



# Aktuelle Themen

## Digitale Ökonomie und struktureller Wandel

# Big Data

## Die ungezähmte Macht

4. März 2014

### Autoren

Thomas F. Dapp  
+49 69 910-31752  
thomas-frank.dapp@db.com  
Veronika Heine

### Editor

Lars Slomka

Deutsche Bank AG  
Deutsche Bank Research  
Frankfurt am Main  
Deutschland  
E-Mail: marketing.dbr@db.com  
Fax: +49 69 910-31877

[www.dbresearch.de](http://www.dbresearch.de)

DB Research Management  
Ralf Hoffmann

Big Data wird zu einem volkswirtschaftlich relevanten Produktions-, Wettbewerbs- und somit Wachstumsfaktor. Moderne Analysetechnologien halten Einzug in sämtliche Lebensbereiche und verändern den Alltag. Insbesondere die Sensorik, biometrische Erkennungsverfahren und der allgemeine Trend zur Konvergenz von Informations- und Kommunikationstechnologien stimulieren die Big-Data-Bewegung.

Um die Potenziale optimal zu heben, sind enorme Herausforderungen zu bewältigen. Neben zunehmenden Datenmengen, unterschiedlichen Datenstrukturen sowie Echtzeitverarbeitungen spielen auch Datensicherheit, dringend reformbedürftige Datenschutzbestimmungen sowie wachsende Qualifizierungsanforderungen der Akteure eine wesentliche Rolle. Es fehlen vielerorts geeignete Digitalisierungs-Strategien, um auf den digitalen Strukturwandel zu reagieren.

Daten haben einen ökonomischen Wert. Die Begehrlichkeiten vieler Akteure, z.B. nach personenbezogenen Daten, steigen. Der Internetnutzer selbst könnte zum gehandelten Gut werden, ohne dass es ihm zu jeder Zeit bewusst ist oder er dadurch einen monetären Nutzen erfährt.

Die Chancen für Wirtschaft, Wissenschaft und Politik sind facettenreich: Sie reichen von der Produktivitätserhöhung, der Kostenersparnis sowie einer optimalen Kundenansprache bis hin zu besseren Prognosen von sich ausbreitenden Krankheiten oder wissenschaftlichen Modellanalysen. Zudem bieten öffentliche Verwaltungseinheiten vermehrt nicht personenbezogene Datensätze an, um neue Produkte, Dienstleistungen und Geschäftsmodelle zu generieren.

Dementsprechend sind die Risiken nicht zu unterschätzen: Viele beklagen nicht nur das wachsende Gefühl des „Information overloading“, sondern auch den zunehmenden Verlust ihrer Datenhoheit. Diverse Akteure sammeln, speichern und werten u.a. personenbezogene Daten zu unterschiedlichen Zwecken x-fach aus. Durch unverhältnismäßiges Datensammeln einiger Akteure könnte das Vertrauen in digitale Kanäle verlorengehen, zumindest spürbar sinken. Dies hätte fatale Auswirkungen auf die Innovationsfähigkeit sowie das Wachstum nicht nur im Bereich des noch jungen Forschungsfeldes der Informationstechnologie.

Die Big-Data-Bewegung ist nicht aufzuhalten. Nun gilt es, die notwendigen regulativen Rahmenbedingungen so zu setzen, dass sich die moderne Technologie und Methodik optimal entfalten kann. Der Wirtschaftsstandort Deutschland, aber auch die Europäische Union haben jetzt die Chance, mit ihren Erfolgen im Bereich Sensorik und Biometrie sowie im Bereich Datenschutz eine internationale Vorreiterrolle in Bezug auf sichere IT-Infrastrukturen einzunehmen.



## Big Data – die ungezähmte Macht

---

|   |    |
|---|----|
| 1. Big Data – Eine logische Evolutionsstufe im Internet ..... | 3  |
| 2. Akteure, Daten-Charakteristik, Quellen und Treiber .....   | 5  |
| 3. Sensorik und Biometrie erobern die Massenmärkte .....      | 10 |
| 4. Zur Rolle digitaler Ökosysteme in der Big-Data-Welt .....  | 13 |
| 5. Der ökonomische Wert von Daten .....                       | 18 |
| 6. Big Data und der Datenschutz .....                         | 21 |
| 7. (Big) Data in der Praxis: Es gibt keinen Königsweg .....   | 24 |
| 8. Grenzen von Big Data .....                                 | 33 |
| 9. Fazit .....  | 35 |
| Literaturverzeichnis .....                                    | 37 |
| Weiterführende Literatur .....                                | 38 |



## Big Data – die ungezähmte Macht

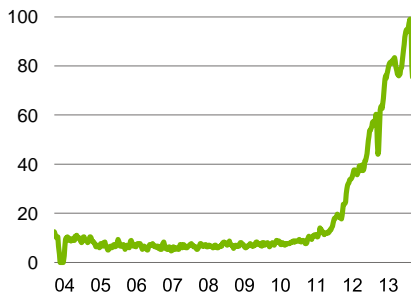
### 1. Big Data – Eine logische Evolutionsstufe im Internet

„Technologie ist weder gut noch schlecht. Sie ist auch nicht neutral.[...] Technologie interagiert in einer Weise mit ihrer sozialen Umwelt, dass technologische Entwicklungen häufig ökologische, soziale und menschliche Konsequenzen nach sich ziehen, die weitreichender sind als die unmittelbaren Ziele der Geräte und Praktiken, um die es eigentlich geht.“ [Melvin Kranzberg, 1986]

Websuche Begriff "Big Data"

1

Weltweites Interesse, 2004-heute, geglättet



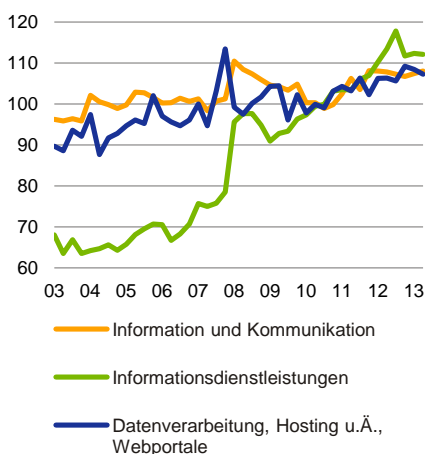
Quelle: Google Trends (www.google.de/trends/)

#### Eine neue Art des Denkens über Daten bricht an

Umsatzentwicklung IKT

2

Deutschland, 2010=100, arbeitstäglich- und saisonbereinigt



Quelle: Statistisches Bundesamt

Existiert zwischen Buch- und Schuh-Käufen beim Online-Händler Amazon zu bestimmter abendlicher Uhrzeit womöglich ein Zusammenhang? Können Kreditkartenfirmen aus einzelnen Transaktionen ihrer Kunden vorhersagen, bei wem eventuell Ehekrisen bevorstehen? Kann durch die GPS-Ortung mobiler Endgeräte in Kombination mit der Auswertung von Micro-Blogging-Diensten in nahezu Echtzeit eine drohende Epidemie vorhergesagt, dadurch verhindert oder zumindest abgemildert werden? Können von einer mit dem Internet verbundenen elektronischen Zahnbürste künftig während der Mundhygiene Informationen direkt an den Zahnarzt weitergeleitet werden, um gegebenenfalls einen zeitnahen Behandlungstermin zur Karies-Entfernung zu bekommen?

Diese Beispiele wirken zugegeben auf den ersten Blick etwas befremdlich, erfahren aber aufgrund der kontroversen und immer breiter geführten Diskussion über das Megathema „Big Data“ eine völlig neue Dimension. Mit Hilfe der Sensorik, einer Vielzahl von Datensätzen sowie bestimmter Algorithmen können sich künftig automatisierte Vorhersagen über bestimmte Verhaltenspräferenzen (nicht nur online) anhand einfacher Korrelationen abbilden lassen. Im Ozean der Daten sind vermutlich jede Menge ungeahnte, überraschende und wertvolle Korrelationen zu entdecken, an die heute kaum jemand denkt.

Neben den technologischen Möglichkeiten werden sich auch allmählich das Verständnis und die Einstellung der Menschen zu Daten sowie zur Datenanalyse verändern. Das Gestaltungspotenzial dessen, was mit soliden Daten, intelligenten Sensoren und Filterung alles gemessen und analysiert werden kann, ist aufgrund moderner Internet-Technologien noch nie so vielversprechend und lukrativ gewesen wie jetzt, in den Anfangsjahren des digitalen Zeitalters.

Wir befinden uns inmitten des digitalen Strukturwandels. Die Relevanz web-basierter Technologien nimmt in allen Wirtschaftszweigen stetig zu. Beinahe jede geschäftliche und betriebliche, aber auch viele private Transaktionen sind inzwischen an moderne, digitale Informationstechnologien gebunden, bei gleichzeitig sinkenden Transaktionskosten. Die Digitalisierung verändert den sozialen und wirtschaftlichen Alltag der Menschen, und die Akteure aus Wirtschaft, Wissenschaft oder Politik reagieren darauf – zugegeben mit unterschiedlicher Geschwindigkeit. Der volkswirtschaftlich relevante Produktionsfaktor, der hinter dieser digitalen Entwicklung steckt und stetig wächst, sind einzelne, unterschiedliche Datenpunkte, die vermehrt auch in Echtzeit nachgefragt werden.<sup>1</sup>

Big Data ist nach einer Reihe logischer Entwicklungsstufen im Internet, wie der Individualisierung, der Verlagerung der Daten in die Cloud oder des rapide steigenden Wunsches nach (digitaler) Mobilität, das neue kontrovers diskutierte Thema. Es schlägt Brücken zu den vorhergehenden. Prinzipiell geht es darum, unterschiedliche Datenmengen mit neuen Datensätzen zu kombinieren, eventuelle Muster in diesen kumulierten Daten mit intelligenten Softwareprogrammen aufzuspüren, um anschließend die richtigen (möglichst wirtschaftlich lukrativen) Schlüsse aus den Ergebnissen zu ziehen. Einmal erhobene Primärdatensätze können zu unterschiedlichen Zwecken und für unterschiedliche Akteure zweit-, dritt-, oder x-fach ausgewertet werden. Dadurch erweisen sich die Daten einerseits als Quelle für Innovation, Kreativität sowie „Out-of-the-box-Denken“ und münden idealerweise in neuen Geschäftsideen, Produkten oder Dienstleistun-

<sup>1</sup> Aus einem Datenberg etwas Wertvolles zu extrahieren bzw. ein Muster zu erkennen, wird in der Wissenschaft auch als Data-Mining bezeichnet.

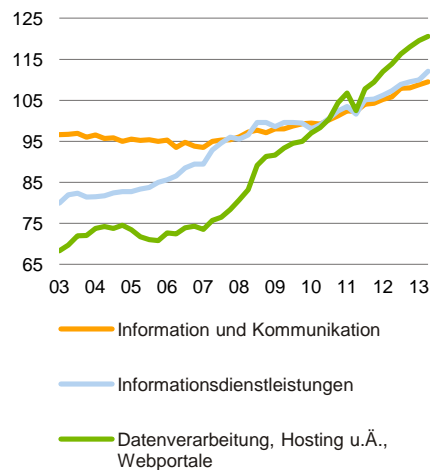


## Big Data – die ungezähmte Macht

Beschäftigung im Dienstleistungsbereich (IKT)

3

Deutschland, 2010=100



Quelle: Statistisches Bundesamt

gen. Andererseits können und werden sie den Menschen aber auch viele Sorgen und Ängste bereiten, weil die eigene Datenhoheit, also die informationelle Selbstbestimmung, schnell verlorengehen kann, wie viele Beispiele zeigen.

Am Ende des Big-Data-Prozesses gilt es, die richtigen Schlüsse aus der Kombination von Technologie und Methodik zu ziehen. Hier lauert für die Entscheidungsträger eine weitere Herausforderung, denn hier treffen zwei Welten aufeinander: Die Bits und Bytes, also die maschinellen Algorithmen auf der einen Seite, und der Mensch mit seinen Instinkten, seiner Erfahrung und seiner Intuition auf der anderen Seite. Zählen im Entscheidungsmoment die Ergebnisse aus der Statistik oder die langjährigen Erfahrungen bzw. das individuelle Bauchgefühl? Das Potenzial der maschinellen Algorithmen, wertvolle Informationen aus den kombinierten Datensätzen zu gewinnen, steckt sicherlich noch in den Kinderschuhen und ist ausbaufähig. Es bleibt aber zu hoffen, dass sich trotz des technologischen Fortschritts und der kaum fassbaren Dynamik in diesem Bereich ein ausgewogenes Mensch-Maschine-Verhältnis herausbilden wird. Abgesehen von den bis dato ungenügenden Datenschutzrichtlinien in diesem noch jungen (internationalen) Forschungsbereich möchten sicherlich die Wenigsten in einer Welt leben, in der Entscheidungsfindungsprozesse ausschließlich programmierten (womöglich selbstlernenden) Algorithmen überlassen werden. Dagegen können moderne IK-Technologien natürlich die Effizienz erhöhen, Kosten reduzieren sowie menschliches Versagen minimieren – ein Trade-Off!

Dieser nicht-trivialen Herausforderung stehen zunehmend Akteure aus Wirtschaft, Wissenschaft, Politik und Gesellschaft gegenüber. Die ökonomische Bedeutung von (Echtzeit-)Daten, wie Finanz- und Handelstransaktionsdaten, Sensor- bzw. Messdaten, Gesundheitsdaten, Social-Networking-Analysen sowie deren Auswirkungen auf die geschäftlichen Abläufe in einer global integrierten Wirtschaft, aber auch im privaten und sozialen Umfeld nimmt kontinuierlich zu. Es werden immer mehr Daten von unterschiedlichen Akteuren erzeugt, die auch gefiltert, zu unterschiedlichen Zwecken analysiert und mehrfach ausgewertet werden können – Tendenz: exponentiell steigend.

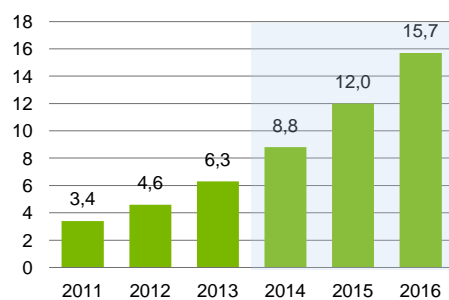
### Der Wunsch der Menschen, die Welt zu quantifizieren

Obwohl wir uns erst am Anfang der Big-Data-Entwicklung befinden, nähern wir uns mit moderner Technologie und Methodik langfristig einem Zustand, bei dem menschliche Handlungen, Aktivitäten sowie Gefühlszustände mehr und mehr messbar gemacht werden können. Das bedeutet, dass Daten in eine maschinenlesbare Form gebracht werden können, um sie gegebenenfalls (mit anderen Datensätzen) zu kombinieren und auszuwerten. Durch die Möglichkeiten der Sensorik sowie durch biometrische Erkennungsverfahren kommen wir dadurch dem seit Menschheitsgedenken bestehenden Wunsch näher, die Welt und die darin lebenden Individuen zu quantifizieren. Damit kommen wir aber auch den Visionen der humanoiden Robotik und der künstlichen, lernfähigen Intelligenz näher, wie sie bereits in vielen Science-Fiction-Filmen vorgestellt wurden und auch für die Massenmärkte Realität werden könnten.

Marktpotenzial Big Data global

4

Mrd. EUR



Quellen: Bitkom, Experton Group

Wie bei jeder neuen und (r)evolutionären Bewegung eröffnen sich Chancen, aber auch Risiken. Beide Ausprägungen sollten jetzt, in diesem frühen Stadium der Entwicklung transparent und breit diskutiert werden. Aus deutscher bzw. europäischer Perspektive sollten aus Gründen der Wettbewerbsfähigkeit jetzt günstige Rahmenbedingungen geschaffen werden, um die Potenziale der Big-Data-Bewegung zu stimulieren. Gleichzeitig müssen unverhältnismäßige und rechtlich sowie moralisch fragwürdige Praktiken des Datenmissbrauchs durch europäische, idealerweise über Europa hinaus geltende datenschutzrechtliche Bestimmungen eingeschränkt werden.

**In welchen Schritten nähern wir uns Big Data?** Im folgenden Kapitel werden die unterschiedlichen Akteure vorgestellt sowie die Grundlagen der Daten-Charakteristik erläutert. Darüber hinaus stehen die Quellen und Treiber der wachsenden Datenflut im Fokus und werden anhand einiger Beispiele erklärt.



## Big Data – die ungezähmte Macht

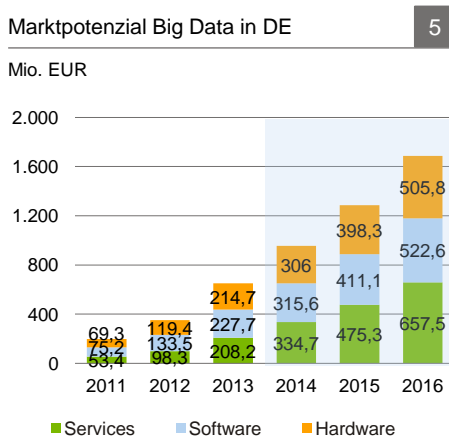
Das Kapitel 3 stellt insbesondere die Bereiche Sensorik und Biometrie in den Mittelpunkt. Beide Technologien werden zunehmend massentauglich und versprechen künftig lukrative Wachstumspotenziale. Als branchenübergreifender Trend schafft die Konvergenz von Informations- und Kommunikationstechnologien zusätzliche Impulse und Raum für Experimente sowie künftige neue Geschäftsmodelle, innovative Produkte und Dienstleistungen. Die Rolle der digitalen Ökosysteme in der Big-Data-Diskussion wird im 4. Kapitel analysiert. Einzelne große Internetfirmen erweitern ihre Kernkompetenzen mit Hilfe moderner Technologien und Methoden im Bereich Datenanalyse. Dadurch steigen nicht nur das Spektrum ihrer Produkte und Dienstleistungen, sondern auch der Lock-in-Effekt sowie die Wechselkosten für die Konsumenten. Daten haben einen ökonomischen Wert. Diverse Datensätze können von unterschiedlichen Akteuren zu unterschiedlichen Zwecken (mehrfach) verwendet, monetarisiert, aber auch missbraucht und zweckentfremdet werden. Grundsätzlich besteht also die Möglichkeit, verblüffende (oder gar beängstigende) detailgetreue (digitale) Profile der Menschen zu erstellen – größtenteils ohne deren Erlaubnis. Die für Außenstehende schwer nachvollziehbare und unverhältnismäßige Sammelaktivität einzelner Akteure wird in Kapitel 5 diskutiert.

Die für die Big-Data-Bewegung ungenügenden, aber elementaren Datenschutzrichtlinien sowie einige Lösungsansätze sind Themen des Kapitels 6. Das Kapitel 7 gewährt einen kurzen Einblick in die unterschiedlichen Big-Data-Projekte im Bereich Wirtschaft, Wissenschaft und öffentliche Verwaltung. Die Grenzen der Big-Data-Bewegung, allem voran die zunehmende Komplexität durch die Verschmelzung unterschiedlicher Datensätze, sind in Kapitel 8 erwähnt. Die Studie schließt in Kapitel 9 mit einem kurzen Resümee sowie einem Ausblick, wohin die Reise in den kommenden Jahren gehen könnte.

## 2. Akteure, Daten-Charakteristik, Quellen und Treiber

*„I keep saying the sexy job in the next ten years will be statisticians. People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s?“* [Hal Varian, Google's Chief Economist on Statistics and data, January 2009]

### Big Data ist mehr als nur IT: Akteure aus diversen Bereichen



Quellen: Bitkom, Experton Group

Viele Entscheidungsträger haben erkannt, dass Big Data längst kein reines IT-Thema mehr ist. Big Data entwickelt sich vielmehr zu einer Bewegung, die das Zusammenspiel moderner Internet-Technologien und Analyse-Methoden umfasst, welche die Erfassung, Speicherung und Auswertung großer und erweiterbarer, vor allem unterschiedlich strukturierter Daten ermöglicht. Big Data erfährt dadurch eine breite, internationale Dimension mit unterschiedlichen wissenschaftlichen Erkenntnissen und Erwartungen hinsichtlich Wachstums- oder Effizienzsteigerungen. Big Data öffnet vor allem den Raum für Experimente, Innovation und Kreativität, bietet einen reichen Fundus an neuen möglichen Datenkombinationen und steht somit auch für die Entdeckung ungeahnter Zusammenhänge. Dadurch birgt Big Data das Potenzial, neue Geschäftsmodelle, Produkte und Dienstleistungen zu schaffen bzw. Innovationen voranzutreiben.

Aufgrund der Ressourcenknappheit sowie der Stärke wissensintensiver Produkte und Dienstleistungen in Deutschland, hat die Big-Data-Bewegung durchaus Potenzial, sich zu einem künftig stabilen komparativen Vorteil für den deutschen Wirtschaftsstandort zu etablieren. Big Data kann zu einem volkswirtschaftlich relevanten Produktions- und Wettbewerbsfaktor mit damit einhergehenden neuen Möglichkeiten der Wertschöpfung werden und für ähnlich (r)evolutionäre Bewegungen wie die massenhafte Nutzung des Internet selbst sorgen. Allem voran versprechen sich viele Anbieter neuer und effektiver Datenanalyse-Tools



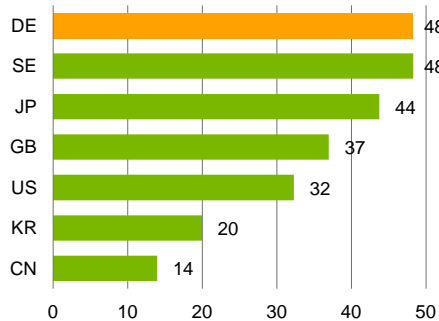


## Big Data – die ungezähmte Macht

### DE und SE: Höchste Raten an erteilten IKT-Patenten

6

%, Anteil gewählter Patente an den Patentanmeldungen im Durchschnitt 2003-2012



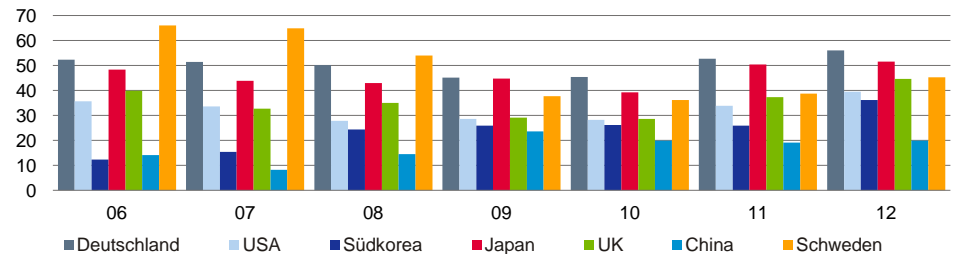
Quelle: EPO

langfristige Effizienzsteigerungen und insbesondere lukrative Umsatzchancen, u.a. durch die Gewinnung und Analyse detaillierter kundenspezifischer Informationen, d.h. **Big Data verspricht** am Ende vor allem auch **Big Business**.

### Patente im IKT\*-Bereich

7

%, Anteil gewählter Patente an den Patentanmeldungen



\* Definition EPO: Audio-visuelle Technologie, Telekommunikation, digitale Kommunikation, elementare Kommunikationsprozesse, Computertechnologie, IT-Methoden für das Management, Halbleiter und elektrische Maschinen, Apparate, Energie

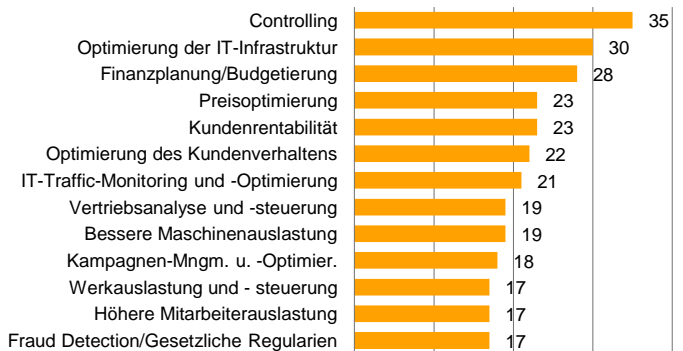
Quelle: EPO

Neben der IT-Branche sehen sich also auch Akteure aus Wirtschaft, Wissenschaft oder öffentlicher Verwaltung zunehmend mit Big Data konfrontiert. Die Analyse der Datenberge, die in den unterschiedlichen Bereichen entstehen, kann nicht nur den volkswirtschaftlichen Nutzen erhöhen, sondern wird zunehmend auch dafür eingesetzt, die gesellschaftlichen sowie humanitären, z.B. medizinischen, Probleme zu mildern, wie ein Beispiel in Kapitel 7 zeigt. Dies kann aber nur unter der zwingenden Voraussetzung gelingen, dass sich die modernen Technologien und Entwicklungen innerhalb eines verbindlichen Rechtsrahmens bewegen. Dieser muss den Datenschutzrichtlinien auf europäischer, bestenfalls internationaler Ebene genügen bzw. das Recht auf informationelle Selbstbestimmung in den Vordergrund rücken.

### Geschäftsbereiche mit größtem Nutzen

8

% der Befragten\* (n=254), 2012

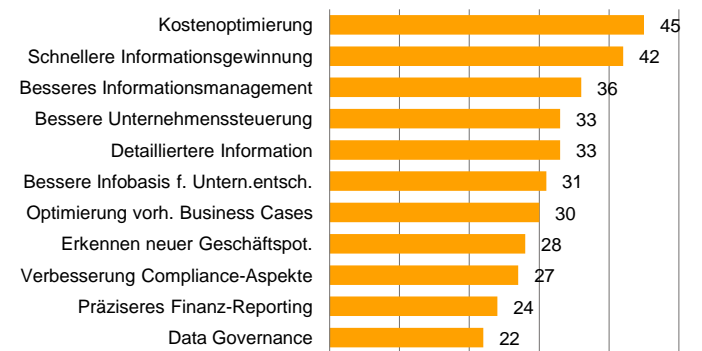


\*Mehrfachnennungen möglich  
Quelle: IDC

### Unternehmerisches Potenzial von Big-Data

9

% der Befragten\* (n=254), 2012



\*Mehrfachnennungen möglich  
Quelle: IDC

### Die Charakteristik von Daten: komplex und heterogen, aber wertvoll

In der Big-Data-Diskussion werden den unterschiedlichen Daten im Wesentlichen drei Eigenschaften zugeordnet, die gemäß ihren englischen Bezeichnungen als die drei „Vs“ bekannt sind und im Folgenden anhand einiger Beispiele und Illustrationen vorgestellt werden:



## Big Data – die ungezähmte Macht

### — Volume (Datenmenge)

Das erste Kriterium behandelt die Datenmengen, die im privaten sowie geschäftlichen, wissenschaftlichen oder öffentlichen Bereich anfallen. Aufgrund der fortschreitenden Digitalisierung sind beinahe alle Bereiche des modernen Lebens involviert. Alleine im privaten Bereich speichern und archivieren die Menschen eine enorme Datenmenge im dreistelligen Gigabyte-Bereich. Hierzu zählen z.B. digitale Fotos, Dokumente, Tabellen, E-Mails, Musiktitel oder Kinofilme. Unternehmen, wissenschaftliche Institute oder öffentliche Verwaltungen speichern, archivieren und werten Datenmassen ebenfalls aus, allerdings in viel größeren Dimensionen.

### Mehr Bits und Bytes wie Sand am Meer

Der Speicherplatz auf mobilen Endgeräten (Smartphone, Tablet-PC, E-Reader) oder auf Notebooks ist im Gigabyte-Bereich und somit kapazitätsmäßig noch vorstellbar.<sup>2</sup> Aber Unternehmen, wissenschaftliche Institutionen oder auch Geheimdienste, die täglich hohe Datenbestände generieren, verwalten und auswerten, rechnen mittlerweile im Tera-, Peta-, Exa-, und Zettabyte-Bereich. Im vergangenen Jahr wurden weltweit 1,8 Zettabyte an Daten produziert. Ein Zettabyte entspricht einer Trilliarde Bytes. Das ist eine Zahl mit 21 Nullen: 1.000.000.000.000.000.000.

### Übersicht Maßeinheiten für Datengrößen

10

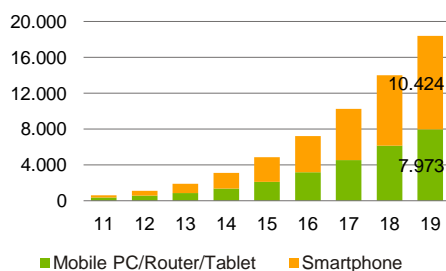
|                  |   |                       |   |                                   |
|------------------|---|-----------------------|---|-----------------------------------|
| 1 Byte           | = | 8 Bits                | = | 1                                 |
| 1 Kilobyte (kB)  | = | 10 <sup>3</sup> Byte  | = | 1.000                             |
| 1 Megabyte (MB)  | = | 10 <sup>6</sup> Byte  | = | 1.000.000                         |
| 1 Gigabyte (GB)  | = | 10 <sup>9</sup> Byte  | = | 1.000.000.000                     |
| 1 Terabyte (TB)  | = | 10 <sup>12</sup> Byte | = | 1.000.000.000.000                 |
| 1 Petabyte (PB)  | = | 10 <sup>15</sup> Byte | = | 1.000.000.000.000.000             |
| 1 Exabyte (EB)   | = | 10 <sup>18</sup> Byte | = | 1.000.000.000.000.000.000         |
| 1 Zettabyte (ZB) | = | 10 <sup>21</sup> Byte | = | 1.000.000.000.000.000.000.000     |
| 1 Yottabyte (YB) | = | 10 <sup>24</sup> Byte | = | 1.000.000.000.000.000.000.000.000 |

Quelle: Eigene Darstellung

### Datenverkehr mobile, web-basierte Endgeräte

11

Petabytes pro Monat



Quelle: Ericsson 2013

„Experten haben hochgerechnet, dass die Menschheit vom Beginn der Zeitrechnung bis zum Jahre 2003 rund fünf Milliarden Gigabyte an Daten erzeugt hat. [...] Im Jahr 2011 sammelte sich dieselbe Datenmenge – 4,4 Exabyte – bereits alle 48 Stunden an.“<sup>3</sup> Unter Fortsetzung dieses Trends, insbesondere aufgrund leistungsfähigerer Rechner, günstigerer Speichermedien sowie intelligenterer Algorithmen, fielen diese Datenmengen im Jahr 2013 bereits alle zehn Minuten an. Glaubt man den Experten, so verdoppelt sich das Volumen also alle zwei Jahre. Der „digitale Heustock“ wird folglich immer größer, und die Suche nach den relevanten Stecknadeln immer schwieriger. Dabei wird zu Recht die kritische Frage gestellt, welche unterschiedlichen Akteure eigentlich welche Stecknadeln im permanent wachsenden Heuhaufen suchen. Vor allem ist interessant zu erfahren, wer genau Zugang zu welchen Datensätzen hat und wer dabei für welche Zwecke, welche Analysemethoden anwendet. Mitunter liegt die Herausforderung auch an den beschleunigten Entstehungsraten der Daten, welche die Unternehmen möglichst in Echtzeit auswerten möchten. Dies bildet die Grundlage für die zweite Charakteristik:

### — Velocity (Geschwindigkeit)

Während die Daten im analogen Zeitalter noch in überschaubaren Frequenzen anfielen, erfasst und ausgewertet wurden, fließen die Datenströme im digitalen Zeitalter aufgrund der Vielzahl an vernetzten Sensoren und der Verschmelzung

<sup>2</sup> Auf zwei Gigabyte verfügbaren Speicherplatz passen z.B. rund 360 MP3s in guter Klangqualität, ein eineinhalb Stunden langer Film in HD (High Definition) sowie etwa 150 Fotos mit einer 5-Megapixel-Auflösung.

<sup>3</sup> Heuer, S. (2013). Kleine Daten, große Wirkung. Big Data einfach auf den Punkt gebracht. Landesanstalt für Medien Nordrhein-Westfalen (LfM). Düsseldorf.



der Informations- und Kommunikationstechnologien (IKT) rund um die Uhr quer über den Erdball. Um auch zeitkritischen Geschäfts- und Entscheidungsprozessen gerecht zu werden, wird es immer wichtiger, die anfallenden Datenströme sowohl von internen als auch externen Quellen in Echtzeit zu integrieren bzw. zu analysieren. Gerade hiervon versprechen sich viele Entscheidungsträger enorme Wettbewerbs- und Effizienzvorteile. Auch die Wissenschaft profitiert von Echtzeitdaten, wenn es beispielsweise darum geht, dass Krankheitserreger zeitnah erkannt werden sollen, um eventuelle präventive Schritte einzuleiten, bevor es zu einer starken Verbreitung der Krankheit kommen kann.

— **Variety** (unterschiedliche Datenbeschaffenheit)

Es geht aber nicht ausschließlich um die Datenflut, sondern auch um die Vielfalt an unterschiedlichen Datentypen und -strukturen, die möglichst einheitlich und harmonisiert verarbeitet werden müssen, um sie miteinander kombinieren zu können. Grundsätzlich kann zwischen strukturierten, halbstrukturierten und unstrukturierten Daten unterschieden werden.

Während bisher größtenteils strukturierte Informationen in relationalen Datenbanken verarbeitet werden, verlieren neuere Daten, z.B. aus sozialen Netzwerken, zunehmend an Struktur und können kaum mit herkömmlichen Datenbanken nutzenstiftend verwertet werden. Zu den strukturierten Daten zählen beispielsweise Kundenstammdaten, wie Geschlecht, Geburtsdatum, Adresse etc. Unstrukturierte Daten sind z.B. Bilder, Videos und Audiodateien. E-Mails wiederum zählen zu den halbstrukturierten Daten, weil sie in der Kopfzeile einen Absender, einen Empfänger und einen Betreff, also strukturierte Daten, haben. Der Rumpf der E-Mail ist hingegen unstrukturiert, weil für den Text einer E-Mail keine vorgegebene Struktur existiert. Experten schätzen, dass heute nur 15% der Daten strukturiert und ca. 85% der Daten unstrukturiert sind.<sup>4</sup>

Die aktuelle Herausforderung für IT-Netzwerkarchitekten besteht also darin, heterogene Datenformate aus unterschiedlichen Datenquellen miteinander kompatibel zu machen, um sie anschließend analysieren, verknüpfen und auswerten zu können. Die Integration dieser Daten ist aber komplex. Es müssen Informationen aus Bildern, Videos, MP3-Dateien, Streaming-Diensten, Micro-Blogging-Diensten, Blogs, Foren, Suchmaschinen, Klickprotokollen, E-Mails, Internet-Telefonie, digitaler Korrespondenz und/oder aus Sensoren lesbar gemacht werden, damit am Ende aus den unterschiedlichen Rohdaten maschinenlesbare Daten werden. Gerade bei Ton, Video- oder Bilddateien ist die Umwandlung in einheitliche maschinenlesbare Daten technisch besonders herausfordernd. Hier müssen die zugrundeliegenden Algorithmen z.B. Sprache(n) verstehen und transkribieren, Gesichter oder Firmenembleme erkennen oder auch urheberrechtlich geschützte digitale Inhalte erkennen können.<sup>5</sup>

In der von Tim Berners-Lee<sup>6</sup> initiierten Diskussion um das semantische Web<sup>7</sup> (Web 3.0) zur Weiterentwicklung des Internet der Dinge (Ubiquitous Computing) wird gefordert, dass die von Maschinen zusammengetragenen Informationen auch von anderen Maschinen gelesen und richtig eingeordnet werden können. Beispiele: Meint der Begriff „Golf“ im Kontext jetzt die Sportart, das Fahrzeug oder eine Meeresbucht im Atlantik; wie wandelt die elektronische Zahnbürste via Sensordaten eventuelle Zahnstein- oder Karies-Bildungen in maschinenlesbare

#### Big Data und der US-Wahlkampf

12

Während der letzten Wahlkampagne von Barack Obama kamen einige moderne Big-Data-Analyse-Tools zum Einsatz. Im Vordergrund standen Fragen, inwiefern der Wirkungskreis des Wahlkampfes verbessert werden kann bzw. wie die richtigen Wähler effektiv von jenen, die bereits verloren waren, zu unterscheiden sind. Hierfür wurden soziodemografische Daten herangezogen, also z.B. Daten aus dem Bereich Konsumverhalten (Informationen aus Loyalty-Programmen, Einkaufsverhalten, Klickverhalten im Internet, etc.) sowie Daten aus sozialen Netzwerk-Plattformen (z.B. Profile bei Facebook). Sämtliche Daten wurden mit den Wählerlisten korreliert, um eventuelle Muster von jenen Menschen zu erkennen, die nicht zur Wahl gehen wollen, aber dennoch für Obama stimmen würden. Mit Hilfe verschiedener Prognose-Instrumente (Predictive Analytics) wurden bis zu 100 Variablen in das System eingefüttert um somit das Wahlverhalten zu ermitteln. Datenschützer kritisieren, dass sich viele der Anwendungen in einem Graubereich befinden und die informationelle Selbstbestimmung der Individuen verletzt wird. Da die Gesetzeslage in den USA aber nicht mit deutschen oder europäischen Datenschutzrichtlinien übereinstimmt, konnten die diversen Datensätze der Wähler in den USA kombiniert und zu Wahlzwecken ausgewertet werden.

Quelle: <http://www.predictiveanalyticsworld.com/patimes/eric-siegel-explains-how-the-obama-2012-campaign-used-predictive-analytics-to-influence-voters-2/>

<sup>4</sup> TNS Infratest GmbH – Geschäftsbereich Technology: Quo Vadis Big Data – Herausforderungen – Erfahrungen – Lösungsansätze, 2012.

<sup>5</sup> Vgl.: Heuer, S. (2013). Kleine Daten, große Wirkung. Big Data einfach auf den Punkt gebracht. Seite 12.

<sup>6</sup> Begründer des World Wide Webs.

<sup>7</sup> Während das Internet eine Möglichkeit bietet, Daten miteinander zu vernetzen, zeigt das semantische Web einen Weg auf, um Informationen auf der Ebene ihrer Bedeutung miteinander zu verknüpfen. Die Daten in einem semantischen Web sind strukturiert und in einer Form aufbereitet, welche es Computern ermöglicht, sie entsprechend ihrer inhaltlichen Bedeutung zu verarbeiten. Zudem erlaubt ein semantisches Web Computern aus den vielen Informationen der weltweiten Daten Wissen herzuleiten und vor allem neues Wissen zu generieren.



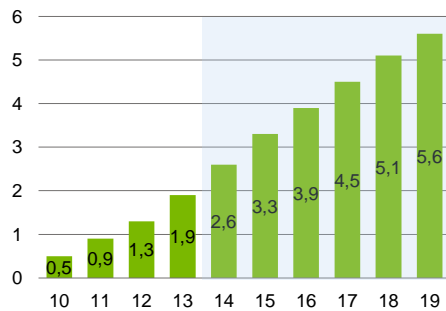


## Big Data – die ungezähmte Macht

### Weltweiter Bestand an Smartphones

13

Mrd.



Quelle: Ericsson Mobility Report 2013

Sprache um, damit die Infos, wenn sie beim Zahnarzt ankommen, auch ausgewertet werden können?

Besonders interessant sind auch subjektive Äußerungen in Text-, Wort-, oder Videobeiträgen. Jede PR-, oder Wahlkampfagentur sowie Werbe- oder Marketingabteilung ist daran interessiert die Stimmungen oder Meinungen der Menschen zu gewissen Themen möglichst in Echtzeit einzufangen. Um die zum Teil emotionalen Äußerungen zu Produkten, Marken oder Ereignissen maschinenlesbar zu machen, werden intelligente Analyse-Instrumente und Algorithmen benötigt, die wertende Aussagen erkennbar machen. Dies wird künftig sicherlich die Königsdisziplin im Bereich Big Data werden, weil es technisch besonders herausfordernd ist und weiterer Forschung bedarf.

Zusätzlich zu den gängigen drei Charakteristika wird von Experten inzwischen ein weiteres „V“ diskutiert:

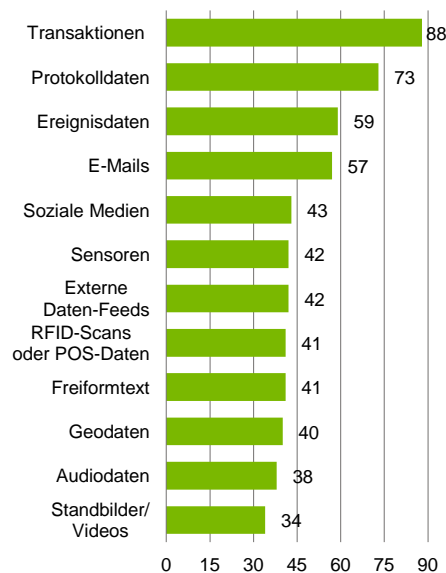
— **Veracity**<sup>8</sup> (Richtigkeit, Zuverlässigkeit von Daten)

Insbesondere bei subjektiven Stimmungs- oder Meinungsäußerungen sind bestimmte Ereignisse und Daten naturgemäß nicht vorhersehbar. Ähnlich verhält es sich bei Wetterdaten, seismologischen Messdaten, makroökonomischen Indikatoren oder Prognosedaten. Gerade für komplexe Berechnungen von Szenarien in der Wissenschaft sind aber derartige Prognosedaten elementar wichtig. Doch trotz der Relevanz dieser Daten kann die Unsicherheit – analytisch gesehen – nicht durch irgendwelche Bereinigungsverfahren eliminiert werden.

### Hauptsächlich interne Quellen

14

% der Befragten (n pro Nennung=557-867 aus 1.144), 2012



Quellen: IBM Institute for Business Value, Saïd Business School Oxford

Wie so oft bei jungen Technologien steckt die wissenschaftliche Potenzialanalyse auch bei Big Data (noch) in den Kinderschuhen. Daher bleiben einige Fragen weiterhin unbeantwortet. Aus den Datenanalysen können natürlich Risiken abgeleitet werden, die mit Hilfe entscheidungstheoretischer Ansätze und unterschiedlichen Wahrscheinlichkeiten benannt werden können; es gibt aber auch unsichere Faktoren, die nicht vorhersehbar sind, sogenannte schwarze Schwäne.<sup>9</sup> Die Notwendigkeit, Unsicherheiten zu erkennen und sie möglichst realistisch einzubeziehen, ist eine weitere Eigenschaft von Daten und eine permanente Herausforderung, für die auch Big Data kein Allheilmittel darstellt.

### Quellen/Treiber von Big Data: Branchenübergreifender digitaler Strukturwandel

Die Fülle an möglichen Datenquellen ist schier unerschöpflich. Sie lassen sich nach ihrem Ursprung grob in drei Gruppen unterteilen. Oft werden die Daten auch durch eine Kombination der drei Gruppen generiert:<sup>10</sup>

- **Von Maschinen generierte Daten:** z.B. Sensor- oder Logdaten<sup>11</sup>, Sprache/Audio/Video, Klickstatistiken, Datendienste.
- **Von Menschen generierte Daten:** z.B. Social Networking, Korrespondenz, Publikationen/Patente, Bilder, Freitext, Formulare, Protokolle, Open Data/Web Content, Sprache/Audio/Video;
- **Geschäftsdaten:** z.B. Stamm- und Falldaten, CRM<sup>12</sup>- und Transaktionsdaten.

<sup>8</sup> Schroeck, M. et al. (2012). Analytics: Big Data in der Praxis. Wie innovative Unternehmen ihre Datenbestände effektiv nutzen. Seite 4f.

<sup>9</sup> In Anlehnung an „Der schwarze Schwan: Die Macht höchst unwahrscheinlicher Ereignisse“, von Nassim Nicholas Taleb (2007). Im Prinzip geht es darum, dass bei Unsicherheit keine Eintrittswahrscheinlichkeiten verfügbar sind, während in Risikosituationen Eintrittswahrscheinlichkeiten für Ergebnisse bzw. Umweltzustände gegeben sind.

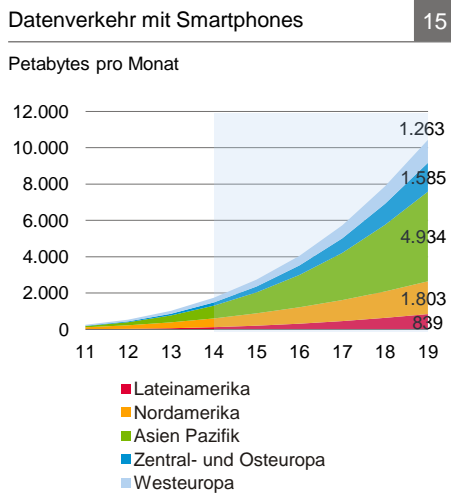
<sup>10</sup> Vgl. Fraunhofer IAIS (2012). Big Data – Vorsprung durch Wissen. Seite 20.

<sup>11</sup> Eine Logdatei/Protokolldatei enthält das automatisch geführte Protokoll aller oder bestimmter Aktionen von Prozessen auf einem Computersystem.

<sup>12</sup> CRM = Customer Relationship Management (dt. Kundenbeziehungsmanagement oder Kundenpflege).



## Big Data – die ungezähmte Macht



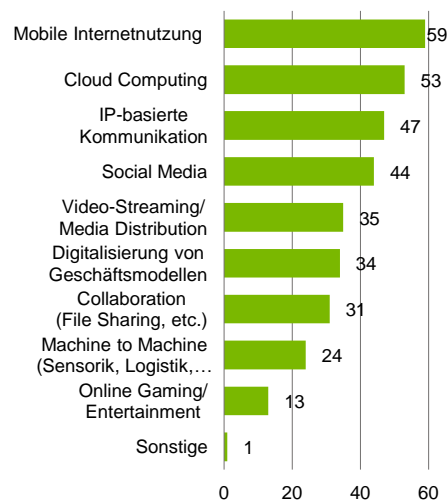
Zu den dominanten Treibern zählen sicherlich die flächendeckende Durchdringung der Digitalisierung (digitaler Strukturwandel) von Infrastrukturen in den Bereichen Energie, Gesundheit, Verkehr, Bildung oder der öffentlichen Verwaltung sowie die zunehmende private und berufliche Nutzung mobiler, web-basierter Endgeräte. Sei es beim Surfen im Internet, wo rund um die Uhr einzelne Klicks der Nutzer mit Hilfe von Tracking-Software verfolgt und gespeichert werden, oder bei der Erfassung von Orts- und Bewegungsmustern die über RFID<sup>13</sup>, Infrarot oder Wireless-Technologien (z.B. NFC [Near Field Communication]) erfasst werden, um Aufenthaltsorte von Individuen sichtbar zu machen. Es werden permanent Film- und Musikdateien hoch- und runtergeladen sowie Streaming-Dienste in Echtzeit genutzt. Gleichzeitig interagieren Millionen von Nutzern auf sozialen Netzwerk-Plattformen und produzieren digitale Inhalte, die ebenfalls in Echtzeit viral gestreut werden. All das führt zu einer explodierenden Datenmenge, die einerseits wertvoll sein kann, andererseits aber auch hohe Anforderungen an die Infrastruktur und Architektur der IT-Systeme stellen.

Tatsächlich verhält es sich jedoch so, dass die umfangreichsten Datenbestände nicht von sozialen Netzwerk-Plattformen stammen. So sind z.B. alle Kurznachrichten einer Micro-Blogging-Plattform (Tweets) zu bestimmten Themen längst nicht so umfangreich wie Datensätze aus der Industrie und/oder der Wissenschaft. Große Datensätze stammen z.B. von geophysischen Messanalysen auf Ölplattformen oder von Seismographen und Wetterstationen sowie aus komplexen Berechnungen von Szenarien der naturwissenschaftlichen oder Ingenieurs-Disziplinen. Ein Teil der großen Datensätze stammt sicherlich auch aus dem global vernetzten Finanzsektor. Alleine die digitalen Handelssysteme generieren sekundlich Millionen von Transaktionen, die gespeichert und in Echtzeit ausgewertet werden könn(t)en.

## Globale Treiber für Big Data

16

Umfrage dt. Unternehmen > 500 MA, Mehrfachnennungen möglich [n=100], %, 2012



Quelle: Eine Studie der Experton Group AG im Auftrag der GmbH & Co. oHG

Der digitale Strukturwandel als Haupttreiber für Big Data bietet insbesondere für die Sensorik<sup>14</sup> sowie die Biometrie als essentielle Schlüsseltechnologien ein enormes experimentierfreudiges und lukratives Einsatzgebiet. Diese Technologien haben quasi die Funktion eines Turboladers im Bereich Big Data, weil Sensoren an immer mehr alltäglichen (massenmarktauglichen) Gegenständen, Maschinen oder sonstigen Objekten angebracht werden. Damit kommen die Menschen ihrem Ziel, alles um sie herum messbar zu machen, näher. Diesem relevanten und dynamischen Treiber widmen wir daher das kommende Kapitel:

### 3. Sensorik und Biometrie erobern die Massenmärkte

Die Eroberung der Massenmärkte durch Sensoren und biometrische Erkennungsverfahren ist mitunter dadurch zu erklären, dass sich mittlerweile eine Vielzahl diverser Sensoren in jedem Smartphone/Tablet oder sonstigen mobilen, web-basierten Endgeräten befindet. Immer mehr Nutzer kommen auf diesem Wege mit der Sensortechnologie oder der Messung individueller Körpermerkmale in Verbindung. Durch die stabiler und schneller werdenden Mobilfunknetze sind viele Menschen mit ihren mobilen Endgeräten auch permanent mit dem Internet verbunden, was der Konnektivität „Maschine zu Mensch“ bzw. „Maschine zu Maschine“ einen zusätzlichen Schub verleiht. Die eingearbeiteten Sensoren unterstützen die Entwicklung neuer Funktionen bei Internetdiensten, beispielsweise via Apps, und werden täglich millionenfach genutzt.

Gemessen werden können Temperaturen, Distanzen, Längen, Licht, Zeit, Tiefen, Geschwindigkeiten, Körpermerkmale, Gewichte, etc. Kein Bereich der

<sup>13</sup> Die RFID-Technologie (engl. radio frequency identification) ermöglicht die automatische Identifizierung und Lokalisierung von Gegenständen und Lebewesen und erleichtert damit erheblich die Erfassung von Daten.

<sup>14</sup> Die Sensorik umfasst in der Technik die Wissenschaft und die Anwendung von Sensoren zur Messung und Kontrolle von Veränderungen von umweltlichen, biologischen oder technischen Systemen.



## Big Data – die ungezähmte Macht

### Mobile Endgeräte bieten eine Fülle an Sensoren ...

### ... sowie zunehmend auch biometrische Erkennungsverfahren

Mess-, Steuer-, Regelungs- oder Automatisierungstechnik kommt heutzutage ohne Sensoren aus. Moderne Mikroelektronik, die Miniaturisierung von Sensoren, die somit immer kleiner, leistungsfähiger und günstiger werden, verstärken diese Entwicklung zusätzlich.

Mobile Endgeräte beinhalten u.a. Bewegungssensoren, Lichtsensoren, Höhenmesser, digitale Kompass, Fingerabdruck-, Gang- oder Spracherkennungssensoren, Näherungssensoren etc. Letztere sorgen beispielsweise dafür, dass sich die berührungsempfindliche Oberfläche des Gerätes (Touchscreen) automatisch deaktiviert, wenn es ans Ohr gehalten wird. Zunehmend wird in mobilen Endgeräten auch mit biometrischen Erkennungstechnologien experimentiert. Mit einem Bewegungssensor oder Mikrofon kann z.B. das Gehen, Laufen, Sprechen oder Fahren gemessen werden. Durch bestimmte Algorithmen lassen sich dann Gang-, Sprach- oder Fahrprofile erstellen und im Smartphone oder anderen mobilen Endgeräten speichern, d.h. Smartphones erkennen ihre Besitzer am Gehen, Sprechen oder Fahren.<sup>15</sup>

Die Sensorik und die Biometrie spielen folglich eine entscheidende Rolle für die Informationsgewinnung aus der umgebenden Umwelt, aus Prozessen und Verfahren in der Medizin, aus der Robotik oder der Fahrzeugtechnik sowie der Haushalts- und Bürotechnik. Die damit einhergehende stetig steigende Informationsflut der smarten Verbindung von Gegenständen wird das bisherige Vorstellungsvermögen vieler Menschen sprengen und neue Möglichkeiten der Vernetzung mit sich bringen. Die somit generierten Datenmengen können nur noch mit spezialisierter Hard- und Software verarbeitet werden. Sowohl für die Wirtschaft als auch für die Wissenschaft bieten die Technologien viel Raum für Experimente, Forschung und Entwicklung, gepaart mit hohen Wachstumsaussichten.

All die einzelnen Datenpunkte stellen neue und vor allem steigende Anforderungen an bestehende IT-Netzwerkarchitekturen, um die wachsenden Datenmengen möglichst rasch, flexibel und zu niedrigen Kosten verarbeiten zu können. Die Zukunft wird für jene Entscheidungsträger lukrativ werden, die diese Daten nicht nur erheben und speichern, sondern in maschinenlesbare Strukturen umwandeln, Muster erkennen und die richtigen Schlüsse ziehen.

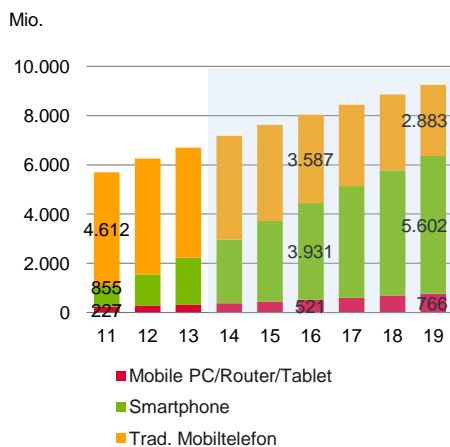
Um aber die gewünschten Effizienzsteigerungen zu erreichen, müssen zuerst sämtliche Datenquellen a) in ein maschinenlesbares Format gebracht werden und b) miteinander in Verbindung gesetzt werden. Denn nur so können die unterschiedlichen Daten mit weiteren Standortdaten von Smartphones oder RFID-Technologien, mit Transaktionsdaten aus dem Online-Handel oder mit Profildaten aus sozialen Netzwerken korreliert werden. Daraus können dann neue ungeahnte, versteckte und überraschende Zusammenhänge herausgefiltert werden, die den jeweiligen Akteuren einen Mehrwert stiften werden.

### Die IKT-Konvergenz als Basis für ungeahntes Innovationspotenzial

Künftig wird die Vernetzung von Alltagsgegenständen, die miteinander kommunizieren und den Menschen im Alltag einen höheren Nutzen stiften, vermehrt in den Mittelpunkt rücken. Die jeweiligen Technologien werden in die Umgebung und in die tägliche Handlung des Menschen möglichst flexibel eingebunden und bieten dadurch als moderne Schnittstelle zwischen Mensch und Maschine neue (digitale) Geschäftsmodelle. Die Verschmelzung von IKT geht auch einher mit der Diskussion um den Begriff „Internet der Dinge“. Dahinter verbirgt sich die Vision, dass das Internet eine Brücke zwischen der realen und der virtuellen Welt schlägt, und viele Alltagsgegenstände aus den Arbeits- und Lebenswelten

Weltweiter Bestand an mobilen Endgeräten

17



<sup>15</sup> Dadurch kann sich der Besitzer eines Smartphones relativ leicht für diverse Transaktionen oder Aktivitäten im Netz authentifizieren ohne die Eingabe von Passwörtern oder Codes, weil individuelle Bewegungsmerkmale einzigartig sind. Siehe hierzu auch Dapp, T. (2012). Der vermessene Mensch. Biometrische Erkennungsverfahren und mobile Internetdienste, Seite 7ff.



## Big Data – die ungezähmte Macht

---

der Menschen „smart“ zusammenführt (Smart Home, Smart City, Smart Glass, Smart Grid, Smart Car, Smart „Everything“). Das können Gegenstände wie Smartphones, Tablets, häusliche Elektronik, Fahrzeuge, Ampeln, Parkplätze, Sportgeräte, Kleidung, aber auch Gebäude, Liefercontainer und Maschinenanlagen aus der Industrie sein (auch bekannt unter dem Begriff Industrie 4.0).

Die Gegenstände werden durch Programmierbarkeit, Speichervermögen, Sensoren und Kommunikationsfähigkeit intelligent und können u.a. web-basiert sowie kabellos eigenständig Informationen austauschen. Durch die gemeinsame Kommunikation können diverse Aktionen ausgelöst sowie wechselseitige Steuerung vorgenommen werden. Dabei dienen die mit dem Internet verbundenen Gegenstände auch als physische Zugangspunkte zu diversen Internetdiensten, womit der zunehmende Wunsch vieler Menschen nach digitaler Mobilität befriedigt werden kann. In diesem eher jungen Forschungsfeld schlummern noch weitreichende und ungeahnte Möglichkeiten, um künftige Wertschöpfung voranzutreiben, wie folgende Beispiele zeigen:

### Waschmaschinen kommunizieren mit Kleidung

Beispielsweise können im Bereich der häuslichen Umgebung web-basierte Jalousien oder vernetzte Heizkörper sowie Klimaanlage via Smartphone ferngesteuert werden. Waschmaschinen aktivieren künftig ihren Waschvorgang automatisch erst dann, wenn der Strom gerade günstig ist. Thermometer in der Waschmaschine sind in der Lage, die Wassertemperatur so zu steuern, dass die Enzyme im Waschmittel optimal wirken können. Intelligente Kleidung kann der Waschmaschine übermitteln, dass sie nur mit 40° und im Schonwaschgang gewaschen werden darf und die Farbe Blau hat.

### Im Alter noch selbstbestimmt handeln können

Wenn es darum geht, im hohen Alter noch selbstbestimmt und möglichst lange in den eigenen vier Wänden wohnen zu können, bieten Sensoren bzw. kognitive Assistenzsysteme künftig aktive Unterstützung in den Bereichen Mobilitätserhaltung, physische Assistenz, Kommunikationsunterstützung, medizinisches Monitoring sowie medizinische Behandlung.<sup>16</sup> So lassen sich Bewegungssensoren in den Fußboden einarbeiten, die automatisch einen Notalarm auslösen, sobald ein Mensch stürzt. Monitoring-Systeme können permanent medizinisch relevante Daten, wie Blutdruck, Herzschlag, Sauerstoffsättigung des Blutes oder den Stresslevel eines älteren Patienten erfassen. An einem Badezimmerspiegel angebrachte Miniatur Sensoren werden demenzkranke Menschen daran erinnern, dass sie ihre Medikamente noch einnehmen oder sich noch die Zähne putzen müssen. Mit Hilfe von Sensoren kann z.B. auch gemessen werden, wenn Menschen beim Führen eines Pkw einen Schweißausbruch bekommen oder sonstige Symptome zeigen, die auf eine plötzliche körperliche Störung bis hin zum Herzinfarkt oder eine Krankheit hinweisen. Damit lassen sich dann Technologien entwickeln, die auf diese erfassten Muster reagieren und z.B. die Geschwindigkeit drosseln, das Fahrzeug via Autopilot sicher am Straßenrand zum Halten bringen und automatisch den Notdienst alarmieren.

### Vernetzungspotenzial in der Industrie

Ein ähnliches Vernetzungspotenzial existiert auch bei Maschinenanlagen und Produktionsketten in branchenübergreifenden Industrien. Im Automobilsektor wird beispielsweise in Bezug auf das „vernetzte Fahren“ daran gearbeitet, dass die Elektrik und Elektronik im Fahrzeug mit dem Internet verbunden werden. Die

---

<sup>16</sup> Vgl.: Beck, S., M. Grzegorzek et. al. (2013). Mit Robotern gegen den Pflegenotstand.



## Big Data – die ungezähmte Macht

Mobilitätsbedürfnisse der Fahrzeughalter ändern sich. Sie erwarten in ihren Fahrzeugen einen ähnlichen Komfort sowie ähnliche Funktionen wie bei mobilen und vernetzten Endgeräten. Künftig zählen also nicht nur Qualität, Sicherheit und zuverlässige Leistung des Fahrzeugs, der digitale Alltag der Menschen soll in einer möglichst personalisierten Art und Weise mit dem Führen eines Fahrzeugs verbunden werden. Im Vordergrund dieser Entwicklung stehen mehrere Trends: Zum einen wird die Vision des automatisierten Fahrens mittelfristig Realität, zum anderen werden sich langfristig effizientere, emissionsärmere Fahrzeuge durchsetzen. Durch den Brückenschlag zwischen industrieller und digitaler Welt werden neue Geschäftsmodelle den Markt erobern, welche moderne und vor allem personalisierte Dienstleistungen für den Fahrer in den Mittelpunkt stellen. Dazu zählen z.B. die cloud-basierte Spracherkennung, der Echtzeit-Austausch von Verkehrsflussdaten oder das vorausschauende, durch Online- und Navigationsdaten unterstützte Fahren, wie automatisiertes Bremsen, Gas geben und sogar lenken (teilautomatisierter Autopilot).

### Alltagsgegenstände bekommen eine eigene Internetadresse

Wenn künftig immer mehr Alltagsgegenstände über das Internet ansprechbar und kontrollierbar werden, genügen die Kommunikationsfähigkeiten der RFID-Technologien nicht mehr aus, um das ganze Konvergenz-Potenzial optimal zu nutzen. Die Alltagsgegenstände bekommen daher künftig eine eigene IP-Adresse, um mit anderen smarten Objekten und Netzknoten kommunizieren zu können. Aufgrund der großen Zahl benötigter IP-Adressen wird bei dieser Entwicklung künftig auf die neuere IPv6<sup>17</sup>-Version mit 128-Bit-Adressen zurückgegriffen, die allmählich die bisherige IPv4-Version ablöst. Das bedeutet aber auch, dass weitere Anstrengungen im Bereich Infrastruktur und Netzausbau unternommen werden müssen.

### Datenschutz-Experten warnen

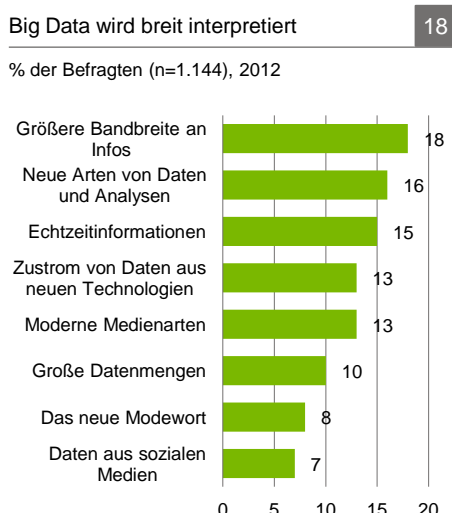
Allerdings beurteilen Datenschutzexperten die Entwicklung zum Internet der Dinge durchaus kritisch, weil viele Sensoren auch ohne Einwilligung der Menschen Daten an unterschiedliche Akteure zur Auswertung weiterleiten können. Somit wird schon heute die informationelle Selbstbestimmung verletzt.

## 4. Zur Rolle digitaler Ökosysteme in der Big-Data-Welt

Der Begriff Big Data wird je nach Verwendung unterschiedlich diskutiert und interpretiert. Ein Reduzieren auf kommerzielle Interessen des Handels oder die Möglichkeiten, die mit Social-Networking-Anwendungen digitaler Ökosysteme, wie Amazon, Twitter, Facebook, Apple oder Google, möglich sind, wird dem Begriff sicherlich nicht gerecht.

Denn Big Data bietet auch Potenzial, gesellschaftliche Probleme, z.B. im Bereich der Medizin, des internationalen Cybercrimes oder im Klimawandel zu lösen oder zumindest zu lindern. Es steht auch für einen grundsätzlichen Wandel in der Wissenschaft. Wenn Forscher bisher vorgefasste Modelle genutzt haben, um die Realität u.a. mit Stichprobenanalysen abzubilden, können zusätzlich datengetriebene Beobachtungen in Echtzeit mit einer weitaus größeren Grundgesamtheit ohne große Kosten herangezogen werden, um Prognosen mit mehr Informationen anzureichern. Im IT-Sicherheits- oder Finanzbereich bieten sich z.B. neue Möglichkeiten, Kreditkartenbetrüger schneller dingfest zu machen<sup>18</sup> bzw. Risiken und Compliance-Angelegenheiten besser einzuschätzen; die Produktion eines Unternehmens im Verarbeitenden Gewerbe kann mit mehr Effizienz gefahren werden; Logistikketten lassen sich optimieren sowie sensible Infrastrukturen sicherer gestalten. Medizinische Forschungsergebnisse lassen sich mit anderen Forschungsfeldern kombinieren, um eventuelle versteckte Er-

### Die gute Seite der Macht



Quellen: IBM Institute for Business Value, Saïd Business School Oxford

<sup>17</sup> <http://de.wikipedia.org/wiki/Ipv6>.

<sup>18</sup> Aus einer Vielzahl von Daten zur Kreditkartenkriminalität wird das Verhaltensmuster von Kreditkartenbetrüger herausgefiltert, um es mit aktuellen Fällen zu vergleichen. Das Ziel ist, herauszufinden, mit welcher Wahrscheinlichkeit ein Betrüger seine nächste Tat begeht.





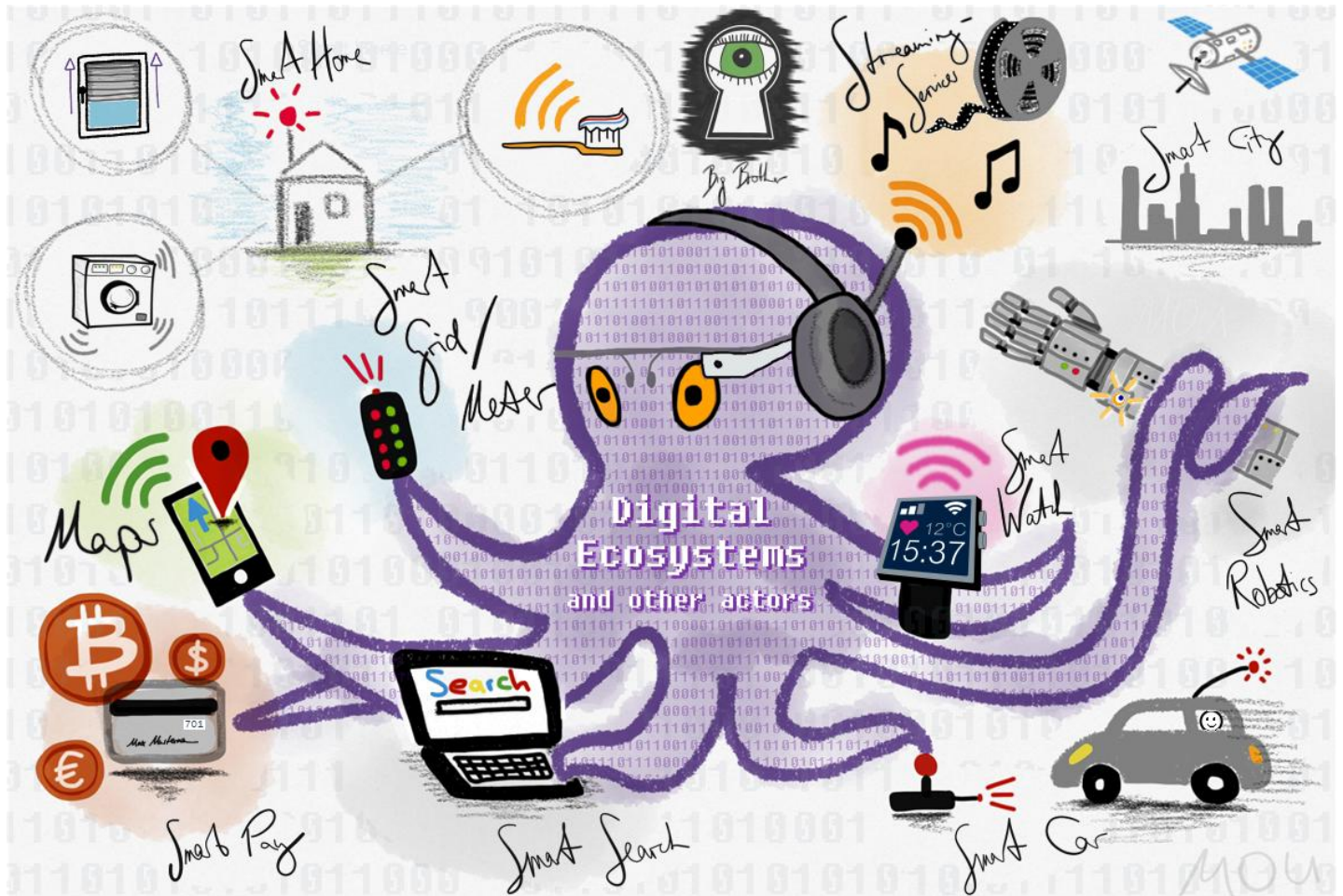
## Big Data – die ungezähmte Macht

kenntnisse zu gewinnen und/oder seltene Krankheiten zu heilen. Darüber hinaus lassen sich Kosten dauerhaft senken, die Entscheidungsgrundlage optimieren, das Konsumverhalten und die Zufriedenheit der Kunden genauer analysieren und/oder zielgruppenspezifische Produkte und Dienstleistungen erstellen.

### Die dunkle Seite der Macht

Big Data haftet aber auch eine dunkle Seite (der Macht) an. Im Zuge der Big-Data-Debatte werden zunehmend auch Orwell'sche Big-Brother-Methoden à la „1984“<sup>19</sup> bzw. unverhältnismäßige Überwachungsvorhaben diskutiert. Hier entstand auch die Vision des gläsernen Menschen. Im Gegensatz zu Orwell dominiert in der Big-Data-Diskussion allerdings kein diktatorisches System und die Menschen partizipieren zunächst freiwillig, während sich im Hintergrund diverse Interessen unterschiedlicher Akteure (Handel, Wissenschaft, NGOs, Geheimdienste etc.) tummeln. Im Fokus steht aber die durchaus nachvollziehbare Sorge, dass Bürger und Konsumenten zunehmend zu einem Spielball einiger Akteure werden bzw. sich (unfreiwillig) zum gläsernen Individuum entwickeln. So scheint Orwells Roman z.B. mit Blick auf die Geheimdienst-Affäre einigen Beobachtern näher an der Realität als jemals zuvor.

### Digitale Ökosysteme bescheren uns „Smart Everything“



Grafik: Oliver Ullmann. Deutsche Bank Research.  
Quelle: Dapp, T. (2014). Big Data – die ungezähmte Macht. Deutsche Bank Research. Frankfurt am Main.

<sup>19</sup> Der Roman von George Orwell (1949) wird häufig genannt, wenn unverhältnismäßige, staatliche Überwachungsmaßnahmen diskutiert werden.



## Big Data – die ungezähmte Macht

Die Ziele digitaler Ökosysteme sind weitreichender als von vielen angenommen

Einerseits entstehen also viele nützliche Produkte und Dienste, die den Menschen bei alltäglichen und routinemäßigen Handlungen unterstützen sollen, wie die oben erwähnten Sensorik-Beispiele im Pflegebereich. Andererseits stehen aber digitale Ökosysteme im Verdacht, Geschäftspraktiken zu betreiben, die nicht ausschließlich den primären Nutzen des Produktes oder die Dienstleistung für den Kunden in den Mittelpunkt stellen. So können einmal erhobene Daten teils zweckentfremdet und ohne den Nutzer zu informieren mehrfach genutzt bzw. monetarisiert werden.

Viele Monetarisierungsstrategien

Wenn z.B. Google eine web-basierte Brille anbietet oder an der Entwicklung autonom fahrender Fahrzeuge forscht, dann umfasst dies mehr als *nur* das Angebot eines selbstfahrenden Autos und mehr als *nur* das Angebot einer mit dem Internet verbundenen Brille. Beide Produkte erlauben einen beinahe schrankenlosen Zugriff auf teils intime und persönliche Daten. Ein einzelnes Unternehmen bekommt Einblicke in die täglichen Handlungsweisen der Menschen, deren Aktivitäten, Neigungen und Identitäten. Eine web-basierte Brille offenbart z.B., wohin die Blicke der Menschen zielen, wie lange und wie oft sie ihre Blicke auf Gegenstände, Werbung oder Menschen richten. Damit erhält das Unternehmen die Möglichkeit, die Big-Data-Analyse-Instrumente optimal in Echtzeit einzusetzen, um wertvolle Muster aus den gesammelten Daten zu ziehen, diese gegebenenfalls zu monetarisieren bzw. Menschen individuelle Werbebotschaften zu senden. Die Begehrlichkeiten der großen Internetunternehmen, diese zusätzlichen Informationen für unterschiedliche Zwecke mehrfach zu nutzen, nehmen zu, während die gesellschaftlichen Folgen (bis jetzt) weitgehend unerforscht sind.

Automatisiertes Fahren: Einst eine Vision, heute schon Realität

Auch das autonom fahrende Fahrzeug wird diejenigen, die es nutzen, völlig gläsern erscheinen lassen.<sup>20</sup> Während des Gebrauchs, aber auch über die gesamte Nutzungsdauer hinweg, können alle möglichen Informationen gemessen, gesammelt und anschließend ausgewertet werden. Dabei werden mittelfristig natürlich Muster zu beobachten sein, die folgende Fragen künftig beantworten können: Wohin fahren die Menschen? Zu welcher Tages- und Nachtzeit nutzen sie das Fahrzeug? Wo arbeiten sie und was sind beliebte Freizeitziele? Wie lange sitzen sie durchschnittlich im Fahrzeug, welche Musik wird gehört, welche Temperaturen werden präferiert, mit wem wird kommuniziert/telefoniert und welche digitalen Inhalte werden während der Fahrt konsumiert? Halten sie sich an die vorgeschriebene Geschwindigkeit, konsumieren sie während der Fahrt Lebensmittel/Getränke/Tabak etc? Jede beschriebene Aktion könnte folglich von einem Unternehmen monetarisiert werden. Mit jeder Aktion wird Geld verdient, indem entweder maßgeschneiderte Produkte und Dienste aus einer Hand angeboten werden oder die persönlichen Daten an interessierte Akteure branchenübergreifend weiterverkauft werden. Vor allem gelingt es so, die Menschen künftig auch im realen Leben auf Schritt und Tritt zu begleiten und somit die entscheidende Brücke zu schlagen zwischen virtueller und realer Welt.

Die Zahlungsströme müssen also nicht zwingend von den Endkonsumenten eines gewissen Angebots direkt an den Anbieter fließen. Stattdessen kann dieser seine Einnahmen auch dadurch maximieren, dass die gesammelten Daten dritten Unternehmen zur Verfügung gestellt werden. Für Start-Ups oder Nischenanbieter wird es somit möglich sein, sich an Ökosysteme mit hohen Marktanteilen anzudocken, um komplementäre Produkte und Dienste anzubieten, die das Leistungsangebot für den Nutzer noch attraktiver machen. Denkbar sind natürlich auch Übernahmen durch digitale Ökosysteme.<sup>21</sup> Somit ändert sich

<sup>20</sup> Morozov, E. (2013). Big Data. Warum man das Silicon Valley hassen darf. FAZ.10.11.2013.

<sup>21</sup> Siehe z.B. <http://www.spiegel.de/wirtschaft/unternehmen/google-kauft-nest-labs-fuer-3-2-milliarden-dollar-a-943362.html>.



## Big Data – die ungezähmte Macht

auch die Wettbewerbssituation und einzelne Unternehmen bauen ihre Marktstellung aus.

### Das Silicon Valley als globaler Innovationstreiber

#### Die Ökosystemisierung<sup>23</sup> im Internet ...

Die digitalen Fußabdrücke der Internet-Nutzer werden immer deutlicher, präziser und leichter aufzuspüren. Alleine auf Plattformen der beliebten digitalen Ökosysteme wie Google, Twitter, Facebook, LinkedIn, Amazon oder Apple hinterlegen sie teils ohne Scheu private und intime Informationen, wie Beziehungsstatus, Freizeitinteressen, Reisegewohnheiten, Konsumgewohnheiten, präferierte Musik- und Filmgenre, Fotografien und „liken“ querbeet alles, was ihnen im Netz unter den Finger bzw. den Cursor kommt.

Jede gefolgte oder selbst erstellte Kurznachricht auf Micro-Blogging-Diensten steht für eine individuelle Präferenz. Durch die Auflistung der gesamten Tweets kann relativ einfach ein Profil über politische oder gesellschaftliche Neigungen der Individuen erstellt werden. Auf Facebook oder vergleichbaren sozialen Netzwerken können anhand der geposteten Beiträge und der Vernetzung zu anderen Nutzern relativ einfach Aufenthaltsorte, Interessen, Hobbies oder sexuelle Neigungen abgelesen und zusammengeführt werden.

#### ... schreitet branchenübergreifend voran

Viele Internetnutzer kommen an einigen digitalen Ökosystemen kaum noch vorbei, egal, ob sie privat oder beruflich im Internet surfen. Für alle möglichen Lebens- und Arbeitsbereiche werden nützliche und beliebte Internet-Dienste „aus einer Hand“ angeboten, z.B. Suchmaschinendienste, Betriebssysteme für diverse Endgeräte, E-Mail-Dienste, Navigation, Cloud-, Streaming-, und Wiki-Dienste sowie eigene Märkte für digitale Güter (iStore, Playstore, etc.). Selbst ohne eine Registrierung bzw. ohne ein E-Mail-Konto können die Suchanfragen der Nutzer gespeichert und ausgewertet werden, um Profile zu erstellen. Dahinter verbirgt sich u.a. das Ziel, so viele Informationen (z.B. personenbezogene Daten) wie möglich digital zu organisieren. Jedoch entfaltet sich das gesamte Nutzungspotenzial erst dann, wenn sich die Internetnutzer registrieren.

#### Im Internet werden pro Minute...

19

|   |
|---|
| 100 Stunden Videomaterial auf YouTube hochgeladen |
| 4.112.500 Google-Suchanfragen getätigt            |
| 3.300.000 Einträge auf Facebook geteilt           |
| 347.000 Tweets auf Twitter versandt               |
| 48.000 Apps bei Apple heruntergeladen             |
| 38.200 Fotos auf Instagram geteilt                |

Quellen: Youtube, Google, Facebook, US Securities and Exchange Commission, Apple, Instagram

Der Ratschlag zur Datensparsamkeit, sollte sie bevorzugt werden, bietet auch keine Lösung – vor allem, weil auch die Nicht-Teilnahme in der digitalen Welt Spuren hinterlässt, die Rückschlüsse auf einzelne Personen und deren Verhalten erlauben (z.B. durch Re-Anonymisierungsmaßnahmen). Es zeigt sich, dass die Anonymisierung bei der Weitergabe personenbezogener Daten nicht ausreicht, um Privatsphären zu schützen, denn bereits einzelne Suchanfragen (evtl. kombiniert mit weiteren Daten) lassen exakte Rückschlüsse auf die Urheber der Daten zu – auch dann, wenn Username oder IP gelöscht wurden, so geschehen bei der anonymisierten Weitergabe von Suchanfragen bei AOL.<sup>22</sup>

### Das Internet dient als Spielwiese für eingezäunte, digitale Ökosysteme

#### Starker Lock-In-Effekt

Nicht nur Google, sondern auch weitere Internetfirmen, wie Apple, Facebook oder Amazon, werben mit vielen attraktiven, modernen (sozialen) und teils kostenfreien IKT-Diensten. Diese Ökosysteme profitieren von einer relativ hohen, loyalen Kundenzahl im dreistelligen Millionenbereich und verfügen über ausreichend liquide Mittel, mit denen sie weitere Innovationen vorantreiben können. Zudem erhöhen sie Attraktivität, indem sie für eine hohe Verfügbarkeit von zusätzlichen, komplementären Internet-Diensten und Apps innerhalb ihrer umzäunten Spielwiesen sorgen. In regelmäßigen Abständen werden komplementäre Internet-Dienste mittels strategischer Allianzen mit Dritt- oder Nischenanbietern oder durch Übernahmen auf dem Markt angeboten. Die Innovationsrate

<sup>22</sup> [http://www.nytimes.com/2006/08/09/technology/09aol.html?\\_r=0](http://www.nytimes.com/2006/08/09/technology/09aol.html?_r=0).

<sup>23</sup> Vgl.: Bahr, F. et al. (2012). Schönes neues Internet? Chancen und Risiken für Innovation in digitalen Ökosystemen. Seite 3ff.





## Big Data – die ungezähmte Macht

dieser technologiegetriebenen Internetunternehmen ist zumindest beachtlich und sucht auf deutscher bzw. europäischer Ebene ihresgleichen.

Zudem ist zu beobachten, dass die großen Internet-Plattformen ihre Fühler zunehmend auch außerhalb ihrer ohnehin größer werdenden Kernkompetenz ausstrecken und somit branchenübergreifend bestehende Wettbewerbskonstellationen aufwirbeln und traditionelle Unternehmen unter Druck setzen. So investiert Amazon in eigene mobile Zahlungsverkehrslösungen, in eigene TV-Produktionen und Cloud-Dienste für Premiumkunden<sup>24</sup>, während Google verstärkt im Bereich Haushaltsgegenstände und Robotik investiert und sogar eine neue Sparte mit humanoider Roboterentwicklung ins Leben ruft.<sup>25</sup> Digitale Ökosysteme sind und bleiben beliebte und akzeptierte Plattformen. Durch die Nutzung der Dienste digitaler Ökosysteme entstehen Lock-in-Effekte, die relativ starke Wechselkosten nach sich ziehen. So steigen die Möglichkeiten für die Plattformen, die individuelle Aufmerksamkeit bzw. die persönlichen Daten der Nutzer im Netz mit Hilfe von Walled-Garden-Strategien zweckzuentfremden bzw. zu monetarisieren, weiter an.<sup>26</sup>

### Ambivalentes Verhalten der Menschen im Internet

Bei all der durchaus berechtigten Kritik an digitalen Ökosystemen zeigt die Realität aber, dass viele Internetnutzer (noch) bereit sind, partielle Freiheitsgrade an die Plattformbetreiber abzugeben, um die bevorzugten Internetdienste bequem und aus einer Hand zu konsumieren. Dabei ist die Präferenz hinsichtlich Kostenfreiheit und Komfort offensichtlich häufig so groß, dass der Nutzer unwiderrprochen persönliche und intime Daten von sich preisgibt, die Weitergabe an Dritte erlaubt, die heimliche Speicherung biometrischer Scans von Sprache und Gesicht zulässt sowie personalisierte Werbebotschaften in Kauf nimmt.

### Manipulation und Intransparenz

Die Nutzer solcher Plattformen haben kaum Kontrolle über die Sicherheit des angebotenen Systems, über eventuelle Zugriffe auf persönliche Daten, über die Sicherheit und Verwendung ihrer Daten sowie deren Löschung. Dennoch überlassen sie den großen Internet-Plattformen in den meisten Fällen die Kontrolle über ihre Daten und gehen davon aus, dass die Plattformen für die nötige Sicherheit sorgen und die Nutzer vor Missbrauch schützen. Das Verhalten der Menschen auf digitalen Kanälen ist also durchaus ambivalent. Kritiker betonen vermehrt, es erhöhte sich auch der Verdacht, dass der Internetnutzer selbst zum gehandelten Gut wird.

Wofür nutzen die Menschen ihr Mobilfunkgerät?

20

|     |  |
|-----|--|
| 54% | surfen im Internet                     |
| 37% | halten sich in sozialen Netzwerken auf |
| 17% | shoppen online                         |
| 21% | teilen Dokumente, Videos etc.          |
| 26% | schauen Videos und hören Musik         |
| 16% | tätigen mobile Bezahlung               |

n=4.500 aus 9 ausgewählten Ländern (DK, FR, DE, IT, NL, PL, RU, SE, GB)

Quelle: Norton, 2012

Für viele Unternehmen bedeutet dies vor allem ein lukratives Geschäft. Der Großteil der Konsumenten wird weiterhin die Vielfalt an Services und personalisierten Diensten trotz etwaiger Datenschutzbedenken nutzen. Einige werden allmählich den Eindruck gewinnen, dass sie zu einem hohen Grade manipulierbar und transparent geworden sind, oder sich sogar ihrer Datenhoheit beraubt fühlen. Es bleibt zudem abzuwarten, ob die Menschen den relativ hohen Preis dieser individualisierten Dienste (umfassendes, persönliches, digitales Profil) mittel- bis langfristig weiterhin kritiklos zahlen oder versuchen, ihre Nachfrage nach Diensten und Produkten mehr und mehr zu verweigern bzw. zu substituieren. Aufgrund der momentanen marktlichen Wettbewerbssituation stellt sich jedoch zu Recht die Frage, ob sich bei einigen Internet-Diensten mittelfristig überhaupt ernstzunehmende Substitutionsmöglichkeiten auf dem Markt für den

<sup>24</sup> Beispiel: Amazon steigt in den Markt für mobile Zahlungsverkehrslösungen ein. <http://www.businessinsider.com/report-amazon-has-bought-square-competitor-gopago-2013-12>.  
<sup>25</sup> <http://www.zeit.de/digital/2013-12/google-roboterfirma>. sowie <http://www.zeit.de/wirtschaft/unternehmen/2013-12/google-roboter-android-erfinder-andy-rubin>.  
<sup>26</sup> Vgl.: Dapp, T. (2013). Die Zukunft des (mobilen) Zahlungsverkehrs: Banken im Wettbewerb mit neuen Internet-Dienstleistern. Seite 21ff.



## Big Data – die ungezähmte Macht

### Schiefelage in der Datenhoheit

Kunden etablieren werden/können und ob die Adaptionsgeschwindigkeit moderner Technologien sinkt aufgrund der Praktiken.

Datenschützer sehen die Verknüpfung teils privater und intimer personenbezogener Daten zwischen einzelnen Internetfirmen unisono mit großer Besorgnis. Denn es wird zunehmend kritisch diskutiert, ob durch die eingezäunten Spielwiesen der digitalen Ökosysteme Innovationen, Kommunikations- und Informationsströme kontrolliert sowie Datenschutzbestimmungen und Nutzerautonomie unterminiert werden. Weiterhin merken Datenschützer an, dass auf intransparente Weise Nutzeraktivitäten zweckentfremdet monetarisiert werden.<sup>27</sup>

## 5. Der ökonomische Wert von Daten

### Wir haben einen „digitalen Zwilling“

*„Sie verstehen nicht oder sie wollen nicht wissen, dass sie so mit bauen an einem digitalen Zwilling ihrer realen Person, der neben ihren Stärken eben auch ihre Schwächen enthüllt – oder enthüllen könnte. Der ihre Misserfolge und Verführbarkeiten aufdecken oder gar sensible Informationen über Krankheiten preisgeben könnte. Der den Einzelnen transparent, kalkulierbar und manipulierbar werden lässt für Dienste und Politik, Kommerz und Arbeitsmarkt.“* [Bundespräsident Joachim Gauck, 3. Oktober 2013, Stuttgart]

### Die Kombination aus diversen Datensätzen verspricht neue Erkenntnisse

Für eine optimale Analyse im Bereich Big Data sind einzelne Datensätze im Besitz eines einzelnen Unternehmens, isoliert betrachtet, nur bedingt nützlich. Erst, wenn mehrere Datenpunkte teils aus unterschiedlichen Quellen zusammengeführt werden, können bestimmte Muster herausgefiltert werden. Nur wenige Unternehmen bieten so viele diverse Produkte und Dienste aus einer Hand an, um ein breites Spektrum an unterschiedlichen Kundeninformationen unter einem Dach zu versammeln.<sup>28</sup> Sofern es sich um personenbezogene Daten handelt, können dadurch umfassende und relativ detailgetreue Profile von Menschen erstellt werden. Die unterschiedlichen Daten, die gespeichert, archiviert und ausgewertet werden, stammen größtenteils aus digitalen Werbe-, Informations-, Transaktions- oder sonstigen Netzwerkkanälen. Relativ einfach lassen sich die Informationen über die jeweiligen Internetprovider, IP- oder E-Mail-Adressen sowie verwendete Suchmaschinen sammeln. Bei jedem Besuch auf der Amazon-Plattform werden diese Informationen z.B. gespeichert und in Kombination mit dem jeweiligen historischen Klick- und Konsumverhalten der Menschen ausgewertet. Nur so kann Amazon seinen Online-Kunden auch auf fremden Internetadressen bei jedem Besuch individualisierte Kaufempfehlungen anbieten, bzw. nur so funktioniert die Individualisierung im Internet per se.

Durch die wachsende Datensammlung der Akteure lassen sich dann mit relativ geringem Aufwand virtuelle und reale Welt verknüpfen, sodass auch Wohnadressen und Telefonnummern der Individuen relativ leicht zu identifizieren sind. Darüber hinaus sind weitere Rückschlüsse, z.B. auf Immobilien- bzw. Mietpreise einer Wohngegend, Informationen über den Arbeitgeber und damit einhergehend Daten zum Gehalt (zumindest innerhalb branchenüblicher Gehaltsbänder) und zum verfügbaren Einkommen möglich. Was würde geschehen, wenn die persönlichen Daten zusätzlich mit weiteren Informationen aus sozialen Netzwerken, mit Daten zum Konsumverhalten, mit Ortungsdaten (GPS-Tracking), Daten zum Beziehungsstatus oder zum Einkommen in Relation gesetzt werden? Die sich ergebenden Datenkorrelationen erlauben wiederum neue, relevante Rückschlüsse auf die persönlichen Neigungen und Identitäten.

<sup>27</sup> Siehe Walled-Garden-Strategien in Kapitel 6 aus Dapp, T. (2013). Die Zukunft des (mobilen) Zahlungsverkehrs. Seite 21ff.

<sup>28</sup> Google ist hier sicherlich die Ausnahme, da das Unternehmen aufgrund seiner Reichweite und seiner überdurchschnittlich hohen Datenmenge in der Lage ist, eigene Daten zu kombinieren.





## Big Data – die ungezähmte Macht

### Digitale Fußabdrücke werden zu handelbaren und begehrten Profilen

Personenbezogene Daten haben einen hohen Preis

Denn sie wissen, was wir tun!

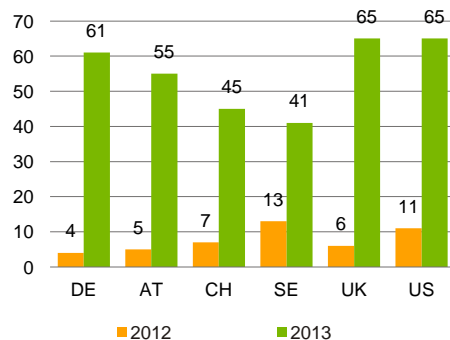
Stellen Sie sich folgendes Szenario vor: Sie gelangen aufgrund einer Datenpanne oder dank eines Whistleblowers zufällig an einen Datensatz zu Ihrer Person. Ihr Profil weist ziemlich detailgetreue Informationen auf, die von einem auf Datenerhebung spezialisierten Unternehmen oder von einem „befreundeten“ Geheimdienst stammen. In Ihrem Profil lesen Sie dann folgendes über sich:

„X ist männlich, 43 Jahre, promovierter Ingenieur, verheiratet mit V, weiblich, 37 Jahre; X lebt mit Individuum V in Ein-Familienhaus in Frankfurt und hält eine Katze; X bezieht regelmäßig das Magazin „Schöner Wohnen“ sowie das digitale Abo des „Maschinenbauers“ direkt auf sein mobiles Endgerät der Marke A; X kauft regelmäßig Fach-Literatur, Kleidung und Elektronikgegenstände beim Onlinehändler A; X interessiert sich für diverse Umweltthemen und unterstützt mittels Dauerauftrag einige internationale NGOs; X interagiert mehrmals täglich auf diversen sozialen Netzwerk-Plattformen über mobile und stationäre Geräte und unterhält überdurchschnittlich viel Kontakt zu den Individuen H (männlich, 41 Jahre), S (weiblich, 29 Jahre), und K (weiblich 35 Jahre); X konsumiert Video-on-demand-Angebote und surft im Schnitt 1 bis 2 Stunden täglich ab 22:00 Uhr im Internet; X unterhält mehrere Kontoverbindungen zu den deutschen Banken D und P sowie der schweizerischen Bank U; X bevorzugt digitale, webbasierte Bezahlmethoden; X besitzt zwei Kreditkarten der Anbieter M und A, Kreditkarte M wird auch von Individuum V (Ehefrau) mitbenutzt, Umsätze laufender Monat der Kreditkarte A: Reiseführer Paris EUR 9,95, Flugreise nach Paris für 2 Personen EUR 379,00 Hotel Paris EUR 440,00 Gastronomie Paris EUR 90,00, Schmuck Paris EUR 399,00, Umsätze laufender Monat der gemeinsam genutzten Kreditkarte M: keine Umsätze; X kauft regelmäßig Sportartikel für Outdoor-Aktivitäten; X lässt sich vermehrt Angebote diverser Kfz-Anbieter zumailen (ist folglich im Begriff, ein Fahrzeug zu kaufen/zu leasen/zu mieten); X vergleicht Preise von Stromanbietern (wägt eventuell einen Stromanbieterwechsel ab); X bucht jährlich online Winterurlaube in Österreich und leiht im Voraus online die Ski-Ausrüstung bei zwei Anbietern (neigt ansonsten zu Fernreisen mit Individuum V und kurzen Städtereisen, zuletzt mit Individuum S (X unterhält höchstwahrscheinlich partnerschaftliche Beziehung zu Individuum S)<sup>29</sup>; X bezieht mehrmals im Quartal rezeptpflichtige Medikamente (X ist Allergiker, neigt zu hohem Blutdruck), X besucht regelmäßig unterschiedliche Foren und partizipiert an Diskussionen zu seltenen Krankheiten (X benutzt hierfür entweder den anonymen Avatar „neugierig“ oder „curious\_1970“). X besitzt Kfz-Haftpflicht-, Rechtsschutz und Hausratversicherung von Unternehmen A; X bezog 2012 juristische Dienste von Unternehmen A in einem Verkehrsdelikt-Prozess, der Prozess ist noch im Gange. Echtzeit-Zusatzinformation: Im Moment hält sich Individuum X in der Seitensprungstraße 7 auf. Hier wohnt Individuum S.“

Angst vor Datendiebstahl

21

Umfrage, in %



DE (n=571); AT (n=586); CH (n=476); SE (n=346); UK (n=435); US (n=409); Mehrfachnennungen möglich

Quelle: eGovernment MONITOR 2013

Ein derartiges Personenprofil ist hinsichtlich der vielen digitalen Fußabdrücke, die der Mensch im Netz mittlerweile (auch unbewusst) hinterlässt, sowie der hierfür notwendigen Datenanalyse-Instrumente sicherlich realistisch und technologisch erstellbar. Auf diese Weise ist es den Akteuren möglich, kontextübergreifend zeitliche Abläufe und digital erfasste Profile über Aufenthaltsorte, soziale Beziehungen, Konsum- und Mediennutzungsverhalten, Gesundheit, Einkommen, Beruf etc. zu erstellen.<sup>30,31</sup>

Jede einzelne Information in diesem fiktiven Profil hat einen ökonomischen Wert mit wertvollen Anknüpfungs- bzw. Monetarisierungspunkten für unterschiedliche

<sup>29</sup> An dieser Stelle wird nochmal auf die eingangs erwähnte Frage hingewiesen, ob Kreditkartenanbieter aus den einzelnen Transaktionen ihrer Kunden vorhersagen können, bei wem eventuell eine Ehekrise bevorsteht. Die Antwort lautet: Ja, sie können.

<sup>30</sup> Vgl.: Weichert, T. (2013). Big Data: Das neue Versprechen der Allwissenheit.

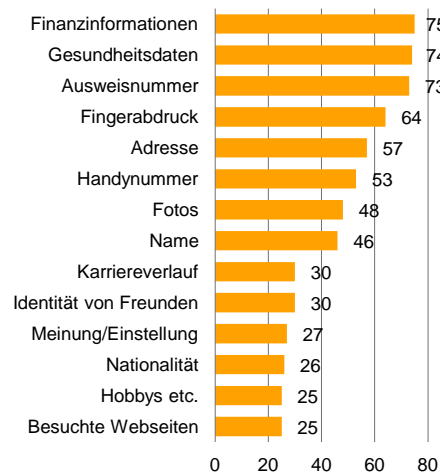
<sup>31</sup> Verwendet werden die englischen Begriffe „personalizing“, „scoring“, „profiling“ oder „tracking“.



## Big Data – die ungezähmte Macht

### Was sind personenbezogene Daten? 22

% der Befragten, EU 27, 2011 (n=26.574)



Quelle: Europäische Kommission

Akteure aus Wirtschaft (z.B. Versicherungsunternehmen, Konsumgüterhersteller oder diverse Dienstleister) Wissenschaft (Neurowissenschaften, Soziologie, Verhaltenswissenschaften, etc.) oder Politik (z.B. Finanzamt, öffentliche Verwaltungseinheiten, Geheimdienste). Die Menschen werden durch ihre digitalen Fußabdrücke zu transparenten (gläsernen) Individuen im Netz, ohne dass ihnen dies unter Umständen zu jedem Zeitpunkt ihrer Onlineaktivität bewusst wäre (z.B. (De)Aktivierung des GPS-Signals).

Die Summe dieser Informationen wird zu individuellen digitalen Profilen aggregiert, die je nach Angebot und Nachfrage einen Preis haben. Welcher monetäre Wert könnte einem einzelnen Kundenprofil bzw. diesem immateriellen Wert zugrundeliegen, vor allem, wie könnte er ermittelt werden? Ein einfaches Rechenbeispiel kann dies illustrieren: Die Marktkapitalisierung z.B. von Facebook beträgt derzeit ca. EUR 70 Mrd.<sup>32</sup> Facebook wirbt mit einer Anzahl an Facebook-Nutzern von über 1,2 Mrd.<sup>33</sup> Der monetäre Wert eines durchschnittlichen Facebook-Nutzerprofils könnte sich durch simple Division auf ca. EUR 58 belaufen. Unter der weiteren Annahme, dass nur ca. zwei Drittel der Konten tatsächlich aktiv sind, erhöht sich der Wert auf ca. EUR 88. Dieser kalkulierte Wert könnte die Basis für Verhandlungen über den Preis einzelner Nutzerprofile zwischen Investoren oder sonstigen Akteuren sein. Ob sich diese Annäherung allerdings auch für die Berechnung eines „Return on Investments“ eignet, sprengt den Rahmen dieser Studie, zeigt aber auf, dass es bei Big Data bzw. beim Handel von Daten zum großen Teil um monetäre Interessen geht.

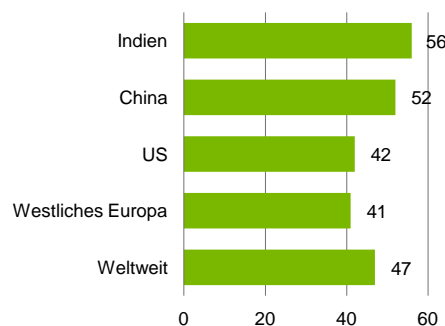
### Personenbezogene Daten werden zur virtuellen Währung im Netz

### Datengetriebener Alltag

Die Menschen bewegen sich also immer tiefer in einen datengetriebenen Alltag. Tägliche, routinemäßige Handlungen werden mit modernen web-basierten Technologien verbunden und erleichtern z.B. mit sogenannten Applikationen (Apps, Internet-Diensten) das Leben. Mit jedem Klick auf Onlineportalen, mit jedem Sprachbefehl auf mobilen Endgeräten oder jeder GPS-Ortung kann Zeit eingespart und können Suchkosten reduziert werden. Diese Dienste erhöhen somit die Effizienz und den Komfort (Convenience) im Alltag. Ihre Nachfrage ist ungebrochen. Die Dienste sind aber nicht umsonst und haben einen im Verhältnis durchaus hohen Preis. Denn die Nutzung vieler beliebter web-basierter Dienste ist zwar kostenfrei, wird aber mit der i.d.R. (freiwilligen) Preisgabe individueller, digitalisierter Persönlichkeit bezahlt, was viele Nutzer im Moment ihrer Entscheidung unberücksichtigt lassen. Denkbar sind künftige Geschäftsmodelle, die den eigentlichen Datenlieferanten, also den Internetnutzern selbst, auch einen monetären Nutzen in Aussicht stellen.

### Ungeschützte Daten: Globales Problem 23

%, geschätzter Anteil ungeschützter Daten an schutzwürdiger Gesamtdatenmenge, 2012



Quelle: IDC Digital Universe Study

Gerade die Nutzung personenbezogener Daten weckt bei den Akteuren viele diverse Begehrlichkeiten und nährt die Hoffnung auf steigende Umsätze. Personenbezogene Daten können einem Individuum direkt zugeordnet werden. Hierbei handelt es sich z.B. um Konto- oder Bonitätsdaten, Kreditkarteninformationen, medizinische Diagnose- oder Gesundheitsdaten, Daten zu Konsum- und Mediennutzungsverhalten, Daten aus Videoüberwachungen oder biometrischen Erkennungsverfahren, Nutzungsdaten aus Applikationen (Apps) oder aus sozialen Netzwerken, Sensordaten mobiler Endgeräte, Standortdaten, E-Mails, Chat- oder anderen Internet-Kommunikationsdaten. Viele Menschen sind sich der möglichen Gefahr (noch) nicht bewusst, dass eine Vielzahl unterschiedlicher Akteure und Algorithmen mit heterogenen Interessen und Zielsetzungen digitale Informationen sammeln und speichern, um sie zu einem späteren Zeitpunkt auszuwerten und eventuell zu monetarisieren. Im Internet werden die Menschen an unterschiedlicher Stelle und zum Großteil unbewusst individuell umworben

<sup>32</sup> Stand: 24.01.2014 auf <http://www.onvista.de/aktien/Facebook-Aktie-US30303M1027>.

<sup>33</sup> [http://allfacebook.de/zahlen\\_fakten/infografik-10-unglaubliche-zahlen-zur-facebook-nutzung-10-jahre-facebook](http://allfacebook.de/zahlen_fakten/infografik-10-unglaubliche-zahlen-zur-facebook-nutzung-10-jahre-facebook).



## Big Data – die ungezähmte Macht

(z.B. Werbebanner bei Internet-Providern), mit Hilfe von kleinen Softwareprogrammen online verfolgt und ausgespäht (Cookies<sup>34</sup>). Sind die Daten erst mal gespeichert, können sie auch erst Jahre später wieder auftauchen und im schlimmsten Fall beruflichen Karrieren oder privaten Beziehungen dauerhaft schaden. Denn Profile lassen sich natürlich auch manipulieren, um Menschen bewusst zu schaden oder zu diskriminieren, z.B. beim Abschluss einer bestimmten Versicherung, bei Bewerbungsprozessen oder durch die Verweigerung weiterer Dienste.

Solange es also keine Lösungen gibt, informierte Entscheidungen zu treffen um die Herausgabe privater, persönlicher Daten auch zu verweigern oder sie löschen zu lassen, scheint die Entwicklung dieses Teilaspekts von Big Data hinsichtlich des Datenschutzes alles andere als erfreulich.

### Personenbezogene Daten wecken viele Begehrlichkeiten

#### Wer kontrolliert die vielversprochenen Algorithmen?

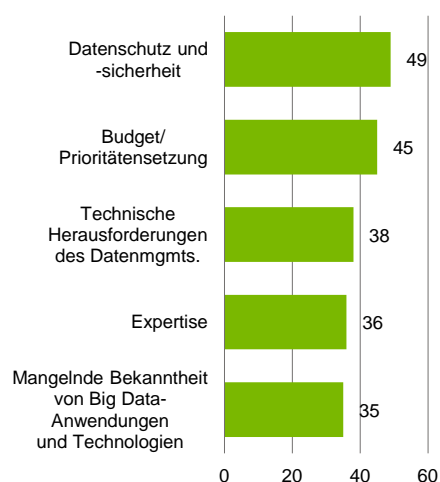
Big Data wird inzwischen von vielen Akteuren aus Politik, Wirtschaft oder Wissenschaft aus unterschiedlichem Interesse propagiert. Bei all den lukrativen Wachstumschancen, die Big Data bzw. das breite Angebot moderner web-basierter Technologien mit sich bringen, werden Risiken und Probleme gerne ausgeklammert. Insbesondere in die neuen, hochgehandelten Algorithmen zur Reduzierung von Komplexitäten oder zur Erstellung von Berechenbarkeitsanalysen werden hohe Erwartungen gesteckt. Die Realität zeigt aber, dass Fragen hinsichtlich dahinterliegender Interessen, Machtverhältnisse, Ethik und Moral, Kontrolle, Rechte und Pflichten oftmals nur eine untergeordnete Rolle spielen.

Das Geschäftsmodell digitaler Ökosysteme beruht zu einem großen Teil darauf, Daten zu monetarisieren. Es spielt aber eigentlich keine Rolle, ob Google, Facebook oder andere Internet-Dienstleister, die Polizei, Finanz- oder Statistikämter, Krankenkassen, Versicherungen, Banken oder – wie jüngst kontrovers debattiert – Geheimdienste auf diese personenbezogenen Daten zugreifen. Für die Nutzung und den Zugriff auf nicht anonymisierte Daten gelten für alle grundsätzlich gesetzliche Vorgaben des Datenschutzrechts.

Hauptprobleme im Bereich Big Data

24

% der Befragten (n=82\*), 2012



\*KMU und Großunternehmen aus verschiedenen Branchen

Quelle: Fraunhofer IAIS

## 6. Big Data und der Datenschutz

„IT-Sicherheit wird zu einer wesentlichen Voraussetzung zur Wahrung der Freiheitsrechte. Die gesellschaftlichen Chancen und ökonomischen Potenziale der Digitalisierung dürfen nicht gefährdet werden. [...] Die Themen IT-Sicherheit und die Abwehr von Wirtschaftsspionage sollen darüber hinaus eine besondere Rolle spielen“, so steht es geschrieben im aktuellen Koalitionsvertrag von CDU, CSU und SPD „Deutschlands Zukunft gestalten“.<sup>35</sup>

Ein ambitioniertes Ziel der künftigen Bundesregierung. Seit Bekanntwerden der Abhörskandale im Juni 2013 sind allerdings noch keine konkreten Handlungen erfolgt. Grundsätzlich gibt es aber Datenschutzrichtlinien auf nationaler sowie europäischer Ebene. In Deutschland ist dies das *Grundrecht auf informationelle Selbstbestimmung* sowie das *Grundrecht auf Gewährleistung der Vertraulichkeit und Integrität informationstechnischer Systeme*. Diese sollen sicherstellen, dass persönliche Daten nicht außerhalb ihres ursprünglichen Kontextes und ohne die Einwilligung der jeweiligen Person gespeichert, weitergegeben oder analysiert werden.

Das Bundesverfassungsgericht hat im Zuge der Volkszählungsentscheidung 1983 das erste Grundrecht auf Datenschutz verabschiedet. Das sogenannte Recht auf informationelle Selbstbestimmung besagt, dass Individuen grundsätz-

<sup>34</sup> <http://de.wikipedia.org/wiki/HTTP-Cookie>.

<sup>35</sup> <https://www.cdu.de/sites/default/files/media/dokumente/koalitionsvertrag.pdf>, S. 139.

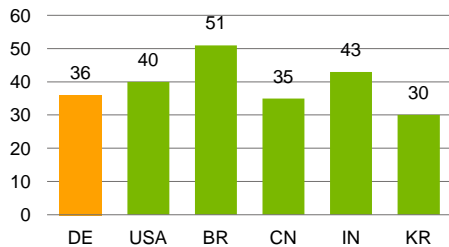


## Big Data – die ungezähmte Macht

"Ich versuche so wenig persönliche Informationen im Internet preiszugeben als möglich"

25

% der Befragten



DE: n=1.213, USA: n=1.218, BR: n=1.215, CN: n=1.207, IN: n=1.211, KR: n=1.214

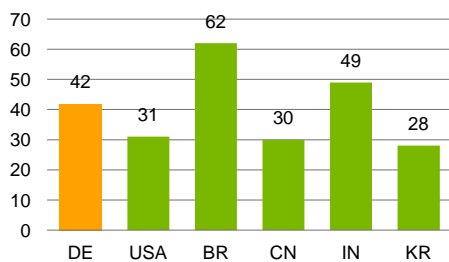
Quelle: Münchner Kreis 2013, www.zukunft-ikt.de

lich selbst die Befugnis darüber haben, wer was wann und bei welcher Gelegenheit über sie weiß, also wem welche Daten für welchen Zweck zur Verfügung gestellt werden. Die Verfügung von persönlichen Daten erfolgt i.d.R. durch informierte Einwilligungen betroffener Personen. Weiterhin gilt das Zweckbindungsprinzip, wonach personenbezogene Daten grundsätzlich nur für jene Zwecke verwendet werden dürfen, für die sie ursprünglich erhoben wurden. Einschränkungen des Grundrechts, insbesondere durch staatliche Stellen, sind nur im Allgemeininteresse zulässig und müssen dem Prinzip der Verhältnismäßigkeit entsprechen. Die Datensparsamkeit bei Informationssystemen sieht zudem vor, dass möglichst wenige Daten erhoben und verarbeitet werden sollen. „Außerdem fordert die Verfassungsrechtsprechung technisch-organisatorische und verfahrensrechtliche Vorkehrungen zum Schutz des Rechts auf informationelle Selbstbestimmung. Zentral für die Verwirklichung der Selbstbestimmung sind die Betroffenenrechte, also der Anspruch auf Auskunft über die in Bezug auf die eigene Person gespeicherten Daten sowie – bei Bedarf – auf Berechtigung, Sperrung und Löschung. Zur Umsetzung der Regelungen zur informationellen Selbstbestimmung bedarf es unabhängiger Kontrollinstanzen (BVerfGE 65, 1ff.).“<sup>36</sup>

"Unternehmen sollen transparent machen, wie und für was sie persönliche Daten nutzen"

26

% der Befragten



DE: n=1.213, USA: n=1.218, BR: n=1.215, CN: n=1.207, IN: n=1.211, KR: n=1.214

Quelle: Münchner Kreis 2013, www.zukunft-ikt.de

### Keine zeitgemäße Grundlage: Richtlinie aus dem Jahr 1995

Das Grundrecht auf Datenschutz wurde 2009 in die Europäische Grundrechtecharta unter Artikel 8 aufgenommen. Unter Verwendung des EU-Rechts ist das Datenschutzgrundrecht nicht nur von den Stellen der EU, sondern auch den Mitgliedstaaten zu beachten. Reformen sind aber unvermeidlich, da die Datenschutzrichtlinien nicht von allen Mitgliedsländern einheitlich umgesetzt und angewendet werden und die eher veralteten Datenschutz-Richtlinien (von 1995) auch nicht mit der Adaptiongeschwindigkeit web-basierter Technologien und generell mit der Evolution des Internets mithalten können. Im Januar 2012 hat die EU-Kommission einen Vorschlag für eine EU-Datenschutz-Grundverordnung vorgelegt. Sie soll die Datenschutzrichtlinie aus dem Jahre 1995 ersetzen und wird seither im Rat der Union und im Europäischen Parlament diskutiert.<sup>37</sup>

Der Änderungsvorschlag sieht vor, den Geltungsbereich der Grundverordnung künftig nicht mehr ausschließlich an den Umstand zu knüpfen, wo geographisch gesehen die verantwortliche Stelle niedergelassen ist und wo die Verarbeitung der Daten erfolgt, sondern auch danach zu fragen, ob hiervon personenbezogene Daten von Personen in der EU betroffen sind. Das sogenannte Marktortprinzip wird dafür sorgen, dass sich große datenspeichernde US-Unternehmen, wie digitale Ökosysteme, nicht mehr darauf berufen können, das europäische Recht binde sie nicht.<sup>38</sup>

### Kampf der Kulturen

Anscheinend haben die USA eine andere Auffassung von datenschutzrechtlichen Aspekten. Da sich aber Millionen europäischer und deutscher Bürger täglich auf sozialen Netzwerk-Plattformen und Onlinehandelsplattformen amerikanischer Internetfirmen tummeln, treffen hier natürlich zunehmend unterschiedliche Datenschutz-Kulturen aufeinander. Da ist Konfliktpotenzial programmiert. Angesichts der jüngst bekannt gewordenen Verstöße gegen das Datenschutzgesetz sind die strukturellen und regelungstechnischen Fehler und Defizite im deutschen, europäischen aber auch im internationalen Datenschutzrecht schnellstmöglich zu beseitigen. Nur so kann der Schutz der Privatsphäre gewährleistet und das Vertrauen der Bevölkerung in die digitale Welt erhöht werden. Ein höheres Vertrauen in digitale Infrastrukturen würde dem Internet als optimaler Katalysator für Innovation und Wachstum noch mehr Schub verleihen.

<sup>36</sup> Vgl.: Weichert, T. (2013). Big Data: Das neue Versprechen der Allwissenheit.

<sup>37</sup> Siehe hierzu ausführlich: Schaar, P. (2013). Big Brother und Big Data – Was heißt eigentlich Datenschutz auf Amerikanisch? Seite 5.

<sup>38</sup> Vgl.: Schaar, P. (2013). Seite 6.





## Big Data – die ungezähmte Macht

### Lösungsansätze sollten möglichst international gelten

#### Anonymisieren und aggregieren von personenbezogenen Datensätze

Einige Big-Data-Analysepraktiken verletzen grundsätzliche Konzepte des Datenschutzgesetzes, insbesondere jene Datenanalysen mit Personenbezug. Dieser Konflikt könnte entschärft werden, wenn eine Anonymisierung der Datensätze erreicht wird.<sup>39</sup> Diese muss die Möglichkeit eines Personenbezugs für die Dauer der Speicherung und Verarbeitung ausschließen. Allerdings bleibt die Gefahr der Re-Identifizierung grundsätzlich bestehen. Denn je mehr individuelle Merkmale in einen Datensatz fließen, desto größer ist das Risiko, dass mit entsprechendem Zusatzwissen Rückschlüsse auf Individuen möglich sind. Eine weitere Anonymisierungsmaßnahme stellt die Datenaggregation dar, also die Zusammenfassung individueller Datensätze zu Gruppendatensätzen. Daten könnten absichtlich „unscharf“ gemacht werden, damit bei einer Analyse keine exakten Zahlen, sondern eher Näherungswerte gewonnen werden können. Damit wird die Rückgewinnung individueller Erkenntnisse deutlich schwieriger.<sup>40</sup>

#### Forderung nach mehr Transparenz

Darüber hinaus können bei der Anwendung und Analyse von Datensätzen Maßnahmen zur Transparenz dem mangelnden Vertrauen der Internetnutzer entgegenwirken und Datenschutzverletzungen minimieren. Die Transparenz sollte sämtliche Analyse-Schritte umfassen, d.h. sowohl die Erhebung der Daten, deren Verschmelzung mit weiteren Datensätzen sowie die eigentliche Analyse und die anschließende(n) Ergebnisverwendung(en). Wenn die unterschiedlichen Akteure, entweder durch einen regulativen Rahmen oder durch selbst-aufgelegte Maßnahmen, ihre Analysepraktiken transparenter gestalten, erhalten die Menschen die Möglichkeit, informiert und selbstbestimmt über die Weitergabe ihrer Daten zu verfügen bzw. einer Analyse zuzustimmen. Dadurch wird dem „Black Box“-Charakter von Big Data entgegengewirkt.

#### Internationale Algorithmen-Abkommen bieten Standardisierung und Zertifizierung

Beim Handel zwischen Staaten gelten in der Regel international vereinbarte Handelsabkommen. Dadurch sollen Welthandel und Weltwirtschaft gefördert werden. Ähnlich wie diese Handelsabkommen den internationalen Warenverkehr regeln, könnte auch ein Bündnis von Ländern internationale Algorithmen-Abkommen verabschieden, die den Handel mit personenbezogenen Daten vereinheitlichen und durch z.B. externe „Algorithmiker“<sup>41</sup> zertifizieren lassen. Dadurch könnten Analysemethoden standardisiert und transparent gemacht werden. Dies würde nicht nur das Bürger-Staat-Verhältnis (in Sachen Geheimdienstpraktiken) einbeziehen, sondern beträfe auch die international tätigen digitalen Ökosysteme.

#### Die Basis ist „Vertrauen“

Die Änderung des Datenumfangs hat auch zu einer Änderung des Datenwessens geführt, was den Schutz der Privatsphäre vor eine große Herausforderung stellt. Die Ausmaße und Potenziale von Big Data sind zum jetzigen Zeitpunkt nur zu erahnen. Daher gestaltet es sich schwierig, alle Eventualitäten juristisch abzudecken. Das Potenzial kann sich dann optimal entfalten, wenn die agierenden Menschen Vertrauen in web-basierten Technologien haben, d.h. ihre Privatsphäre unangetastet bleibt und ihre Grundrechte respektiert und gewahrt werden.

### Das Vertrauen in digitale Kanäle ist bereits angekratzt

Seit im Juni 2013 die Dokumente von Ex-Geheimdienstmitarbeiter Edward Snowden veröffentlicht wurden, wird die Diskussion um Big Data kaum noch ohne die Überwachungspraktiken von Geheimdiensten geführt. Kritiker glauben,

<sup>39</sup> Vgl.: Weichert, T. (2013). Big Data: Das neue Versprechen der Allwissenheit.

<sup>40</sup> Vgl.: Weichert, T. (2013). Big Data – eine Herausforderung für den Datenschutz.

<sup>41</sup> Vgl.: Mayer-Schönberger, V. und Cukier, K. (2013).





## Big Data – die ungezähmte Macht

---

dass alles, was technologisch möglich ist, auch getan wird, um an persönliche Daten zu gelangen, unabhängig von etwaigen (nationalen) datenschutzrechtlichen Grenzen.

*„If you're looking for the needle in the haystack, you have to have the entire haystack first“* [Deputy Attorney General James Cole, stellvertretender US-Justizminister]

### Nationale Datenschutzbestimmungen können umgangen werden

Problematisch ist, dass hierzulande geltendes Datenschutzrecht durch den Einsatz von Technologien umgangen werden kann. Beispielsweise läuft ein Großteil des deutschen Telekommunikationsverkehrs über das Ausland (hauptsächlich die USA), weil die inländischen Serverkapazitäten offensichtlich mit der Datenfülle überfordert sind. Es gibt Berichte, dass Daten abgegriffen, gespeichert und weiterverwendet werden, ohne dass offiziell das deutsche Datenschutzrecht verletzt wird.<sup>42</sup> Inwiefern große Teile der digitalen Kommunikation auch bewusst über ausländische Server umgelenkt wurden und ob sich deutsche Telekommunikationsunternehmen den Begehrlichkeiten der Geheimdienste überhaupt entziehen können, bleibt wohl, wie viele Einzelheiten der gesamten Ausspähhaktion, ein wohlbehütetes Geheimnis.

### Abhörpraktiken gefährden womöglich die Innovationskraft

### Volkswirtschaftlicher Schaden sollte nicht unterschätzt werden

Die jüngst diskutierten Abhörpraktiken könnten jedoch weitreichende volkswirtschaftliche Konsequenzen mit sich bringen und sollten nicht unterschätzt werden. Die ökonomische Gefahr liegt vor allem darin, dass die Menschen, aufgrund sinkenden Vertrauens und zunehmender Unsicherheit ihr Mediennutzungs- und Konsumverhalten mittel- bis langfristig anpassen sowie ihre Adaptionsgeschwindigkeit web-basierter Technologien reduzieren könnten. Das mag der Entwicklung web-basierter Technologien eventuell einen Wachstumsdämpfer bescheren, insbesondere für die digitalen Ökosysteme, aber auch für viele Nischenanbieter und Start-Ups im Bereich IKT, die bereits jetzt verstärkt damit beschäftigt sind, ihren Kunden sichere IT-Infrastrukturen und Betriebssysteme zu versprechen und mögliche Sicherheitslücken zu schließen.

### Sichere IT-Architekturen gewinnen zunehmend an Bedeutung

Selbst wenn viele Menschen die bekannt gewordenen Abhörpraktiken der Geheimdienste nicht als direkte Gefahr einstufen, könnte sich aufgrund der dauerhaften medialen Berichterstattung unterschwellig eine gewisse Unsicherheit einschleichen. Wenn mehr Menschen sich im Internet beobachtet und abgehört bzw. kontrolliert fühlen, hat das Auswirkungen auf die individuelle Entfaltung, Freiheit, Kreativität und zuletzt sicherlich auch auf die Innovationskraft und die Wettbewerbsfähigkeit einer gesamten Volkswirtschaft. Eine 100%ige Datensicherheit ist und bleibt natürlich Illusion. Dennoch wird die Sicherheit von IT-Infrastrukturen für die Nutzer künftig an Bedeutung gewinnen. Unternehmen, die glaubwürdig sichere Internet-Dienste und -Technologien anbieten, dürften davon profitieren.

## 7. (Big) Data in der Praxis: Es gibt keinen Königsweg

Die durch Datenexplosion und -auswertung entstehenden neuen Nutzungs- und vor allem Kommerzialisierungsmöglichkeiten stellen die Akteure folglich vor die Frage, ob es sich für sie lohnt, die prinzipiell zur Verfügung stehenden Daten zu filtern, zu analysieren bzw. generell in Big-Data-Projekte zu investieren. Diese Frage kann sicherlich bejaht werden. Denn ähnlich, wie es sich die Anbieter von Produkten und Dienstleistungen heutzutage nicht mehr leisten können, Diskussionen und Bewertungen über ihre Unternehmensleistungen auf sozialen Netzwerk-Plattformen und/oder Foren durch ihre (potenziellen) Kunden zu igno-

---

<sup>42</sup> <http://www.heise.de/security/meldung/Raetselhafte-Entfuehrungen-im-Internet-2053503.html>.

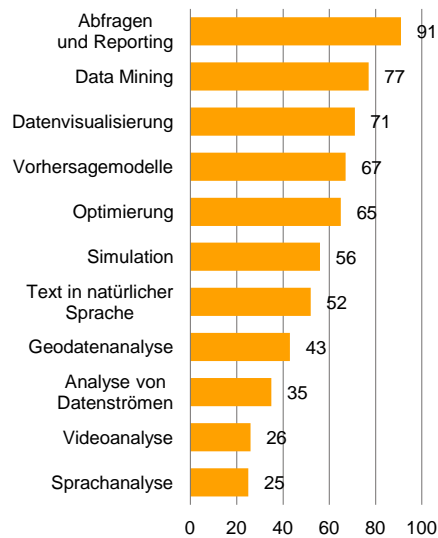


## Big Data – die ungezähmte Macht

### Big Data Analyse-Tools

27

% der Befragten (n pro Nennung=508-870 aus 1.144), 2012



Quellen: IBM Institute for Business Value, Saïd Business School Oxford

rieren<sup>43</sup>, so ist es ratsam, vorhandene interne und/oder externe Datensätze auszuwerten. Anderenfalls könnten den Unternehmen mangels relevanter Informationsvorsprünge mittelfristig wertvolle Wettbewerbsvorteile verlorengehen.

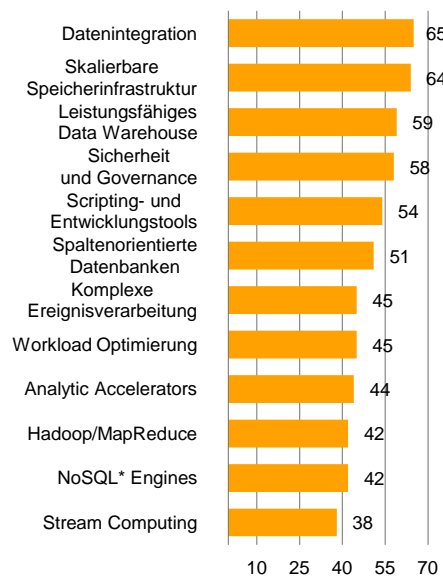
Die Akteure sind daher gut beraten, wenn sie die Fülle an Informationen und Datenströmen als potenzielles Wachstumsfeld anerkennen und schnell lernen, sie effektiv zu kanalisieren. Hierfür brauchen die Akteure eine anpassungsfähige Digitalisierungsstrategie. Beispielsweise werden neue IT-Systeme künftig so in bestehende IT-Architekturen eingebettet, dass sie Trends und Vorhersagen schneller aufspüren können, um z.B. Analysen und Prognosen bezüglich der Verhaltensmuster der Kunden, Geschäftspartner und/oder Wettbewerber durchführen zu können. Die Möglichkeit, Prozesse und Strukturen effizienter zu gestalten, IT-Systeme zu optimieren, eventuelle Synergien zu erreichen und somit Kosten zu senken, sind ebenfalls nicht zu vernachlässigende Vorteile.

Um wettbewerbsfähig zu bleiben, sollten die Akteure lernen, ihren steigenden Datenpool mit modernen Managementmethoden zu verbinden und in adäquate Geschäftsstrategien sowie entsprechend angepasste Produkte und Dienstleistungen zu transformieren. Mag auch einigen Akteure anfangs der große Wurf nicht gelingen, so schlummert doch in vielen Geschäftsprozessen branchenübergreifend Effizienzpotenzial, das auch in kleinen Schritten gehoben werden kann. Für Finanzinstitute könnte das z.B. bedeuten, dass sich ihre Aufmerksamkeit weniger auf Finanzprodukte (z.B. Derivate) und andere neue, virtuelle Finanzprodukte richten, sondern vielmehr auf die Bankdienste, die rund um den Finanzbereich entstehen (web-basierte Beratung, Informationsdienste, Foren etc.). Im Angebot der Bank werden künftig sicherlich mehr Dienste zu finden sein, die auf gefilterten Kundeninformationen basieren.

### Welche Plattformkomponenten werden verwendet?

28

% der Befragten (n pro Nennung=297-351 aus 1.144), 2012



\*NoSQL = Nicht-relationale Datenbanken, geeignet für datenintensive Applikationen, wie z.B. die Indexierung großer Dokumentenmengen

Quellen: IBM Institute for Business Value, Saïd Business School Oxford

### Aufholpotenzial für den deutschen Mittelstand

Gemäß der Innovationspotenzialanalyse des Fraunhofer Instituts für Intelligente Analyse- und Informationssysteme (IAIS) gaben die Teilnehmer der Umfrage an, dass sie branchenübergreifend und langfristig vor allem auch Potenzial für die Individualisierung von Dienstleistungen (z.B. tägliche Gesundheitsdiagnosen, individuelles Entertainmentprogramm on demand, etc.) sehen.<sup>44</sup> Zudem wünschen sie sich vermehrt intelligente Produkte (z.B. selbstlernende und selbstregulierende Häuser, autonom fahrende Fahrzeuge etc.). Gleichzeitig wurde aber auch von 95% der Umfrageteilnehmer der Wunsch nach Förderung in Form von Best Practices oder Schulungen geäußert, um bestehende Wissensdefizite im Bereich Big Data zu reduzieren.

Der vielversprechende Datenschatz, der aus unterschiedlichen Quellen stammt, wird von vielen (mittelständischen) Unternehmen (noch) weitgehend ignoriert. Das wird sich in Zukunft sicherlich ändern. So stehen auch deutsche mittelständische Unternehmen vor der enormen Herausforderung, die anfallenden Datenmengen zu erfassen, zu speichern und gewinn-, bzw. effizienzbringend zu verarbeiten. Im Gegensatz zu den großen Internetfirmen fehlen aber vielen mittelständischen Unternehmen die notwendigen digitalen Strategien bzw. das Know-how, ihre internen sowie externen Wertschöpfungsprozesse mit modernen Datenanalysetechnologien auszustatten. Wer nicht zeitnah auf diese Entwicklung mit adäquaten und vor allem flexiblen Softwarelösungen und Filterinstrumenten reagiert, verpasst unter Umständen wertvolle Wettbewerbsvorteile und büßt womöglich sogar Marktanteile ein. Allerdings handelt es sich bei dem Implementierungsprozess moderner Big-Data-Instrumente um keinen Selbstläufer. Ein zusätzlicher Ressourceneinsatz ist erforderlich. Unabhängig von den

<sup>43</sup> Vgl.: Dapp, T. (2011). Die digitale Gesellschaft. Seite 7ff.

<sup>44</sup> Vgl.: Fraunhofer IAIS (2012). Big Data – Vorsprung durch Wissen.



## Big Data – die ungezähmte Macht

| Die 4 Einführungsphasen bei Big Data |             |  | 29   |
|--------------------------------------|-------------|--|------|
| 1.                                   | Informieren | Konzentration auf Recherche und Marktbeobachtung   | 24%* |
| 2.                                   | Planen      | Entwicklung einer Strategie und Roadmap auf Basis geschäftlicher Anforderungen und Herausforderungen | 47%  |
| 3.                                   | Prüfen      | Start von Big Data-Initiativen zur Überprüfung von geschäftlichem Nutzen und Anforderungen           | 22%  |
| 4.                                   | Umsetzen    | Implementierung von zwei oder mehr Big Data-Lösungen und Einsatz von Advanced Analytics              | 6%   |

\* Prozent aller Befragten, 2012 (n=1.061, Prozentsatz entspricht aufgrund Rundung nicht 100%).

Quellen: IBM Institute for Business Value, Saïd Business School Oxford

unterschiedlichen Akteuren sollten die Entscheidungsträger zu Beginn des Big-Data-Prozesses grob folgende Fragen beantworten können:

- Welche Ziele werden verfolgt? Beispielsweise Effizienz- und/oder Umsatzsteigerung, Optimierung von IT-Netzwerk- und Systemarchitektur, Öffnungsprozesse vorantreiben (z.B. externe Wissensträger implementieren, Daten aufbereiten und veröffentlichen), Kundenprofile erstellen, Szenarien- oder Modellberechnungen optimieren
- Welche Daten sind für die Analyse relevant?
- Welche Daten sind bereits vorhanden; welche Daten können intern/extern zusätzlich aggregiert werden?
- Welche Analysetechnologien und -methoden werden benötigt?
- Welche zusätzlichen Ressourcen werden benötigt und wer verfügt über die Entscheidungshoheit? Know-how (interne/externe Fach- und Managementkompetenzen), interne/externe Hardware/Software (Analysetools, Data Warehousing, Speicherkapazität), Zeit, finanzielle Mittel (interne/externe Finanzierung) etc.

Um das Potenzial sowie die breiten Einsatzmöglichkeiten von Big Data aufzuzeigen, werden im Folgenden kurz einige durchaus erfolgreichen Big-Data-Projekte aus Wirtschaft, Wissenschaft und dem öffentlichen Sektor vorgestellt:

### Big-Data-Projekt im Gesundheitssektor

Bislang mussten sich Gesundheitsbehörden bei der Analyse von Grippewellen auf die Anzahl der meldepflichtigen Krankheitsfälle stützen. Da aber viele Menschen tendenziell nicht gleich bei ersten Krankheitssymptomen zum Arzt gehen und die Dauer des Übermittlungsprozesses der amtlichen Meldung von Krankheitsfällen relativ hoch ist, können die Behörden i.d.R. nur mit einer (gerade für den Gesundheitsbereich) erheblichen zeitlichen Verzögerung von 1 bis 2 Wochen<sup>45</sup> auf den Ausbruch von Krankheitswellen reagieren.<sup>46</sup>

Ein Big-Data-Projekt von Google kann hier Abhilfe leisten: Seit Ende 2008 gibt es inzwischen für mehr als 25 Länder das für Nutzer kostenlos zur Verfügung stehende Onlinetool „Google Grippe-Trends“.<sup>47</sup> Ein auf Daten basierendes Frühwarnsystem von Grippewellen kann für die unterschiedlichen Gesundheitsbehörden als Ergänzung zu herkömmlichen Grippe-Überwachungsmechanismen verwendet werden, um zeitnah, vielleicht sogar präventiv, aktiv zu werden.<sup>48</sup> Die Google-Forscher Ginsberg und Mohebbi<sup>1</sup> entwickelten einen Algorithmus, der die täglich am häufigsten vorkommenden 50 Millionen Suchanfragen nach griperelevanten Begriffen filtert. Suchvolumenanalysen (ebenfalls mit Tools aus der Google-Welt) zeigen, dass sich die Suchanfragen zu Beginn einer Grippeperiode überdurchschnittlich häufen. Die Datenfilterung lässt ein Muster erkennen, das insbesondere saisonale Volatilitäten aufweist. Der nachträgliche Datenvergleich von tatsächlichen Krankheitszahlen, bereitgestellt durch die jeweiligen Gesundheitsbehörden, mit den auf elektronischen Suchanfragezahlen beruhenden Vorhersagen zeigt ein nahezu übereinstimmendes Bild. Da die regionalen Unterschiede im Bereich Grippeausbreitung relativ groß sind, benutzt Google Informationen über Krankheitsfälle von regionalen Anbietern.

### Big Data verschafft Informationsvorsprünge

<sup>45</sup> Ginsberg, Jeremy et al. (2009). Detecting influenza epidemics using search engine query data.

<sup>46</sup> Vgl.: Mayer-Schönberger, V. und Cukier, K. (2013). Big Data. Die Revolution, die unser Leben verändern wird. S. 7ff.

<sup>47</sup> <http://www.google.org/flutrends/intl/de/about/how.html>.

<sup>48</sup> Dieses Analysemodell beschränkt sich bislang auf die Analyse von Grippewellen und das Denguefieber – die Vorhersage anderer Krankheitswellen ist in naher Zukunft nicht geplant.



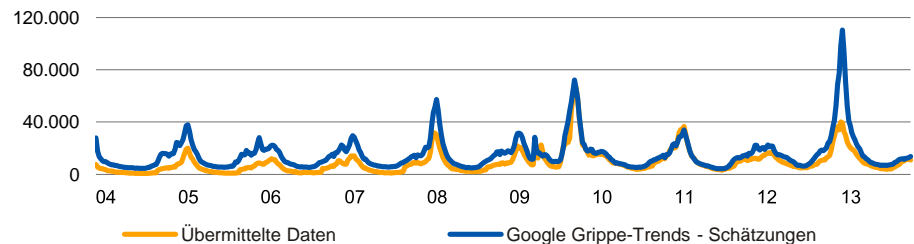
## Big Data – die ungezähmte Macht

Insgesamt ergibt sich ein durchaus realitätsnahes Bild ohne zeitliche Verzögerung – im Vergleich zu den zeitlich hinterherhinkenden herkömmlichen Grippeüberwachungssystemen. Auf dieser Basis lassen sich also exakte Schätzungen zur aktuellen Grippelage in verschiedenen Regionen anstellen. Dieser Informationsvorsprung verschafft den Gesundheitsbehörden wertvolle Zeit, um Grippeepidemien schnell zu erkennen und eventuell einzudämmen.

Grippe-Häufigkeit Vereinigte Staaten

30

Historische Schätzungen



Quellen: U.S. Centers for Disease Control; Google Grippe-Trends (<http://www.google.org/flutrends>)

Das Google-Analysemodell wird jährlich überprüft und ggf. angepasst, sollte sich das elektronische Suchverhalten nach Gesundheitsinformationen ändern, z.B. hinsichtlich der Terminologie.<sup>49</sup> Die Google-Vorhersagen für Anfang 2013 überschätzten die tatsächliche Erkrankungsrate (s. Grafik). Google kündigte an, den Algorithmus an die neuen Suchgewohnheiten anzupassen. Forscher vermuteten, dass die Warnungen der Medien vor einer drohenden Grippewelle zu Beginn des Jahres 2013 die Menschen dazu veranlassten, rein präventiv nach gripperelevanten Begriffen zu „googeln“, ohne jegliche Symptome zu haben. Dies verzerrte die Vorhersage des Google-Tools – zumindest für die USA.

### Misstrauen schwingt dennoch mit

Durch unmittelbare Grippevorhersagen, kann der Gesundheitssektor effektiver gesteuert werden. Somit bietet das Analysetool von Google ein Indiz für die Innovationskraft, die hinter Big Data stecken kann, sowie einen möglichen volkswirtschaftlichen Nutzen. Jedoch darf nicht vergessen werden, dass für die Vorhersagen persönliche Daten gespeichert und ausgewertet werden müssen. Google versichert, dass der einzelne Nutzer nicht identifiziert werden könne, da die Auswertung der Suchanfragen anonymisiert erfolge. Lediglich das große Ganze sei von Interesse, da nur so verlässliche Aussagen möglich seien. Doch genügt in diesem Fall die Vorgehensweise von Google hinsichtlich datenschutzrechtlicher Bestimmungen? Zwar setzt sich die Analyse milliardenfach verschiedener Suchanfragen individueller Nutzer zusammen. Um regionale Vorhersagen treffen zu können, erfolgt die geographische Zuordnung der Suchanfragen jedoch über die jeweiligen IP-Adressen.<sup>50</sup> Daten basierend auf IP-Adressen lassen sich jedoch ohne größeren Aufwand re-anonymisieren. Letztendlich schwingt immer etwas Misstrauen mit. Misstrauen dahingehend, dass solche Tools auch eine elegante Rechtfertigung dafür sind, persönliche Daten dauerhaft zu speichern und gegebenenfalls zu monetarisieren.

<sup>49</sup> Beispielsweise war gleich im Jahr 2009 eine Anpassung des Ausgangsmodells notwendig, um auch die nicht saisonal verlaufene Schweinegrippe-Pandemie bestmöglich analysieren zu können. Siehe: Cook, Samantha et al. (2011). Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic.

<sup>50</sup> Internetprotokolladresse: jedem Gerät in einem Netzwerk wird eine Adresse in Form von Zahlenfolgen zugeordnet; die eine genaue Adressierung und Zustellung von Datenpaketen ermöglicht.



## Big Data – die ungezähmte Macht

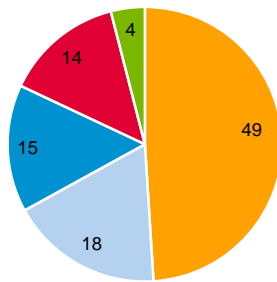
### Big-Data-Projekt im Marketingbereich

„What scares me about this is that you know more about my customers after three months than I know after 30 years.“ [Lord MacLaurin, ehem. CEO von Tesco, Reaktion auf die ersten Ergebnisse aus Test von Treuekarten]

#### Big Data-Ziel: Kundenorientierung

31

% der Befragten, gewichtet und zusammengefasst, n=1.067, 2012



- Kundenorientierte Ergebnisse
- Betriebliche Optimierung
- Risiko-/Finanzmanagement
- Neues Geschäftsmodell
- Mitarbeiterzusammenarbeit

Quellen: IBM Institute for Business Value, Saïd Business School Oxford

Vor allem bei großen (amerikanischen) Einzelhandels-Ketten ist die Analyse-Methode, Daten für Marketingzwecke einzusetzen, inzwischen gängige Praxis. Die noch bis vor wenigen Jahren anonyme Beziehung zwischen großen Einzelhändlern und dem einzelnen Kunden soll durch ein „One-to-one-Marketing“, nach dem Prinzip des Tante-Emma-Ladens in größerer Dimension, in eine individuelle Händler-Kunden-Beziehung umgewandelt werden.<sup>51</sup> Tante Emma wusste aufgrund ihrer detaillierten Kundenkenntnis genau, welchem Kunden sie welche Produkte zu welchem Zeitpunkt empfehlen und schlussendlich verkaufen konnte. Diese eher analoge „Datenbank“ möchten die Supermarktketten in großem Maßstab einsetzen. Hierfür werden kontinuierlich kundenspezifische Daten gespeichert, aktualisiert und analysiert. Die Daten hierfür stammen hauptsächlich aus Kreditkartenkäufen und Einkäufen im jeweiligen Onlineshop – aber auch Daten von Kundenkarten (Loyalty-Programme, Kunden- oder Bonuskarten) geben Aufschluss über Konsumvorlieben.

Die herkömmliche Marktforschung mit ihren personenbezogenen Umfragen wird immer mehr durch eine datenbasierte Analyse des Kaufverhaltens in Echtzeit ersetzt. Um herauszufinden, welche Kundengruppe welches Produkt kauft, braucht es somit gute Programmierer und Datenanalysten sowie lernfähige Algorithmen. Mit den richtigen Rückschlüssen kann dem Kunden zum gewünschten Zeitpunkt das gewünschte Produkt angeboten werden, in Form von Gutscheinen, individuellen Werbeprospekten, Werbe-E-Mails etc. Big-Data-Anwendungen haben also aus Sicht des Handels vor allem das Potenzial, zum lukrativen Marketinginstrument zu werden.

Die US-Amerikanische Kaufhauskette Target untersuchte z.B. das Kaufverhalten von Kundinnen, die sich für den Baby-Geschenke-Service registrieren ließen.<sup>52</sup> Es zeigte sich folgendes Kaufmuster: Schwangere kaufen ab dem dritten Schwangerschaftsmonat verstärkt parfümfreie Lotionen, einige Wochen später spezielle Nahrungsergänzungsmittel. Die sich im Verlauf der Schwangerschaft ändernden Konsumgewohnheiten weisen ein so einheitliches Muster auf, dass sie der Supermarktkette sogar relativ genaue Rückschlüsse auf den jeweiligen Geburtstermin erlauben. Target legt, basierend auf diesen Kenntnissen, Produktlisten an, die als Frühwarnsignal für Schwangerschaften bei Kundinnen dienen. Zeigt das Kaufverhalten einer Kundin plötzlich dieses „Schwangerschaftsmuster“, so ist es bei Supermarktketten mittlerweile üblich, die Kundin dementsprechend durch gezielte Werbung oder Gutscheine anzusprechen. Die individuelle Kundenansprache ist mittlerweile so ausgefeilt und sicher in ihrer Vorhersage, dass sie sogar familiäre Krisen auszulösen vermag. So geschehen in den USA: ein Vater warf der Einkaufskette vor, seine junge Tochter mit gezielter Werbung für Babyprodukte zu einer Schwangerschaft anzustiften. Wie er kurz darauf jedoch feststellen musste, hatte Target jedoch einen Informationsvorsprung. Die Analyse der Kaufgewohnheiten der Tochter entsprach dem Muster, das viele Frauen in den ersten Schwangerschaftsmonaten aufweisen. Die Tochter war tatsächlich schwanger.

Die weltgrößte Einzelhandelskette Walmart analysierte in ihrer bisherigen Big-Data-Geschichte diverse Datensätze zum Kaufverhalten ihrer Kunden und entdeckte dabei erstaunliche und, wie sich herausstellte, umsatzsteigernde Korrelationen: Der amerikanische Snack „Pop-Tart“<sup>53</sup> wird z.B. besonders oft gekauft,

<sup>51</sup> Bloching, B., et al. (2012). Data Unser. Wie Kundendaten die Wirtschaft revolutionieren.

<sup>52</sup> Vgl.: Mayer-Schönberger, V. und Cukier, K. (2013).

<sup>53</sup> <http://de.wikipedia.org/wiki/Pop-Tart>.

### Individuelle Kundenansprache





wenn Unwetterzentralen Hurrikan-Warnungen heraus geben.<sup>54</sup> Auch stellte Walmart fest, dass in den Abendstunden Bier und Pampers vergleichsweise oft im gleichen Einkaufskorb landeten.<sup>55</sup> Walmart setzt seitdem Big-Data-Marketinginstrumente zur Regaloptimierung ein. Windeln und Bier werden bei Walmart-Filialen in benachbarten Regalen aufgebaut und wenn ein Wirbelsturm naht, werden Paletten mit „Pop-Tarts“ direkt neben den Kassen positioniert.

### Big-Data-Projekt in der Wissenschaft

*„Torture the data long enough, and it will confess to anything.“* [Ronald Coase, Economist]

Das „FuturICT Knowledge Accelerator and Crisis-Relief System“<sup>56</sup> ist ein europäisches Forschungsvorhaben im Bereich der IKT. Dieses wissenschaftliche Großprojekt wird von zahlreichen internationalen Forschungsinstitutionen unterstützt, initiiert durch Dirk Helbing, der Professor an der ETH Zürich ist.

Die Philosophie von Open Innovation schwingt mit

Die Vision ist die Entwicklung einer auf einem „Supercomputer“ basierenden Wissensplattform. Hier sollen alle globalen Datenströme zu einem möglichst genauen Abbild der Realität zusammengeführt werden. Wissenslücken technischer, wirtschaftlicher, gesellschaftlicher und ökologischer Art werden durch Computersimulationen und Daten-Analysen minimiert bzw. geschlossen. Insbesondere sollen Wechselwirkungen zwischen den verschiedensten Bereichen aufgedeckt werden. Virtuell simulierte Zukunftsszenarien helfen diese Wechselbeziehungen zu managen und zeigen präventive Maßnahmen auf, die ergriffen werden könnten, um drohende Krisen im Weltgeschehen, wie Finanzkrisen, abzumildern oder sogar zu verhindern. Damit dies möglich wird, soll FuturICT auf drei Pfeilern aufgebaut sein: einem „planetaren Nervensystem“ (Planetary Nervous System), einem „lebendigen Weltsimulator“ (Living Earth Simulator) und einer „globalen Beteiligungsplattform“ (Global Participatory Platform).

Das Planetary Nervous System soll die Visualisierung des aktuellen Weltgeschehens hauptsächlich basierend auf den Daten eines weltumspannenden Netzwerks von Sensoren zum Ziel haben. Diese Darstellung der globalen Gesellschaft basiert auf Reality Mining, also der Echtzeit-Gewinnung von Datenströmen, produziert in allen Lebens- und Arbeitsbereichen. Besonders herausfordernd dabei ist die ständige Anpassung an neue Umstände, denn wie einst der griechische Philosoph Heraklit treffend bemerkte: „Nichts ist beständiger als der Wandel“, d.h. die Menschen und ihre Umgebung befinden sich in einem permanenten Transformationsprozess.

Mit Hilfe des zweiten Pfeilers, des Living Earth Simulators, sollen – basierend auf den im Planetary Nervous System und der Partizipationsplattform gespeicherten Daten – mögliche Zukunftsszenarien simuliert werden. Dieser politische „Windkanal“<sup>57</sup> soll Aufschluss darüber geben, wie sich das globale System verhält, wenn sich unterschiedliche Bedingungen ändern. Hier werden also „Was wäre, wenn...“-Szenarien durchgespielt. Den jeweiligen Szenarien werden Eintrittswahrscheinlichkeiten beigemessen, auf deren Basis sich ein Frühwarnsystem entwickeln lässt. Von besonderem Interesse sind natürlich jene Einflussfaktoren, die das globale System destabilisieren (können). Die Simulation beruht auf dem Experimentieren mit verschiedenen Modellen und deren Kombination (= pluralistische Modellierung), eingebettet in eine Open-Source-Softwareumgebung. Diverse Forscher sollen somit die Möglichkeit erhalten, neu entwickelte Analysetools direkt einspeisen zu können.

<sup>54</sup> [http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html?\\_r=0](http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html?_r=0).

<sup>55</sup> <http://www.forbes.com/forbes/1998/0406/6107128a.html>.

<sup>56</sup> [www.futurict.eu](http://www.futurict.eu).

<sup>57</sup> Geiselberger, H., Moorstedt T. (Hrsg.) (2013). Big Data: Das neue Versprechen der Allwissenheit. Suhrkamp Verlag, Berlin.



## Big Data – die ungezähmte Macht

---

Die Idee der interaktiven Infrastruktur setzt sich auch bei der offenen Partizipationsplattform fort: Sie basiert auf der Idee eines „Prosumers“, also dem koproduzierenden Konsumenten. „Prosumer“ können die auf der Plattform bereitgestellten Datenreihen, Simulationen und Applikationen einsehen und zugleich selbstbestimmt Informationen in das System einspeisen. Jeder, ob Individuum, Unternehmen oder Organisation, behält in diesem komplexen System also die Kontrolle über die eigenen Daten und hat zugleich die Möglichkeit, die eingepflegten Ideen anderer einzusehen. Mit dieser interaktiven Partizipation erfährt das FuturICT-Modell eine Art „Open Innovation“-Komponente, die insbesondere externe Wissensträger involviert und somit für eine laufende Qualitätsüberprüfung sorgt. Kernkonzepte von FuturICT sollen also Offenheit, Transparenz und Partizipationsmöglichkeit sein.

### Finanzierung als Hürde

Wann und in welchem Umfang FuturICT allerdings umgesetzt wird, ist nach der ausgebliebenen Förderung von 1 Mrd. EUR durch das 7. EU-Forschungsrahmenprogramm „Future Emerging Technologies“ (FET)<sup>58</sup> ungewiss. Trotz des offensichtlichen Nutzens für die Gesellschaft ist die Durchführbarkeit von FuturICT auch kritisch zu sehen. Ein globales, komplexes System, auf das eine Vielzahl von Nutzern permanent und zeitgleich zugreifen kann, erfordert immense Serverkapazitäten. Angesichts der aktuellen technischen Rahmenbedingungen ist jedoch fraglich, ob ein Projekt dieser Größenklasse trotz des Fortschritts bei der Serverkapazität verwirklicht bzw. finanzierbar ist.

### Kritik der Freiburger Schule aktueller denn je

Eine andere Hürde ist sicherlich auch eine fehlende fundierte Theorie über das menschliche Verhalten, auf der ein Vorhersagemodell basieren könnte. Kritiker bemängeln zudem, dass die Welt komplexer und auch chaotischer ist als jedes Modell oder alle Kombinationen von Modellen. Das Nichtwissen wird immer größer sein als das Wissen, und der Drang, mehr Kohärenz zu erkennen, als gegenwärtig erreichbar ist, führt schnell zu einer vorschnellen Generalisierung und Vereinfachung von komplexen Zusammenhängen. Schon Friedrich August von Hayek schrieb in seinem Buch über den „Weg zur Knechtschaft“ von dem Vorwurf, dass es sich die gesellschaftlichen Planer, mechanistisch auch als Sozialingenieure bezeichnet, gefallen lassen müssen, dem konstruktivistischen Irrglauben aufzusitzen, dass eine Gesellschaft am Reißbrett entworfen werden könne.<sup>59</sup> Es bleibt zudem fraglich, ob die Handlungsempfehlungen, die sich im Living Earth Simulator ergeben würden, für die Menschen nachvollziehbar wären. Unterschiedlich zu interpretierende Vorschläge, die nicht plausibel sind, werden politisch wohl kaum umgesetzt werden können.

### Open-(Big-)Data-Projekte in der öffentlichen Verwaltung

Viele Big-Data-Projekte werden ausschließlich durch partikularinteressierte Monetarisierungsstrategien getrieben. Die unterschiedlichen Anwendungen von Daten können aber auch, unabhängig von ihrem Datenvolumen, einem Zweck dienen, der einen wertvollen und ökonomischen Nutzen für die Gesellschaft stiftet. „Open Data“<sup>60</sup> oder „Open Government Data“ dienen im Gegensatz zu den üblicherweise diskutierten Big-Data-Anwendungen nicht primär monetären Zwecken, bieten aber ebenfalls einen oft unterschätzten reichen Datenschatz, der von unterschiedlichen Akteuren zu unterschiedlichen Zwecken mehrfach verwendet werden kann. Open Data steht für die Öffnung digitaler (nicht personenbezogener) Informationen, die der Bevölkerung (nahezu) uneingeschränkt zur Verfügung stehen.

---

<sup>58</sup> [http://cordis.europa.eu/fp7/ict/programme/fet\\_en.html](http://cordis.europa.eu/fp7/ict/programme/fet_en.html).

<sup>59</sup> Hayek, F. A. von (2004). Der Weg zur Knechtschaft. 4. Auflage. Mohr Siebeck Verlag, Tübingen.

<sup>60</sup> <http://opendatahandbook.org/de/why-open-data/index.html>.



## Big Data – die ungezähmte Macht

### Innovation durch Daten

Primär handelt es sich um Datensätze (z.B. Karten und Satellitenaufnahmen, Geo- oder Umweltdaten) und weniger um Textmaterial wie Dokumente und Akten(notizen). Getrieben wird diese Bewegung durch die steigende Nachfrage der Bürger nach Transparenz, Interaktion und Kollaboration. Ein weiterer ökonomisch wertvoller Aspekt ist die Innovation durch Daten. Denn die „offenen“ Datensätze dürfen von Bürgern, Konsumenten, Journalisten, Forschern, Unternehmen, Organisationen etc. für unterschiedliche, sowohl private als auch kommerzielle Zwecke verwendet und vor allem mehrfach weiterverarbeitet werden. Die dabei frei zur Verfügung gestellten Daten schaffen in Kombination mit web-basierten Technologien somit einen volkswirtschaftlichen Nutzen in Form neuer Geschäftsmodelle, innovativer Produkte und Dienstleistungen. So reichen beispielsweise Journalisten ihre investigativen Berichte durch unterschiedliche offene Datensätze an und treiben dadurch das Themenfeld „Datenjournalismus“ voran.<sup>61</sup> Open Data lebt davon, dass die Daten auch nachgenutzt bzw. wiederverwendet werden können. Hierfür organisiert z.B. die Open Knowledge Foundation<sup>62</sup> (OKF) in Deutschland regelmäßige sogenannte „Hackathons“.<sup>63</sup> Hier tauscht sich eine bunte Mischung an Interessierten, wie Journalisten, Open-Data-Experten oder Softwareentwickler aus, um vorhandene, frei zugängliche Daten zu analysieren und daraus eventuelle neue web-basierte Dienste in Form von Apps zu programmieren.

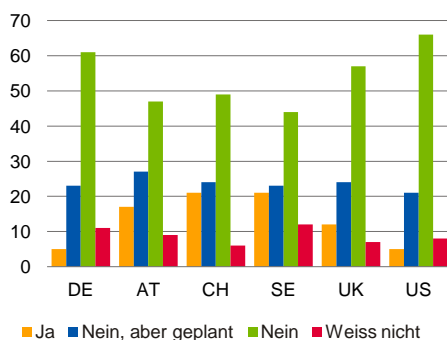
Obwohl Open Data sämtliche erhobenen Daten einschließt, wird der Begriff oft synonym für Open Government Data verwendet. Die allmähliche Öffnung von Verwaltungsdaten wandelt zunehmend das Bürger-Staat-Verhältnis. Denn die kostenlos zur Verfügung gestellten Daten schaffen auch neue demokratische Dimensionen; z.B. werden bislang undurchsichtige Behördenabläufe transparenter. Vor allem die angelsächsischen Staaten ([data.gov](http://data.gov); [data.gov.uk](http://data.gov.uk)), aber auch einige nordische Länder haben eine Vorreiterrolle bei der Öffnung von Verwaltungsdaten. Hier wurden bereits mehrere Open-Data-Projekte erfolgreich implementiert und auch von den Bürgern angenommen.<sup>64</sup>

In den vergangenen Jahren wurde eine Liste an Standards entwickelt, heute bekannt als die *zehn Prinzipien zum Öffnen von Regierungsinformationen*. Diese sollen Regierungen und Verwaltungen dabei helfen, ihre Datenbestände für die Allgemeinheit zu öffnen. Erst Mitte 2013 verpflichteten sich die G8-Mitgliedsstaaten<sup>65</sup> durch die Unterzeichnung einer „Open Data Charter“ dazu, fünf politische Prinzipien<sup>66</sup> zur Öffnung ihrer Datenbestände bis 2015 zu implementieren. Ein Beispiel für Open Data ist die britische Online-Steuerauskunftsdatei<sup>67</sup>. Hier können die Bürger einer übersichtlichen Aufschlüsselung entnehmen, wofür der britische Staat ihre Steuergelder ausgibt. Diese Option gibt es auch in Deutschland: die Homepage [offenerhaushalt.de](http://offenerhaushalt.de), ein weiteres Projekt der OKF, gibt detaillierte Informationen über die Haushaltsausgaben auf Bundesebene, mit Angaben dazu, welchem Ressort welches Budget zur Verfügung steht. Spätestens seit 2005, als das Informationsfreiheitsgesetz<sup>68</sup> erlassen wurde, hat jeder Bürger auch das Recht, sich über solche Sachverhalte informieren zu dürfen. Die Datendarstellung auf der Homepage ermöglicht es, sich schnell über die Ausgaben zu informieren, die Informationen auszuwerten und weiterzuverarbeiten.

Nutzung eGovernment-Dienste via mobile Endgeräte

32

Umfrage in %, 2013



DE (n=600); AT (n=713); CH (n=702); SE (n=725); UK (n=713); USA (n=576)

Quelle: eGovernment MONITOR 2013

### Etablierung von Standards im Bereich Open Data

<sup>61</sup> <http://datajournalismhandbook.org/>.

<sup>62</sup> Die OKF hat z.B. ein Minimum-Set von zehn Datensätzen definiert, die nationale Regierungen als Open Data veröffentlichen sollten und misst entsprechend Länder in ihrem „Open Data Census“. Zuletzt wurde der Zensus, der inzwischen fast 80 Länder umfasst, für den G8-Gipfel erhoben. Siehe <http://census.okfn.org/>.

<sup>63</sup> Siehe z.B. <http://energyhack.de/> oder <http://jugendhackt.de/>.

<sup>64</sup> Siehe hierzu z.B. das Open Data Institute (<http://theido.org>).

<sup>65</sup> Deutschland, Frankreich, Großbritannien, Italien, Japan, USA, Kanada und Russland.

<sup>66</sup> Open Data by Default, Qualität und Quantität, Nutzbar von allen, Veröffentlichung von Daten für besseres Regierungshandeln, Veröffentlichung von Daten für Innovation.

<sup>67</sup> <http://wheredoesmymoneygo.org/>.

<sup>68</sup> Das Gesetz gewährt jeder Person einen voraussetzungslosen Rechtsanspruch auf Zugang zu amtlichen Informationen von Bundesbehörden. Dabei ist eine Begründung durch rechtliche, wirtschaftliche oder sonstige Interessen nicht erforderlich.



## Big Data – die ungezähmte Macht

Die 10 Prinzipien für offene Daten

33

### 1. Vollständigkeit

Sämtliche Rohdaten eines öffentlichen Datensatzes sollen in vollem Umfang zur Verfügung gestellt werden – bestenfalls unter Angabe von Formeln sowie Erklärungen zur Datenberechnung. Als öffentliche Daten werden hierbei Daten verstanden, die nicht berechtigten Datenschutz-, Sicherheits- oder Zugangsbeschränkungen unterliegen.

### 2. Primärquellen

Die Daten sollen an ihrem Ursprung gesammelt werden, und zwar mit dem höchstmöglichen Feinheitsgrad, also nicht in aggregierten oder anderweitig modifizierten Formaten.

### 3. Zeitliche Nähe

Daten werden so zügig wie zur Werterhaltung notwendig zur Verfügung gestellt – bestenfalls direkt nach deren Erhebung und Zusammenstellung.

### 4. Leichter Zugang

Daten werden so vielen Nutzern wie möglich für möglichst viele Verwendungszwecke bereitgestellt. Die gewünschten Informationen sollen hierbei so leicht als möglich eingeholt werden können.

### 5. Maschinenlesbarkeit

Um Daten automatisiert verarbeiten zu können, müssen die Informationen strukturiert in den weitverbreiteten Datenformaten zur Verfügung gestellt werden.

### 6. Diskriminierungsfreiheit

Ohne vorherige Identifikation (Registrierung oder sonstige Mitgliedschafts-Voraussetzungen) sollen die Daten jedem zu jeder Zeit zur Verfügung stehen.

### 7. Verwendung offener Standards

Daten werden in standardisierten, kostenlos verfügbaren Formaten bereitgestellt, über die keine juristische Person die alleinige Kontrolle hat.

### 8. Lizenzierung

Daten unterliegen keinen Nutzungsbeschränkungen wie Urheberrecht, Patent, Markenzeichen oder Geschäftsgeheimnis. Sinnvolle Datenschutz-, Sicherheits- und Zugangsbeschränkungen sind jedoch zulässig.

### 9. Dauerhaftigkeit

Öffentliche Datenströme sollen dauerhaft in Online-Archiven verfügbar gemacht werden.

### 10. Niedrige Nutzungskosten

<http://sunlightfoundation.com/blog/2011/07/14/vivek-kundras-10-principles-for-improving-federal-transparency/>.

<http://wiki.okfn.de/10-Prinzipien-fuer-offene-Daten>

In Großbritannien sind aber nicht nur die Steuerausgaben online abrufbar. Open Data auf lokaler Ebene wird der Allgemeinheit auf der Internetseite OpenlyLocal bereits in vergleichsweise hohem Maße zur Verfügung gestellt.<sup>69</sup> So hat immerhin schon knapp ein Viertel der britischen Gemeinden ihre lokalen Verwaltungsdaten hier öffentlich zugänglich gemacht, und täglich kommen mehr Kommunen hinzu. Zu den veröffentlichten Daten zählen neben den aufgeschlüsselten Gemeindeausgaben auch alle sonstigen öffentlichen Gemeinderatsinformationen sowie Bevölkerungsstatistiken oder lokale Finanzdaten.

Eine Organisation aus London (**MySociety**) hat die Plattform „Fix my Street“ ins Leben gerufen. Hier können die Bürger über eine App Schlaglöcher, defekte Ampeln oder Straßenlaternen direkt an die zuständigen Abteilungen melden. So senden die Bürger ein Foto des Schlaglochs via Smartphone an die entsprechende Behörde, die dann anhand der GPS-Daten einen Reparaturauftrag erteilen kann. Der Bürger bekommt regelmäßig Updates über den Reparaturfortschritt. Diese Plattform wird mittlerweile in vielen europäischen Städten nachgeahmt. Von derartigen Plattformen können also sowohl die Stadtverwaltung als auch die Bürger profitieren – nicht selten scheitert aber die zeitnahe Ausbesserung am Budgetmangel der Kommunen.

In Deutschland fehlt (noch) eine Datenbank, in der alle Gemeindedaten gesammelt werden. Aber vereinzelte deutsche Kommunen, wie Köln<sup>70</sup> oder Frankfurt am Main, bieten inzwischen solche Tools für ihre Stadt an. Auf der von Bürgern geschaffenen elektronischen Plattform „Frankfurt gestalten – Bürger machen Stadt“<sup>71</sup> (ebenfalls ein Projekt der OKF) werden z.B. lokale Informationen der Stadt Frankfurt kostenlos online zur Verfügung gestellt. Die Bürger können sich über die Schnittstelle **OpenStreetMap** aber auch über eine gezielte Filterung der vorhandenen Datensätze zur Lokalpolitik und sonstige Vorkommnisse z.B. Polizei- oder Baustellenmeldungen informieren. Diverse Foren bieten den Bürgern zudem die Möglichkeit, sich über lokale Entwicklungen auszutauschen.

Die genannten Projekte zeigen, dass sich Open Data als alternative Datenverwendung in der Big-Data-Diskussion nicht nur um „Geld verdienen“ dreht, sie stellen vielmehr den gesellschaftlichen Nutzen in den Vordergrund. Somit stellt sich die Frage, inwiefern die Big-Data-Diskussion auch zusätzliche Impulse freizusetzen vermag, um Open Data bzw. die Schaffung von Werten voranzutreiben. Die Öffnung bislang nicht zugänglicher Daten ist nicht nur für den Verwaltungssektor relevant, sondern hat sicherlich auch für andere Bereiche Potenzial. Beispielsweise ist im Bildungswesen (Open Research Data) eine Öffnung schon dahingehend wahrnehmbar, dass Lehrstühle ihre Vorlesungen, ganze Vortragsreihen und die dazugehörigen Unterlagen als Youtube-Video oder auf der jeweiligen Universitäts-Homepage der Allgemeinheit zur Verfügung stellen (wie z.B. die Online-Plattform MITOpenCourseware<sup>72</sup> des Massachusetts Institute of Technology [MIT]).

<sup>69</sup> <http://openlylocal.com/>.

<sup>70</sup> <http://www.offenedaten-koeln.de/>.

<sup>71</sup> <http://www.frankfurt-gestalten.de/>.

<sup>72</sup> <http://ocw.mit.edu/index.htm>.



## 8. Grenzen von Big Data

### Big Data ist kein Allheilmittel

Neben dem datenschutzrechtlichen Nachholbedarf gibt es noch weitere Aspekte, welche die Grenzen und Schwächen moderner Big-Data-Analysemethoden aufzeigen. Mit den modernen Technologien und Methoden im Bereich Big Data eröffnen sich den Akteuren sicherlich neue Wege, bestehende Datenmodelle und Szenarien beispielsweise mit Echtzeit-Daten anzureichern, um somit aussagekräftigere Ergebnisse zu erlangen und bestenfalls Trends zu prognostizieren. Allerdings bewirkt Big Data auch keine Wunder, denn die Gesetze der Statistik lassen sich auch aufgrund der Fülle neuer Daten(korrelationen) nicht aushebeln. Nur weil datengetriebene Analysen plötzlich mit neuen, bisher ungeahnten quantifizierbaren Datensätzen angereichert werden können, sind die Ergebnisse nicht objektiver als in der Analysezeit vor Big Data, d.h. trotz der zunehmenden Quantifizierbarkeit menschlicher Handlungen, trotz der neuen maschinenlesbaren Kaufpräferenzen oder Gefühlsäußerungen der Menschen werden keine zwangsläufigen Tatsachen geschaffen. Dies gilt insbesondere hinsichtlich von Datensätzen aus dem Bereich sozialer Netzwerke.<sup>73</sup> Die Bewertung von Analyseergebnissen aus wissenschaftlichen Datenanalysen, seien die Analysen auch noch so quantifizierbar und leidenschaftslos, sie bleiben Interpretationen von Menschen (Subjekten), die auch irren können.

### Steigende Komplexität bei der Verschmelzung von Datensätzen

#### Steigende Komplexität bei Datenkombinationen

Darüber hinaus gibt es weiterhin methodologische Probleme, wie die Unzuverlässigkeit, die Fehleranfälligkeit und Unvollständigkeit großer Datensätze. Diese Probleme verstärken sich zudem, wenn unterschiedliche Datenquellen kombiniert werden (Datenabgrenzungsproblem). Bei der Zusammenführung diverser Datensätze steigt mit jeder Korrelation die Komplexität der zugrundeliegenden Datenbestände. Das Zusammenführen bedeutet nicht, dass es (nur) zu einer einfachen Addition dieser Daten kommt. Wenn z.B. zehn Informationen aus Umweltsensoren mit zehn Informationen aus dem Verkehrsaufkommen verknüpft werden, resultieren am Ende nicht 20 neue Datensätze, sondern sie multiplizieren sich zu 100 neuen Informationen, die wiederum neu interpretiert werden können. Damit steigen die Komplexität sowie die Anforderungen an den neu entstandenen Datensatz stark an, mit der impliziten Gefahr, aus der zugrundeliegenden Analyse Muster erkennen zu wollen, die gar nicht existieren. Dies könnte zu fehlerhaft abgeleiteten Handlungen führen.

### Ursache und Kausalität

#### Big Data liefert keine Kausalität, sondern nur Korrelation

Obwohl die Daten immer „bigger“ werden, gilt nach wie vor „Qualität vor Quantität“. Big Data kann nicht mit Vollständigkeit gleichgesetzt werden. Die zur Verfügung stehenden Informationen und Datenreihen können unter Zuhilfenahme der bereits vorhandenen Analysetools auf Korrelationen untersucht werden. Doch sind diese durch Algorithmen gefundenen Zusammenhänge nicht immer logisch. So kann die Kausalität der gefundenen Zusammenhänge auch mit dem heutigen Stand der Technik nicht belegt oder hinterfragt werden, sondern dies obliegt nach wie vor dem menschlichen Verstand, seinen Erfahrungen und Intuitionen. Der Mensch muss herausfinden, welche Fragestellungen und Antworten die Daten hergeben können. Algorithmen stellen also lediglich ein Kombinationsstool dar und ermitteln, WELCHE Zusammenhänge die Daten aufweisen, nicht aber, WARUM es diese Zusammenhänge gibt. Datenanalysen bleiben also die Entscheidungsbasis, die Entscheidungsmacht hingegen sollte nicht auf

<sup>73</sup> Vgl.: Boyd, D., Crawford, K. (2013). Big Data als kulturelles, technologisches und wissenschaftliches Phänomen. Sechs Provokationen.





## Big Data – die ungezähmte Macht

---

die Algorithmen übertragen werden. Big Data beruht zwar auf Korrelationen, ist zur Beurteilung deren Kausalität aber ungeeignet.<sup>74</sup>

### Wo bleibt die menschliche Intuition?

#### Es droht der informationelle Missbrauch

Die neuen Technologien und Methoden könnten die Menschen dazu verleiten, Zusammenhänge nicht mehr nur durch Zuhilfenahme bzw. Entwicklung von Modellen zu finden, sondern Daten zunehmend für sich sprechen zu lassen, ohne weitere Dateninterpretationen. Dies könnte die Menschen langfristig in ihrer Denkweise beeinflussen, weil sie sich zu stark auf Big Data verlassen, ohne Zusammenhänge zu hinterfragen. Die Denkmuster werden also potenziell in bestimmte Bahnen gelenkt, was vor allem die menschliche Intuition und Kreativität negativ beeinflussen kann. Beide Eigenschaften stellen jedoch zentrale Größen im Innovationsprozess dar. Dürfen Daten also ohne weitere Interpretationen Handlungsempfehlungen automatisch ableiten? Eine potenziell präventive Nutzung von Datenauswertungen sollte nicht gängige Praxis werden. In einem Extremszenario würde die Polizei z.B. eine Person verstärkt überwachen, nur weil sie aufgrund ihres aktuellen „digitalen Fußabdrucks“ (Tweets, Suchanfragen, Kreditkarte- oder Online-Käufe etc.) in ein durch Algorithmen ermitteltes typisches Verhaltensmuster früherer Bankräuber passt, ohne dass die Person kriminell geworden ist oder dies jemals beabsichtigen würde. Natürlich klingt es zunächst verlockend, wenn mit Hilfe von Big Data ein Verbrechen vorhergesagt werden kann, bevor es verübt wird. Tatsächlich ist es jedoch informationeller Missbrauch, wenn Menschen aufgrund flächendeckender und anlassloser Profilerstellung plötzlich in elektronisch ermittelte Raster fallen und per se unter Generalverdacht gestellt werden. Ein solches Vorgehen gilt generell als Gefahr für ein freiheitliches und demokratisches System und wird daher nicht erst seit den jüngst bekannt gewordenen Spähaffären diskutiert.

### Fragwürdige Datensätze

#### Daten von sozialen Netzwerken liefern keine allgemeingültigen Aussagen über menschliches Verhalten

Einige Akteure, darunter auch Wissenschaftler, nutzen gerne Datensätze von sozialen Netzwerk-Plattformen, um möglichst neue Muster aus Stimmungen und Gefühlslagen der Menschen zu filtern, um mediale Events, politische Proteste oder sonstige gesellschaftliche Bewegungen erklären oder bestenfalls vorherzusagen zu können. Aber selbst die Analyse solcher umfangreichen Datensätze erlaubt nicht das Ableiten allgemeingültigen menschlichen Verhaltens. Auch ein großer Teil z.B. aus der Facebook-Gemeinde ist lediglich eine spezifische Teilmenge. Je nach Analyse und abgeleiteter Botschaft unterscheiden sich die Nutzer sozialer Netzwerke von anderen Menschen. Hier spielen unterschiedliche Identitäten, wie Alter, Geschlecht, Bildungsgrad, Internet-Affinität, die Bereitschaft, persönliche Bilder und Botschaften zu posten, etc., eine große Rolle.

Zudem kann nicht mit Sicherheit festgestellt werden, ob es sich bei den offiziell angegebenen Kontenzahlen sozialer Netzwerke nicht auch um „Fake“-Konten handelt bzw. ob Nutzer und Konto übereinstimmen. Weiterhin kann nicht ausgeschlossen werden, dass einzelne soziale Netzwerk-Plattformen problematische Inhalte der Nutzer zensurieren und somit die Aussagekraft der Meldungen von Anfang an verzerren. Diese methodologischen Fragen müssen im Hinterkopf behalten werden, um zu entscheiden, welche Fragen bzw. welche Analyse und darauf abzielend, welche Interpretationen am Ende der Analyse noch zulässig sind.<sup>75</sup> Daher gilt trotz oder gerade wegen der Big-Data-Bewegung: Auch Smart Data kann zu guten Ergebnissen führen.

---

<sup>74</sup> Vgl.: Mayer-Schönberger, V. und Cukier, K. (2013).

<sup>75</sup> Vgl.: Boyd, D., Crawford, K. (2013).



## Ethik und Moral

Was bedeutet eigentlich „öffentlich zugänglich“?

Eine ausführliche Diskussion über Ethik und Moral innerhalb der Big-Data-Diskussion würde den Rahmen dieser Studie sprengen. Aber natürlich spielen auch ethische und moralische Fragestellungen eine große Rolle, wenn es darum geht, personenbezogene Datensätze aus unterschiedlichen Quellen und Kontexten teils mehrfach und durch unterschiedliche Akteure zu verwerten. Viele Fragen in der Big-Data-Diskussion z.B. bezüglich Zugang, Zweck, Kontrolle, Machtverhältnissen oder Wahrheitsgehalt bleiben unbeantwortet. Welchen Status haben z.B. Daten, die auf diversen sozialen Netzwerk-Plattformen als „öffentlich“ gekennzeichnet sind, also nicht nur für Freunde zugänglich sind, sondern quasi für alle Internetnutzer? Dürfen Interessierte diese Datensätze für Analysen verwenden oder gibt es einen Probandenschutz, was dazu führen würde, dass von jedem Einzelnen zuvor eine informierte Einwilligung eingeholt werden muss? Die Tatsache, dass es digitale Inhalte gibt, die öffentlich zugänglich sind, bedeutet nicht automatisch, dass jeder mit diesen Inhalten machen kann, was er möchte. Das Internet ist kein rechtsfreier Raum.

## 9. Fazit

Big Data ist nicht aufzuhalten

Die Datenflut ist Realität. Big Data lässt sich ebenso wie die Globalisierung nicht aufhalten. Mit Big Data ist auch die IT-Sicherheit ein entscheidender Teil des gesamten Wertschöpfungsprozesses geworden. Während sich die Akteure früher eher mit der Limitiertheit von Daten herumgeschlagen haben, kämpfen sie heute mit der Analyse-Limitiertheit bzw. der Frage, welche wertvollen Informationen den anfallenden Datenmengen noch entlockt werden können.

Obwohl sich die Entscheidungsträger durchaus im Klaren sind, dass Big Data mittel- bis langfristig ein Thema mit hoher strategischer Relevanz und lukrativen Wachstumschancen ist, mangelt es in vielen Fällen an geeigneten Digitalisierungsstrategien, geschultem Personal für die neuen Herausforderungen sowie den notwendigen adäquaten Management-Kompetenzen. Aufgrund fehlender oder veralteter Datenschutzbestimmungen sowie (noch) eher starren Silodenkens bestehen nach wie vor auch Hemmungen, im Bereich Big Data experimentell aktiv zu werden.

Es entstehen neue Jobs und Berufe

Big Data wird die Struktur bestehender Wirtschaftszweige verändern, weil es neue Wettbewerbskonstellationen geben wird. Viele etablierte Geschäftsmodelle werden durch den digitalen Strukturwandel aus den Angeln gehoben und bekommen in ihren jeweiligen Branchen künftig mehr Konkurrenz von branchenfremden Unternehmen, die sich auf Produkte und Dienste im Bereich web-basierter Informations- und Kommunikationstechnologien bzw. auf die Analyse von Daten spezialisiert haben. Zudem werden neue Berufszweige entstehen, wie z.B. Datenanalysespezialisten oder Algorithmiker.

Insbesondere breiten sich Start-Ups und Nischenanbieter aus, die sich auf die Auswertung und Analyse unterschiedlicher, teils öffentlich zugänglicher Daten spezialisieren und daraus Produkte und Dienste kreieren. Dadurch erhöhen sie den Druck auf etablierte Geschäftsmodelle. Eine Applikation für web-basierte Endgeräte, die das individuelle Mediennutzungs- bzw. Leseverhalten der Menschen auswertet, kann in kürzerer Zeit bessere Prognosen darüber abgeben, was die Menschen gerne lesen, als es ein Verlagshaus mit jahrelanger Erfahrung im Kundenservice je konnte.

Daten sprechen lassen

Mühsam und lang erarbeitetes Expertenwissen innerhalb etablierter Geschäftsbereiche wird künftig schneller in Frage gestellt, weil die Geschäftsmodelle neu in den Markt eintretender Technologie-Unternehmen darauf basieren, dass intelligente Algorithmen in kürzerer Zeit und zu geringeren Kosten Prognosen mit einer höheren Eintrittswahrscheinlichkeit abgeben als einzelne Experten, die ihr Wissen jahrelang aufbauen mussten.



## Big Data – die ungezähmte Macht

---

Dies wird rasch zu mehr Konkurrenz sowie einem Umdenken im Markt führen. Etablierte Geschäftsmodelle werden zum Teil schmerzhaft Transformationsprozesse durchlaufen müssen. Zudem werden sich die Qualifizierungsanforderungen auf dem Arbeitsmarkt anpassen, d.h. Wirtschaft, Wissenschaft und auch der öffentliche Sektor stellen aufgrund der gestiegenen Dynamik durch den digitalen Strukturwandel höhere Anforderungen an künftige Mitarbeiter. Bei einer steigenden Nachfrage nach Big-Data-Methoden haben Quereinsteiger aus den Bereichen Statistik, Mathematik, Informatik, Datenanalyse oder künstliche Intelligenz und Robotik gute Chancen, einen lukrativen Job zu bekommen, weil sie branchenübergreifend ihre Fähigkeiten einsetzen können.

### Viele Fragen bleiben unbeantwortet

Bei all dem Potenzial und den positiven Auswirkungen, die aus der Big-Data-Diskussion abgeleitet werden können, bleiben dennoch zentrale Fragen unbeantwortet: Welchen Einfluss werden die modernen Technologien und Analysemethoden künftig auf den Alltag haben? Viele Branchen befinden sich bereits in schmerzhaften Anpassungsprozessen, welche weiteren Branchen sind betroffen und mit welchen Digitalisierungs-Strategien können sie drohenden Wettbewerbsnachteilen entgegenwirken? Wie werden die Akteure und die Bürger künftig mit der Schieflage in der Datenhoheit umgehen? Welche Rolle spielt der Wirtschaftsstandort Deutschland in einer zunehmend digitalisierten, global vernetzten Welt? Als wissensintensive und rohstoffarme Volkswirtschaft hat Deutschland optimale Voraussetzungen, im Bereich moderner Internettechnologien, vor allem im Bereich der IT-Sicherheit, künftig eine internationale Vorreiterrolle einzunehmen.

### Die digitale Welt braucht Regeln

Kommerziell und informationstechnologisch mag die Big-Data-Debatte bereits eine Grundlage haben. Jetzt geht es darum, die zentralen grundrechtlichen und auf Freiheit basierenden Datenschutzaspekte sowie die eventuell drohenden wettbewerbsverzerrenden Effekte durch die digitalen Ökosysteme zu diskutieren. Letzteres spielt vor allem deshalb eine Rolle, weil die bereinigenden Marktkonsolidierungen in der digitalen Welt der großen Internetkonzerne sehr viel schneller vonstatten gehen als z.B. in der Automobilbranche. Entscheidungsträger aus den Bereichen Wirtschaft, Wissenschaft, Politik oder Bürgerrechtsinitiative sind jetzt aufgefordert, die künftig notwendigen regulativen Rahmenbedingungen aktiv mitzugestalten. Insbesondere bei der Fragestellung, wie die datenschutzrechtliche Zweckbindung und die bereits praktizierten Big-Data-Methoden in Einklang gebracht werden können.

### Die vielen unterschiedlichen Interessen müssen jetzt diskutiert werden

Das Internet der Dinge wird kommen, humanoide Roboter werden früher oder später in vielen Haushalten für mehr Effizienz und Unterstützung sorgen. Die Menschen werden in Zukunft Dialoge mit ihren Computern führen (künstliche Intelligenz), um ihren Alltag zu optimieren. Die spannende Frage dabei ist nicht, wann das alles eintreffen wird, sondern wie die (künftigen) Generationen mit den modernen Technologien umgehen werden, welche Rahmenbedingungen sie vereinbaren werden, welche Grenzen definiert werden und ob die Menschen darauf vertrauen können, dass ihre Rechte durchsetzbar sind.

Die unterschiedlichen Hoffnungen und vielgepriesenen Potenziale, aber auch die durchaus berechtigten Ängste machen deutlich, dass die Entwicklungen und Einsatzmöglichkeiten im Bereich Big Data facettenreich sind. Am Ende lassen sich viele unterschiedliche, kommerzielle, ehrbare, aber auch kriminelle oder fragwürdige Ziele mit Big Data verfolgen. Jedem positiven Beispiel lassen sich im selben Atemzug auch besorgniserregende, jedoch durchaus realistische Szenarien zuordnen.

### Gefahr des „Loss of trust“

Mit den notwendigen (idealerweise international geltenden) Rahmenbedingungen soll eine Balance geschaffen werden, welche die Chancen der Entwicklung von Big Data beflügeln, ohne die Risiken zu stark in den Mittelpunkt zu stellen. Dieser Balanceakt wird kein einfacher sein, sollte aber möglichst jetzt in diesem frühen Entwicklungsstadium auf internationaler Ebene angegangen



## Big Data – die ungezähmte Macht

---

werden. Für viele künftige Probleme und Fragen im Bereich Big Data wird es keine perfekten Lösungen oder endgültigen Antworten geben. Die Herausforderungen liegen mitunter darin, einen Weg zu finden, moderne Technologien und Methoden nutzenstiftend in den Alltag der Menschen zu integrieren, ohne dass

- Freiheitsrechte und demokratische Leitlinien verletzt werden,
- Diskriminierungen oder Manipulationen stattfinden oder
- die Angst der Menschen steigt, sich auf digitalen Kanälen aufzuhalten, weil unterschiedliche Akteure massenhaft private Daten ausspähen.

Es liegt an uns, als wertvolle und kreative Ideenlieferanten hinter jeder (r)evolutionären Technologie bzw. Innovation, die Zügel fest in der Hand zu behalten, um einen adäquaten Kurs vorzugeben. Möge die Macht mit uns sein.

Thomas F. Dapp (+49 69 910-31752, thomas-frank.dapp@db.com)

Veronika Heine

### Literaturverzeichnis

- Bahr, F. et al. (2012). Schönes neues Internet? Chancen und Risiken für Innovation in digitalen Ökosystemen. Policy Brief. Stiftung Neue Verantwortung. Berlin.
- Beck, S., Grzegorzek, M. et al. (2013). Mit Robotern gegen den Pflegenotstand. Policy Brief 04/13. Stiftung neue Verantwortung. Berlin.
- Big Data – Vorsprung durch Wissen. Innovationspotenzialanalyse (2012). Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS. Sankt Augustin bei Bonn.
- Boyd, D., Crawford, K. (2013). Big Data als kulturelles, technologisches und wissenschaftliches Phänomen. Sechs Provokationen. In Geiselberger, H., Moorstedt, T. (2013). Big Data. Das neue Versprechen der Allwissenheit. eBook Suhrkamp Verlag Berlin.
- Brown, B. et al. (2011). Big data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute.
- Dapp, T. (2012). Der vermessene Mensch. Biometrische Erkennungsverfahren und mobile Internetdienste. Aktuelle Themen. Deutsche Bank Research. Frankfurt am Main.
- Dapp, T. (2013). Die Zukunft des (mobilen) Zahlungsverkehrs: Banken im Wettbewerb mit neuen Internet-Dienstleistern. Aktuelle Themen. Deutsche Bank Research. Frankfurt am Main.
- Dapp, T. und Heymann, E. (2013). Dienstleistungen 2013. Heterogener Sektor verzeichnet nur geringe Dynamik. Aktuelle Themen. Deutsche Bank Research. Frankfurt am Main.
- Geiselberger, H., Moorstedt, T., Hrsg. (2013). Big Data. Das neue Versprechen der Allwissenheit (2013). eBook Suhrkamp Verlag Berlin.
- Ginsberg, Jeremy et al. (2009). Detecting influenza epidemics using search engine query data. Google Inc. Centers for Disease Control and Prevention in *Nature* Vol. 457.
- Hayek, F. A. von (2004). Der Weg zur Knechtschaft. 4. Auflage. Mohr Siebeck Verlag. Tübingen.
- Heuer, S. (2013). Kleine Daten, große Wirkung. Big Data einfach auf den Punkt gebracht. DigitalKompakt LfM #06. Landesanstalt für Medien Nordrhein-Westfalen (LfM). Düsseldorf.





## Big Data – die ungezähmte Macht

---

- Kranzberg, M. (1986). Technology and history: Kranzberg's laws. In *Technology and Culture*. 27/2. S. 544-560.
- Manig, M. und Giere, J. (2012). Quo Vadis Big Data – Herausforderungen – Erfahrungen – Lösungsansätze. TNS Infratest GmbH. München.
- Mayer-Schönberger, V., Cukier, K. (2013). Big Data. Die Revolution, die unser Leben verändern wird. Redline Verlag. München.
- Morozov, E. (2013). Big Data. Warum man das Silicon Valley hassen darf. Frankfurter Allgemeine. Feuilleton vom 10.11.2013.
- Schaar, P. (2013). Big Brother und Big Data – Was heißt eigentlich Datenschutz auf Amerikanisch? Rede anlässlich der 41. Römerberggespräche am 26. Oktober 2013. Frankfurt am Main.
- Schroek, M. et al. (2012). Analytics: Big Data in der Praxis. Wie innovative Unternehmen ihre Datenbestände effektiv nutzen. IBM Institute for Business Value in Zusammenarbeit mit der Saïd Business School, Universität Oxford.
- Weichert, T. (2013). Big Data – eine Herausforderung für den Datenschutz. In Geiselberger, H., Moorstedt, T. (2013). Big Data. Das neue Versprechen der Allwissenheit. eBook Suhrkamp Verlag Berlin.
- Weichert, T. (2013). Big Data und Datenschutz. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein.

### Weiterführende Literatur

- Becker, K., Stalder, F. (2010). Deep Search. Politik des Suchens jenseits von Google. Bundeszentrale für politische Bildung. [www.bpb.de](http://www.bpb.de). Bonn.
- Bloching, B. et al. (2012). Data Unser. Wie Kundendaten die Wirtschaft revolutionieren. Redline Verlag. München.
- Kurz, C., Rieger, F. (2012). Die Datenfresser. Wie Internetfirmen und Staat sich unsere persönlichen Daten einverleiben und wie wir die Kontrolle darüber zurückerlangen. Fischer Verlag GmbH. Frankfurt am Main.
- Schmidt, J., Weichert, T. (2012). Datenschutz. Grundlagen, Entwicklungen und Kontroversen. Bundeszentrale für politische Bildung. Bonn.
- Shapiro, C. und Varian, H. (1999). Information Rules. A Strategic Guide to the Network Economy. Harvard Business Review Press, Boston Massachusetts.



## Fokusthema Deutschland

- ▶ **Ausblick Deutschland:  
2% Wachstum in 2015 trotz belastender  
Beschäftigungspolitik  
(Aktuelle Themen – Konjunktur)** ..... 28. Februar 2014
- ▶ **Macht Hollande den Schröder –  
und wird es funktionieren?  
(Standpunkt Deutschland)** ..... 24. Februar 2014
- ▶ **Industrie 4.0: Upgrade des Industriestandorts  
Deutschland steht bevor  
(Aktuelle Themen – Branchen)** ..... 4. Februar 2014
- ▶ **Ausblick Deutschland:  
Weiter im Aufwind  
(Aktuelle Themen – Konjunktur)** ..... 27. Januar 2014
- ▶ **Große Koalition – kleinmütige Politik  
(Standpunkt Deutschland)** ..... 16. Dezember 2013
- ▶ **Deutsche Leistungsbilanzüberschüsse –  
Kritik nicht stichhaltig  
(Standpunkt Deutschland)** ..... 12. Dezember 2013
- ▶ **Ausblick: Deutschland:  
Aufbruch in die Vergangenheit  
(Aktuelle Themen – Konjunktur)** ..... 29. November 2013
- ▶ **Ausblick Deutschland:  
Trotz Wachstum in der Kritik  
(Research Briefing – Konjunktur)** ..... 19. November 2013
- ▶ **Mindestlohn von EUR 8,50:  
Eine falsche Weichenstellung  
(Standpunkt Deutschland)** ..... 4. November 2013
- ▶ **Ausblick Deutschland:  
Von Übertreibungen und Ängsten  
(Aktuelle Themen – Konjunktur)** ..... 31. Oktober 2013

Unsere Publikationen finden Sie unentgeltlich auf unserer Internetseite [www.dbresearch.de](http://www.dbresearch.de). Dort können Sie sich auch als regelmäßiger Empfänger unserer Publikationen per E-Mail eintragen.

Für die Print-Version wenden Sie sich bitte an:  
Deutsche Bank Research  
Marketing  
60262 Frankfurt am Main  
Fax: +49 69 910-31877  
E-Mail: [marketing.dbr@db.com](mailto:marketing.dbr@db.com)

Schneller via E-Mail:  
[marketing.dbr@db.com](mailto:marketing.dbr@db.com)

© Copyright 2014. Deutsche Bank AG, Deutsche Bank Research, 60262 Frankfurt am Main, Deutschland. Alle Rechte vorbehalten. Bei Zitaten wird um Quellenangabe „Deutsche Bank Research“ gebeten.

Die vorstehenden Angaben stellen keine Anlage-, Rechts- oder Steuerberatung dar. Alle Meinungsäußerungen geben die aktuelle Einschätzung des Verfassers wieder, die nicht notwendigerweise der Meinung der Deutsche Bank AG oder ihrer assoziierten Unternehmen entspricht. Alle Meinungen können ohne vorherige Ankündigung geändert werden. Die Meinungen können von Einschätzungen abweichen, die in anderen von der Deutsche Bank veröffentlichten Dokumenten, einschließlich Research-Veröffentlichungen, vertreten werden. Die vorstehenden Angaben werden nur zu Informationszwecken und ohne vertragliche oder sonstige Verpflichtung zur Verfügung gestellt. Für die Richtigkeit, Vollständigkeit oder Angemessenheit der vorstehenden Angaben oder Einschätzungen wird keine Gewähr übernommen.

In Deutschland wird dieser Bericht von Deutsche Bank AG Frankfurt genehmigt und/oder verbreitet, die über eine Erlaubnis der Bundesanstalt für Finanzdienstleistungsaufsicht verfügt. Im Vereinigten Königreich wird dieser Bericht durch Deutsche Bank AG London, Mitglied der London Stock Exchange, genehmigt und/oder verbreitet, die in Bezug auf Anlagegeschäfte im Vereinigten Königreich der Aufsicht der Financial Services Authority unterliegt. In Hongkong wird dieser Bericht durch Deutsche Bank AG, Hong Kong Branch, in Korea durch Deutsche Securities Korea Co. und in Singapur durch Deutsche Bank AG, Singapore Branch, verbreitet. In Japan wird dieser Bericht durch Deutsche Securities Limited, Tokyo Branch, genehmigt und/oder verbreitet. In Australien sollten Privatkunden eine Kopie der betreffenden Produktinformation (Product Disclosure Statement oder PDS) zu jeglichem in diesem Bericht erwähnten Finanzinstrument beziehen und dieses PDS berücksichtigen, bevor sie eine Anlageentscheidung treffen.

Druck: HST Offsetdruck Schadt & Tetzlaff GbR, Dieburg