsfähigkeit setzt voraus, dass Daten

nach der Rechtsordnung einem Rechtssubjekt zugeordnet werden können, sodass diesem ein umfassendes Herrschaftsrecht daran zusteht. Eine Zuordnung nach §§ 90 ff. BGB erscheint ausgeschlossen. Das Datum an sich ist keine körperliche Sache im Sinne des § 90 BGB, sondern vielmehr körperlich abhängig. Einer Bewertung des Datums als wesentlicher Bestandteil des Datenträgers nach § 93 BGB steht entgegen, dass dann „Dateneigentum“ und Sacheigentum nicht auseinanderfallen könnten. Genau dies muss aber möglich sein, um Schäden in vernetzten Datenbeständen sachgerecht ausgleichen zu können.

Beispielhaft sei auf die Situation verwiesen, in der der wirtschaftliche Schaden nicht beim Eigentümer des Datenträgers eintritt, weil nicht er sondern eine dritte Person die Daten wirtschaftlich nutzt. Während die Rechtsprechung vor rund 25 Jahren einen Fall des Datenverlusts auf einem eigenen Datenträger durch Rückgriff auf das Eigentumsrecht an diesem lösen konnte, ist diese Konstruktion aufgrund technischer Neuerungen wie Speichern von Daten in der „Cloud“ nicht mehr zeitgemäß. Das in dem damaligen Zusammenhang diskutierte „Recht am eigenen Datenbestand“ als sonstiges Recht im Sinne des § 823 BGB könnte zwar in solchen Situationen seine Berechtigung haben, begegnet jedoch dogmatischen Bedenken. Als ausschließliches Recht müsste das „Recht am eigenen Datenbestand“ nämlich eine den in § 823 Abs. 1 BGB genannten Rechten ähnliche Ausschlussfunktion bieten. Wem dieses Aus-

schlussrecht zustehen soll, ist nicht ersichtlich, sodass die Lösung als Hilfskonstruktion bewertet werden muss, die mehr Probleme aufwirft als sie löst.

Eine eigentumsrechtliche Zuordnung von Daten zu einem Rechtssubjekt kann auch nicht über das Datenschutzrecht erfolgen. Dieses schafft nur eine rechtliche Verantwortlichkeit für Daten, darf jedoch nicht dahin verstanden werden, dass der Betroffene Ausschließlichkeitsrechte an den einzelnen Datensätzen im Sinne eines Eigentums besitzt. Gleiches gilt für eine in Betracht kommende Lösung über das urheberrechtliche sui-generis-Recht an einer Datenbank. Dieses stellt einen Investitionsschutz dar, der vor wirtschaftlicher Ausbeutung fremder Leistung schützt, nicht jedoch darüber hinausgehend Daten einer Person zuordnen soll.

Die Zuordnungsproblematik kann allerdings durch einen Rückgriff auf das Strafrecht gelöst werden. § 303a StGB schützt explizit Daten. Die Zuordnung des Schutzgutes zu einem Rechtssubjekt erfolgt anhand des Skripturaktes, also nach dem technischen Herstellungsprozess. Eine Übertragung dieser Konzeption auf das Zivilrecht ermöglicht also eine eindeutige Zuordnung und damit die Möglichkeit der Etablierung eines „Dateneigentums“.

Es lässt sich somit festhalten, dass die Herleitung eines Eigentums am Datum dogmatisch möglich ist. Ob sich das Institut des Dateneigentums durchsetzt, ist jedoch noch nicht abzusehen, da die Diskussion erst am Anfang steht. Dennoch muss diesem Thema gerade aus Sicht der Big-Data-Unternehmen hohe Aufmerksamkeit gewidmet werden. Denn je nach Ausgang der Debatte könnten unabhängig vom Datenschutz- oder Urheberrecht Ansprüche durch den „Dateneigentümer“ geltend gemacht werden, die den Umgang mit den Datensätzen beeinflussen würden. Zudem hätte eine Klärung der Frage nach dem Eigentum an Daten auch grundlegende Auswirkungen auf insolvenzrechtliche Fragestellungen.

Datenschutzrecht

In der aktuellen wissenschaftlichen Diskussion wird vehement auf das Spannungsfeld von Big Data und dem Datenschutzrecht hingewiesen. Mit Blick auf die datenschutzrechtlichen Grundsätze wie Zweckbindung, Transparenz, Direkterhebung, Datenvermeidung bzw. Datensparsamkeit und dem Verbot mit Erlaubnisvorbehalt wird klar, dass Big Data mit dem Schutz personenbezogener Daten kollidieren kann. Um Datenschutzverstöße zu vermeiden, müssen Big-Data-Unternehmen deshalb vielfältige Problemfelder bei ihrer Entscheidungsfindung beachten.

Ein Problem liegt bereits in der Frage nach der Anwendbarkeit des deutschen Datenschutzrechts. Das grundsätzlich geltende Territorialprinzip bereitet als Anknüpfungspunkt Schwierigkeiten, wenn Daten weltweit auf verschiedenen Standorten verteilt sind, die Datenmengen flüchtig sind und sich der einmal ermittelte Standort binnen kurzer Zeit ändern kann. Big-Data-Unternehmen müssen deshalb womöglich eine Vielzahl unterschiedlicher Rechtsordnungen beim Umgang mit Daten berücksichtigen.

Weiterhin erfordert der Umgang mit personenbezogenen Daten nach deutschem Datenschutzrecht grundsätzlich die Einwilligung des Betroffenen oder eine Legitimation durch gesetzliche Erlaubnisgründe. Dieser Grundsatz des Verbots mit Erlaubnisvorbehalt wird zunehmend als untauglich kritisiert, da angesichts der Ubiquität der Datenverarbeitung in Smartphones etc. potentiell jedermann zum Datenverarbeiter wird. Der Schutz der informationellen Selbstbestimmung wird im Ausgangspunkt über alles gestellt, um in der Folge jedoch viele und zum Teil sehr weitgehende Erlaubnistatbestände festzulegen. Es wird deshalb vorgeschlagen, im Rahmen einer Modernisierung des Datenschutzrechts das Konzept des Verbots mit Erlaubnisvorbehalt zu überdenken. Big-Data-Unternehmen könnten auf diesem Wege von der komplexen Aufgabe der Ermittlung ei-

nes „passenden“ Legitimationstatbestandes entlastet werden.

Im Zusammenhang damit stellt sich auch die generelle Frage nach der Wirksamkeit des Instituts der Einwilligung als Legitimationstatbestand. Zum einen bestehen Bedenken hinsichtlich der Freiwilligkeit der Einwilligung, wenn diese faktisch Gegenleistung für eine „kostenlose“ Dienstleistung wird und sich somit zum Handelsgut entwickelt. Zum anderen ist zu klären, wie eine Einwilligung wirksam eingeholt werden kann und ob diese in ihrer Zeitdauer möglicherweise begrenzt werden kann.

Insgesamt zeigen die hier angedeuteten sowie eine Vielzahl weiterer Probleme, dass das Datenschutzrecht ein Hemmnis für Big-Data-Unternehmen darstellt. Dies ist auf der einen Seite gewollt, um personenbezogene Daten effektiv zu schützen. Auf der anderen Seite ist jedoch auch zu sehen, dass die Grundgedanken des Datenschutzrechts noch aus einer Zeit stammen, in der das Phänomen Big Data nicht existierte. Deshalb sind Modernisierungsvorschläge notwendig, in deren Rahmen die gegenläufigen Interessen in einen angemessenen Ausgleich gebracht werden müssen.

Urheberrecht

Der Umgang mit Big Data wirft auch Fragen aus urheberrechtlicher Sicht auf. Eine allgemeingültige und für jede Big Data-Lösung einheitliche Begutachtung urheberrechtlicher Problematiken ist nicht möglich. Die konkreten Probleme ergeben sich stets aus der Ausgestaltung des Verfahrens im Einzelfall. Zu bedenken ist jedoch, dass der Schutz des Urheberrechts in zwei Richtungen wirken kann. Zum einen ist es möglich, dass sich das datenverarbeitende Unternehmen selbst auf das Urheberrecht – nämlich den Sui-generis-Datenbankschutz – berufen kann. Zum anderen könnte ein Datenumgang jedoch auch gegen Urheberrechte bzw. verwandte Schutzrechte Dritter verstoßen. Gerade letzteres muss bei der Entscheidung über eine konkrete Datenverarbeitung vom Big-Data-Unternehmen berücksichtigt werden.

Wie beim Datenschutzrecht stellt sich auch in diesem Zusammenhang die Frage nach der Anwendbarkeit der deutschen Rechtsordnung. Grundsätzlich wird bei in Deutschland genutzten Big-Data-Lösungen wegen des Territorialitätsprinzips deutsches Recht anwendbar sein. Bei Internet-Sachverhalten, die sich nicht einem bestimmten Schutzterritorium zuordnen lassen, entstehen jedoch kollisionsrechtliche Fragestellungen, die noch nicht vollumfassend geklärt sind. Für Big-Data-Unternehmen bedeutet dies wiederum eine unsichere rechtliche Lage, da es möglich ist, dass eine Vielzahl verschiedener Rechtsordnungen Anwendung findet.

Urheberrechtlicher Schutz wird nur gewährt, wenn eine ausreichende Gestaltungshöhe des Werkes vorliegt. Einzelne Daten haben regelmäßig nicht die dafür erforderliche individuelle Prägung, sodass Urheberrechtsverstöße beim Umgang mit einzelnen Daten zunächst nicht zu befürchten sind. Etwas anderes kann jedoch gelten, wenn nutzergenerierte Inhalte z.B. aus sozialen Netzwerken ausgewertet werden sollen. Solche Daten können als Lichtbilder (§ 72 UrhG), Lichtbildwerke (§ 2 Abs. 1 Nr. 5 UrhG) oder Sprachwerke (§ 2 Abs. 1 Nr. 1 UrhG) geschützt sein. Die Auswertung würde dementsprechend das Vervielfältigungsrecht und das Recht der öffentlichen Zugänglichmachung berühren.

Die Zulässigkeit des Datenumgangs hängt aus urheberrechtlicher Sicht davon ab, ob eine gesetzliche Schrankenregelung zugunsten der datenverarbeitenden Unternehmen greift oder der Rechtsinhaber dieser Handlung zustimmt. Die gesetzlichen Schranken werden den Umgang mit großen Datenmengen jedoch regelmäßig nicht legitimieren. Diese Feststellung wird zum Anlass genommen, darüber zu diskutieren, ob das UrhG um eine weitere Schranke ergänzt werden muss, um den neuen Dimensionen des Datenverkehrs gerecht zu werden. Ohne „passende“ Schranke bleibt den Big-Data-Unternehmen nur der Weg über die Zustimmung des Rechteinhabers. Dabei bestehen jedoch erhebliche praktische Probleme. Die neuen Technologien des

Internetzeitalters führen dazu, dass fast zwangsläufig in zahllose fremde Ausschließlichkeitsrechte eingegriffen wird, so dass eine große Zahl vertraglicher Vereinbarungen getroffen werden müsste. Um dieses Problem zu umgehen, wird die Figur einer faktischen Einwilligung diskutiert. Ob eine solche Konstruktion allerdings tatsächlich dazu geeignet ist, die Handlung zu legitimieren, ist noch ungeklärt. Grundsätzlich ist also die Tendenz zu erkennen, dass im Falle des Eingreifens von Urheberrechten ein Umgang mit Big Data nur unter erschwerten Bedingungen möglich ist.

Diese Feststellung lässt sich auch durch die Tatsache unterstreichen, dass daneben ein weiteres urheberrechtliches Schutzrecht den Umgang mit Big Data hemmen kann. Werden die Daten nämlich aus fremden Datenbanken bezogen, so kann der Sui-generis-Datenbankschutz gem. §§ 87a ff. UrhG den Handlungen des Big-Data-Unternehmens entgegenstehen. Der Sui-generis-Schutz ist unabhängig von einer individuellen Schöpfungshöhe. Vielmehr wird eine Datenbank geschützt, wenn eine wesentliche Investition für ihre Erstellung getätigt wurde. Wann die erforderliche Wesentlichkeit vorliegt, ist eine Frage des Einzelfalls. Daraus resultiert für Big-Data-Unternehmen also wiederum eine Unsicherheit, die vor der Vornahme einer unternehmerischen Entscheidung ausgeräumt werden muss. Denn sollte die erforderliche Schwelle überschritten sein, so dürfen ohne Zustimmung des Datenbankherstellers nur unwesentliche Teile der Datenbank für den eigenen Datenumgang genutzt werden. Unterhalb der gesetzlich vorgesehenen Schwelle können die Daten hingegen ohne urheberrechtliche Beschränkung verwertet werden.

Alles in allem belegen also bereits die hier nur kurz skizzierten Probleme, dass das Urheberrecht mit seinen ungeklärten Rechtsfragen sowie teilweise modernisierungsbedürftigen Regelungen ein Hemmnis für den Big-Data-Sektor sein kann. Jedenfalls darf aus der Sicht der Unternehmen der Fokus auf das Datenschutzrecht nicht den Blick für ebenfalls einschlägige urheberrechtliche Regelungen trüben.

Vertragsrechtliche und haftungsrechtliche Fragen

Weiteren Unklarheiten sind Big-Data-Unternehmen im Hinblick auf die Vertragsgestaltung sowie Haftungsfragen ausgesetzt. Die Vertragsgestaltung ist aufgrund der vielen technischen Fragen sowie der Berührungspunkte zum Datenschutz- und Urheberrecht kompliziert. Pauschalisierte Aussagen lassen sich nur schwerlich treffen. Nichtsdestotrotz lässt sich feststellen, dass Regelungen in allgemeinen Geschäftsbedingungen, die eine Haftung für Datenverlust bzw. -beschädigung ausschließen oder beschränken sollen, Gefahr laufen, im Rahmen der gerichtlichen AGB-Kontrolle als unwirksam bewertet zu werden. Daneben ergeben sich ganz allgemein Fragen bezüglich der Haftung von Big-Data-Unternehmen. Eine solche setzt ein Fehlverhalten des jeweiligen Unternehmens voraus. Dieses könnte unter Umständen in der Übermittlung einer fehlerhaften Information zu sehen sein. Nach welchen Kriterien jedoch bestimmt werden kann, wann eine solche vorliegt, ist bislang nicht abschließend geklärt.

Fazit

Die Entwicklung von Big Data in Deutschland wird maßgeblich vom geltenden Recht beeinflusst. Von besonderer Bedeutung wird die weitere Diskussion um ein etwaiges „Dateneigentum“ sein. Zudem ergeben sich insbesondere aus dem Datenschutzrecht sowie dem Urheberrecht einige Hürden, die den Umgang mit großen Datenmengen erschweren. Im Rahmen von Diskussionen über Modernisierungen dieser Rechtsgebiete sollte überprüft werden, inwieweit diese Regelungen noch zeitgemäß sind und an welcher Stelle Änderungen sachgerecht wären, um den wirtschaftlichen Interessen datenverarbeitender Unternehmen entgegenzukommen ohne jedoch die schutzwürdigen Interessen betroffener Personen zu vernachlässigen. Solange dieser Prozess andauert müssen sich Big-Data-Unternehmen in jedem konkreten Einzelfall damit auseinandersetzen, ob ihre Big-Data-Lösung mit dem Urheber- und Datenschutzrecht im Einklang steht.

Innovationspotential von Big Data

In der Studie wurde das Innovationspotential von Big Data sowohl in ausgewählten Branchen als auch branchenübergreifend untersucht. Im folgenden Abschnitt werden die Ergebnisse zusammengefasst wiedergegeben.

Ergebnisse der Untersuchung aktueller Studien. Für die Studie wurde ein Überblick aktueller Studien zu den Innovationspotenzialen von Big Data erarbeitet. Für diese Untersuchung wurden nur Big Data Studien ausgewählt, die Aussagen zu Deutschland beinhalten, die eine quantitative Datengrundlage haben und vergleichbare Aussagen enthalten, so dass diese zueinander verknüpft werden konnten. Es wurden insgesamt 9 Studien ausgewertet. Die Stichprobengröße der untersuchten Studien (BARC, BITKOM, Computing Research, Experton Group, Fraunhofer IAIS, Interxion, TATA Consultancy Services, TNS Infratest) ist mit $n > 4492$ umfangreich.

Die Kernaussagen der Studien können wie folgt zusammengefasst 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.

Ergebnisse der im Rahmen der Studie durchgeführten empirischen Studie. Ergänzend zu obiger Analyse der existierenden Studien zu Innovationspotenzialen von Big Data wurde für die Studie eine eigene empirische Untersuchung durchgeführt. Das Ziel dieser Untersuchung war es, die bestehenden Erkenntnisse zu validieren und zu ergänzen.

Die Untersuchung wurde im 2. und 3. Quartal des Jahres 2013 durchgeführt. Die gesamte Anzahl von Fragebogenrückläufen betrug 185.

Persönlicher Hintergrund der Studienteilnehmer. Abbildung 5.1 zeigt den Tätigkeitshintergrund der Teilnehmer an der Studie, rund 16 Prozent der Teilnehmer der Studie waren Entscheider. Die Teilnehmer der Studie hatten in der Mehrheit einen (Wirtschafts-) Informatik Ausbildungshintergrund (siehe Abbildung 5.2).

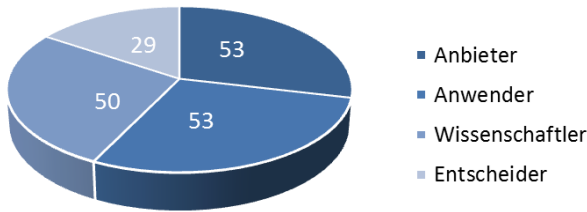


Abbildung 5.1: Tätigkeitsbereiche der Teilnehmer der Studie

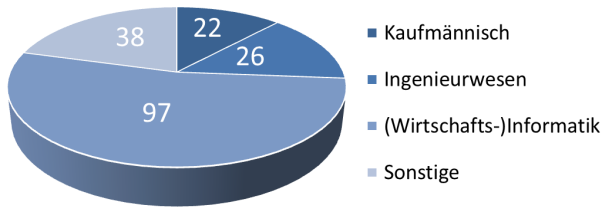


Abbildung 5.2: Ausbildungshintergrund der Teilnehmer der Studie (Anzahl Antworten=183)

Die meisten Teilnehmer waren zwischen 20 - 54 Jahre alt (164 Antworten; 16 Teilnehmer älter als 54) und rechneten sich mehrheitlich der Branche *Informationstechnologie* zu.

Von den 90 Teilnehmern, die sich einem Unternehmen zuordneten, rechneten sich 40 Teilnehmer der Unternehmensgröße kleines und mittelständisches Unternehmen (KMU) zu. 47 Teilnehmer arbeiteten in Unternehmen mit mehr als 1000 Angestellten, 23 in Unternehmen mit 250-1000 Mitarbeitern und 39 mit 250 oder weniger Mitarbeitern (insgesamt 109 Antworten). Bei 38 Unternehmen betrug der Jahresumsatz über 250 Mio. Euro, bei 66 Unternehmen lag der Jahresumsatz darunter, davon hatten 24 Unternehmen einen Jahresumsatz von 5-50 Millionen.

In den folgenden Abschnitten werden die Ergebnisse der Studie näher betrachtet.

Big Data Projekte sind noch in einer sehr frühen Phase. 40 Prozent der Entscheider geben an, in der Informationsphase über Big Data-Technologien zu sein (siehe Abbildung 5.3). Lediglich 8 Prozent der Entscheider haben sich bereits mit der Umsetzung beschäftigt. Ein Viertel der Entscheider gab an, dass sie bereits die Ent-

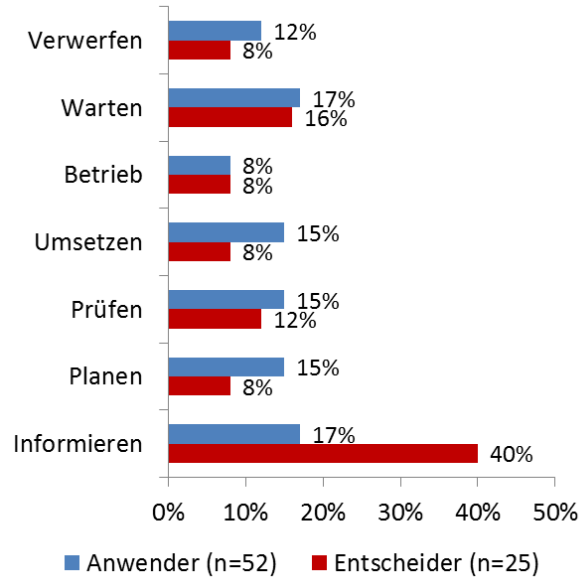


Abbildung 5.3: Status von Big Data Projekten

wicklung 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.

Big Data wird in den kommenden fünf Jahren wichtig... Im Rahmen der Studie wurden Anwender und Entscheider bzgl. ihrer Einschätzung der Wichtigkeit des Themas Big Data in der Praxis gefragt. Während sich 44 Prozent der Anwender (n=50) dem Thema Big Data im Jahre 2013 eine *sehr wichtige* bzw. *eher wichtige* Rolle zumessen, waren es bei den Entscheidern (n=24) 29 Prozent. Für das Jahr 2014 lag die Einschätzung bei den Anwendern bereits bei 61 Prozent und bei den Entscheidern bei 43 Prozent. Die größte Wichtigkeit wurde dem Thema Big Data jedoch in den nächsten fünf Jahren zugemessen (Anwender 78 Prozent und Entscheider 60 Prozent).

...und wird sich in weniger als 10 Jahren etabliert haben Der Großteil der Befragten geht davon aus, dass sich Big Da-

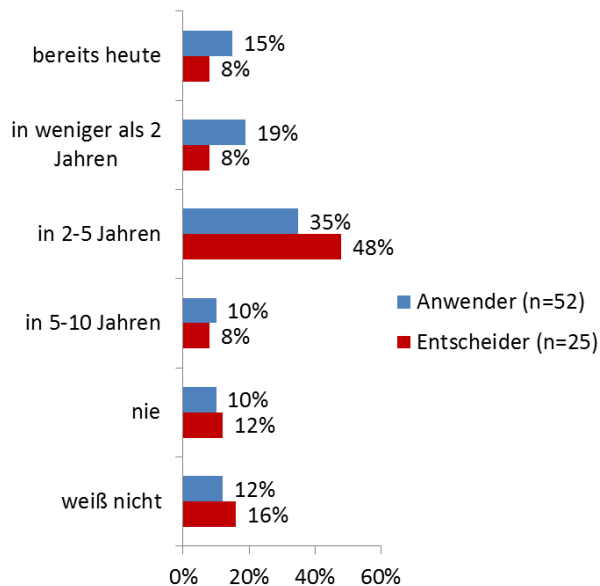


Abbildung 5.4: Wann wird Big Data wettbewerbsentscheidend?

ta in weniger als 10 Jahren in ihrer Branche flächendeckend durchsetzen wird. Diese Einschätzung wird auch von Anbietern und Entscheidern geteilt, wie Abbildung 5.4 zeigt.

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

Messbarer Wertbeitrag von Big Data? 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. 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.

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 (siehe Abbildung 5.5).

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.

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,

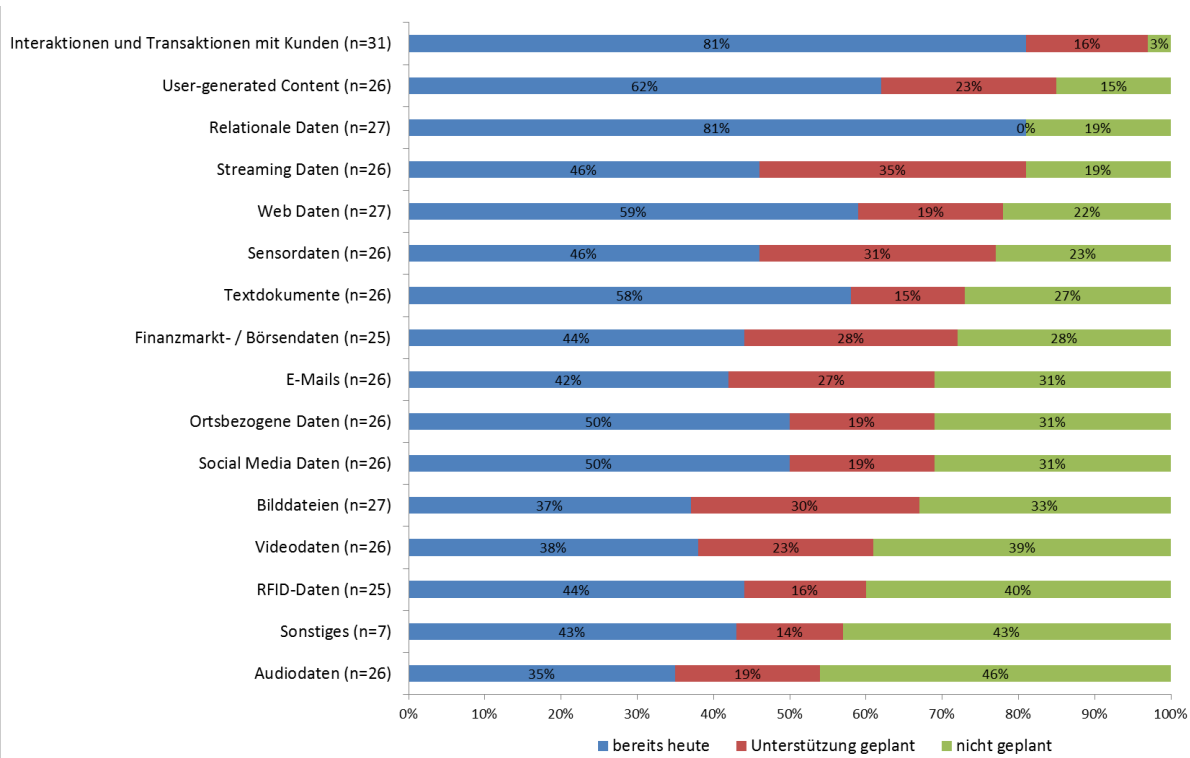


Abbildung 5.5: Angebot an Big Data Werkzeugen anhand der verbreiteten Datentypen (Anbieterperspektive)

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.

Zusammenfassung der Ergebnisse der Studie.

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 werden.
2. Deutsche Anbieter von Big Data Technologie 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 sehr hohe Herausforderungen. Dabei stehen neben einer Anpassung der rechtlichen Rahmenbedingungen an den Stand der Technik vor allem die Aufklärung und ein effektives Erwartungsmanagement.

5. Eine weitere wichtige Herausforderung ist die Verfügbarkeit entsprechend qualifizierten Personals. Geeignete Ausbildungs- und Weiterbildungsangebote sind hier notwendig, um den Unternehmen das entsprechende Knowhow 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.

Branchenübergreifendes und branchenspezifisches Innovationspotential.

Im Folgenden werden aus den untersuchten Studien und der eigenen Studie zunächst allgemeine branchenübergreifende Innovationspotenziale diskutiert, bevor auf die als herausragend klassifizierten einzelnen Branchen eingegangen wird.

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. Dabei lassen sich aus der Literatur unterschiedliche Stärken und Schwächen, sowie Chancen und Risiken ableiten.

In Abbildung 5.6 sind die Ergebnisse der SWOT-Analyse zu Big Data im Allgemeinen dargestellt.

Die folgenden Branchen wurden als besonders herausragend in Bezug auf Big Data Innovationspotential identifiziert:

- öffentlicher Sektor,
- Industrie 4.0,
- Gesundheitssektor/Lifesciences,
- Marktforschung,(Social-) Media und Entertainment,
- Mobilitätsdienstleistungen,
- Energiewirtschaft sowie
- Risikomanagement und Versicherungswesen.

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. 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.

| SWOT-Analyse | | Interne Analyse | |
|-----------------|---------|--|--|
| | | Stärken | Schwächen |
| Externe Analyse | Chancen | <ul style="list-style-type: none"> • Schnellere und genauere Auswertungen • Bessere strategische Entscheidungen • Verbesserte Steuerung operativer Prozesse • Bessere Auswertung mittels Business Intelligence • Kostenreduktion • Flexiblere Dienstleistungen, zielorientiertere Marketingaktionen • Bessere Informationen zum innerbetrieblichen Risikomanagement • Besseres Verständnis des Marktes • Verbesserter Kundenservice | <ul style="list-style-type: none"> • Monitoring führt zur Erhöhung der internen Kontrolle • Reduktion von Datensätze ließ Informationen unbeobachtet, Big Data schafft verlässlichere Daten • Kundenindividuelle Lösungen (gläserner Kunde) |
| | Risiken | <ul style="list-style-type: none"> • Distributive Daten • Betrugserkennung (engl. Fraud Detection) • Unstrukturierte informelle Kommunikation • Fehlendes fachliches und technischen Wissen | <ul style="list-style-type: none"> • Datenschutz • Datensicherheit • Gesellschaftliche Akzeptanz • Ethische Gründe • Fehlende überzeugende Einsatzszenarien • Technische Probleme • Kosten |

Abbildung 5.6: SWOT-Analyse zur Potentialbewertung von Big Data im Allgemeinen.

Analyse von Big Data & Big Data Technologien

Als Resultat der Studie und den Anforderungen der Industrie konnten die folgenden vier Kernanforderungen für das Management von Big Data abgeleitet werden:

1. Umgang mit großen, heterogenen Datenmengen
2. Komplexe Datenanalysealgorithmen
3. Interaktive, oftmals visuell unterstützte Datenanalyse
4. Nachvollziehbare Datenanalyse

Um diese Herausforderungen zu meistern, müssen skalierbare, einfach zu bedienende Datenanalyzesysteme und neue Algorithmen bzw. Paradigmen zur Datenanalyse entwickelt werden, die die verschiedenen Aspekte und Anforderungen gleichzeitig adressieren. Diese Herausforderungen werden von existierenden Datenmanagementsystemen bisher nicht erreicht.

Analyse von Big Data

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. Datenanalyzesysteme 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).

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 Minings, des Text Minings, des Graph Minings, Video Minings und der visuellen Analyse. So muss das Datenanalyzesystem 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 SQLDatenbanksystemen noch den Big Data Lösungen (Hadoop, Pig, Hive, Storm, Lambda Architek-

tur, etc.) realisiert. Hierdurch ergibt sich hohes Innovations und Marktpotential für die Entwicklung und Kommerzialisierung moderner Datenanalyzesysteme, die relationale Datenverarbeitung mit Algorithmen insbesondere des maschinellen Lernens und der Statistik vereint.

Technische Herausforderungen von Big Data Technologien

Der interaktive und iterative Datenanalyseprozess 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 (engl. user-defined functions, kurz UDFs) ausreichend schnell verarbeiten können.

Programmiermodelle für die breite Masse von Analysten. Es gibt neben Hadoop eine Vielzahl von interessanten Forschungsarbeiten zur massiv-parallelen Verarbeitung von datengetriebenen iterativen Algorithmen. 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. 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.

Iterative Datenstromverarbeitung.

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. Batchbasierte Systeme wie Map/Reduce und Spark 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 oder spezielle Graphverarbeitungssysteme zum Beispiel GraphLab und Google Pregel diese Eigenschaft ausnutzen und reduzieren den Berechnungsaufwand mit jeder Iteration.

Verarbeitung von Benutzerdefinierten Funktionen mit geringer Latenz

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, SCOPE, Stratosphere und Spark sind einige Systeme, die ausdrücklichere, 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 Funk-

tionen wie Match, CoGroup und Cross in Stratosphere.

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 oder mittels statischer Codeanalyse zu erhalten.

Stand der Technik

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.

Big Data als Chance für deutsche Technologieanbieter. 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 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 interna-

tionale Preise gewonnen bzw. schlagen die US-amerikanischen Wettbewerber in wichtigen Benchmarks, z.B. im TPC Benchmark.

Innovative Forschungsprototypen an deutschen Universitäten. 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.

Datenmarktplätze

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 viele Unternehmen. Datenmarktplätze 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. 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.

Insbesondere werden es Informationsmarktplätze gerade kleinen und mittelständischen Unternehmen (KMU) ermöglichen, diese Daten zu analysieren und wirtschaftlich zu nutzen und somit einen Wettbewerbsnachteil auszugleichen.

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. In einer globalisierten und eng verzahnten Wirtschaft ist es notwendig, die Rahmenbedingungen zu setzen, die es deutschen Unternehmen ermöglicht 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.

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.

Prämisse 2: Verantwortungsvoller

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

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.

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 daher 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.

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.

Handlungsempfehlung 3: Stärkung von deutschen Technologieanbietern für Big Data (Technology Push). Im Bereich der Big Data-Technologien Deutschland ist 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 einzugreifen, der Anbieter kann sein Produkt oder Dienstleistung auf die konkreten Bedürfnisse der Anwender ausrichten. Das Ziel dabei ist es, die

Erfolgchancen der auf den Markt kommenden deutschen Big Data Technologieanbieter (Technology Push) zu erhöhen.

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.

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.

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.

Impressum

Herausgeber

Prof. Volker Markl, TU-Berlin
Prof. Thomas Hoeren, WWU-Münster
Prof. Helmut Krcmar, TU-München

Autoren

Die vorliegende Kurzfassung ist eine Zusammenfassung einer vom BMWi in Auftrag gegebenen Studie mit dem Titel "Innovationspotenzialanalyse für die neuen Technologien für das Verwalten und Analysieren von großen Datenmengen (Big Data Management)" Die Autoren der Studie sind: Volker Markl, Alexander Löser, Thomas Hoeren, Helmut Krcmar, Holmer Hemsens, Michael Schermann, Matthias Gottlieb, Christoph Buchmüller, Philip Uecker, Till Bitter

Bildnachweis

Foto (S. 3) Holmer Hemsens/TU-Berlin
Grafiken (S. 9-11) Eigene Darstellung basierend auf Ergebnissen und Daten der im Rahmen der Studie durchgeführten Umfrage

Redaktion

TU-Berlin/FG DIMA

Stand

März, 2014; Version 0.99