# Big Data und Datenschutz: Wahrnehmung und Technik

Martin Steinebach

Fraunhofer SIT
steinebach@sit.fraunhofer.de

## Zusammenfassung

Big Data wird in den Medien oft als Revolution in der IT bezeichnet und von vielen Wirtschaftsvertretern als notwendige Technologie für zukünftige Wertschöpfungsketten gesehen. Gleichzeitig sieht die Öffentlichkeit Big Data verbreitet als Bedrohung der Privatsphäre und Herausforderung für den Datenschutz. Der Beitrag beschreibt die Auswertung einer Online-Befragung zum Thema Big Data und Datenschutz und liefert Grundlagen zum Thema, um die Ergebnisse in einem geeigneten Kontext diskutieren zu können.

# 1 Big Data: Einführung und Definition

Unter Big Data wird das Erheben, Speichern, Zugreifen und Analysieren von großen und teilweise heterogenen, strukturierten und unstrukturierten Datenmengen verstanden. Big Data stellt eine neue Herangehensweise an den Umgang mit großen Datenmengen dar. Durch neue Algorithmen, die selbstständig Muster und Zusammenhänge in Daten erkennen können, und durch neue Hardware-Lösungen, die in der Lage sind, eine große Datenmenge zeitnah zu verarbeiten, werden die Möglichkeiten für Datenanalysen erheblich vervielfältigt. Das volle Potenzial entfaltet Big Data dann, wenn Analysten in Echtzeit Zusammenhänge in Daten herstellen und prüfen können, um neue Erkenntnisse aus den Daten zu gewinnen. Auch die Datenquellen, die als Basis für die Analysen dienen, sollten möglichst aktuell sein und als kontinuierlicher Fluss von Informationen dem System zugeführt werden.

## 1.1 Wozu Big Data?

Big Data kann in den unterschiedlichsten Domänen eingesetzt werden. So helfen Big-Data-Anwendungen im Gesundheitswesen und der medizinischen Forschung mittels datengestützter Diagnose und Behandlung. Ebenso werden Big-Data-Technologien für Wettervorhersagen und Klimamodelle verwendet, um dynamische und möglichst echtzeitfähige Modelle zu erstellen. Auch in der Weltraumforschung und bei Teilchenbeschleunigern wird Big Data genutzt. Weitere Anwendungsfelder ergeben sich bei Sicherheits- und Polizeiarbeit sowie bei der Infrastruktur von Mobilfunknetzen, Internet und intelligenten Stromnetzen (sog. Smart Grids). Auch für Meinungs- und Trendforschung mittels Daten aus sozialen Medien verspricht Big Data enormes Potenzial. Offensichtliche Anwendungsmöglichkeiten für Big Data bestehen vor allem in den Bereichen Wirtschaft und Konsum. Ob bei Werbung, Kundenbindung und -analyse oder im Kreditwesen, in der Finanz- und Versicherungsmathematik oder bei der sogenannten Business Intelligence für Unternehmen – Big Data findet hier vielfältige Einsatz- und Optimierungsmöglichkeiten. Einen ausführlichen Blick auf die Möglichkeiten bietet beispielsweise die Studie des BITKOM [BITK14].

P. Schartner · K. Lemke-Rust · M. Ullmann (Hrsg.) · D•A•CH Security 2015 · syssec (2015) 429-440.

## 1.2 Definition

Der Begriff Big Data wird im Kontext der Informationsgewinnung aus „großen“ Datenbeständen verwendet. Dabei ist nicht der bloße Umfang eines Datenbestandes entscheidend, sondern die Kombination verschiedener technischer Herausforderungen für die Datenverarbeitung im Kontext einer immensen „Datenflut“. Big Data wird oft durch die Eigenschaften Volume, Velocity, Variety (kurz „3V“) und die damit verbundenen Herausforderungen charakterisiert, was auf Doug Laney zurückgeht [Lane01].

Teilweise werden weitere Aspekte hinzugefügt und mit einem „V“ beschrieben, beispielsweise die folgenden: **Veracity** steht für Vertrauenswürdigkeit der Daten oder der gezogenen Schlüsse. **Value** betont, dass Big Data letztendlich immer eine Wertschöpfung der Daten beabsichtigt. **Visualization** stellt die intuitive Darstellung der Ergebnisse heraus.

Je nach Anwendung treffen die genannten Eigenschaften mehr oder weniger zu. Gemeinsam ist allen Big-Data-Lösungen letztendlich, dass durch eine Verarbeitung von Daten neue Zusammenhänge erkannt und so neue Erkenntnisse gewonnen werden sollen. Dies ist zwar schon lange ein Ziel der Informatik, durch die technischen Fortschritte im Umfeld von Big Data ist eine Umsetzung allerdings deutlich einfacher geworden.

# 2 Herausforderung Datenschutz

Neue Technologien führen oft zu einem Spannungsfeld zwischen dem technisch Möglichen und dem ethisch Vertretbaren. Die Gesellschaft muss sich erst über die Konsequenzen der Technologie im Klaren werden und dann Regeln für den Umgang mit ihr finden. Ein Beispiel dafür aus der Vergangenheit ist die Situation des Urheberrechts im Internet. Als um das Jahr 2000 herum Dienste wie Napster Musik in Form von MP3-Dateien plötzlich frei verteilbar und so kostenfrei verfügbar machten, begann eine noch heute andauernde Diskussion um eine gerechte Wahrung verschiedener Interessen sowie deren technische und rechtliche Konsequenzen.

Auch Big Data führt zu neuen Herausforderungen im Umgang mit Daten. Konzepte, die ursprünglich als ausreichend zum Schutz der Privatsphäre betrachtet wurden, weichen auf, weil immer mehr Daten miteinander verknüpft werden können. Große Mengen unterschiedlicher Daten werden zusammengefügt, um neue Methoden der Wertschöpfung zu realisieren, ohne dabei von Anfang an auch Aspekte des Datenschutzes zu berücksichtigen.

Von Bedeutung ist hier auch die Sichtweise der Industrie: In einer Umfrage des BITKOM vom Februar 2014 [BITK14] wurde festgestellt, dass etwas mehr als die Hälfte aller befragten Unternehmen den Datenschutz als Hindernis für den Einsatz von Big Data sieht. Ähnlich wurden von den Unternehmen auch die Hürden durch die Anforderungen an die IT-Sicherheit gesehen – hier war es knapp die Hälfte der Unternehmen. In immerhin 17 Prozent der Unternehmen sind keine Prozesse für den Umgang mit personenbezogenen Daten festgelegt. Der BITKOM hat für die Studie 507 Unternehmen mit mindestens 50 Mitarbeitern befragt. Folglich gibt Drehmel [Dreh15], die Bereichsleiterin Datenschutz für den BITKOM, auch die Anregung, den Datenschutz an die Erfordernisse von Big Data anzupassen.

Letztendlich kann Big Data als ein Dual-Use-Phänomen gesehen werden, wobei allerdings weniger die Diskussion zwischen ziviler und militärischer Nutzung im Vordergrund steht, als vielmehr die zwischen gesellschaftlichem Nutzen und dem Risiko des Verlustes von Privatsphäre. Eine klare Abgrenzung nur auf Grundlage der entwickelten Algorithmen kann hier nicht getrof-

fen werden: Die gleichen Daten und Verfahren, die einem Arzt bei der Diagnose eine Krankheit helfen, können eine Krankenkasse dazu bewegen, einen Kunden abzulehnen oder seine Beiträge zu erhöhen. Nur wenn eine konkrete Anwendung diskutiert wird, kann hier eine Aussage getroffen werden, wie viel Chance und wie viel Risiko vorliegt.

## 2.1 Persönliche Daten und Big Data

Big Data ist ein Ansatz, dessen Umsetzung große Mengen von Daten erfordert. Hinsichtlich des Datenschutzes ist zu unterscheiden, ob diese Daten personenbezogen sind oder nicht. Personenbezogene Daten sind alle Daten, die auf eine bestimmbare Person hinweisen oder ihr zugeordnet sind. Einfache Beispiele sind körperliche Merkmale der Person, aber auch ihre Telefonnummer oder ihr Wohnort. Nicht personenbezogene Daten sind Daten, für die (auch in Zukunft) keine Zuordnung zu handelnden oder betroffenen Personen möglich ist. Das gilt unter anderem für Daten, die sich ausschließlich auf Geräte und Produkte, nicht aber auf ihre Nutzer beziehen, beispielsweise Sensordaten zur Ortung von Transportgütern in der automatisierten Logistik. In der Praxis stellt diese Trennung immer wieder einer Herausforderung dar. Durch weitere Datenquellen und eine geschickte Zusammenführung besteht immer wieder das Risiko, dass Daten personenbezogen werden. So kann eine Unterscheidung immer nur nach Stand des Wissens erfolgen.

## 2.2 Informelle Selbstbestimmung

Eine bedeutende Rolle in diesem Konflikt kommt dem Grundrecht auf informationelle Selbstbestimmung zu (BVerfGE 65,1). So ist festgelegt, dass der Bürger selbst bestimmen darf, welche Information über ihn zu welcher Zeit zur Verfügung stehen darf. In der Praxis ist dieses Recht auf Daten mit direktem Personenbezug beschränkt. Ein wesentlicher Aspekt bei Big Data ist jedoch, dass häufig Aussagen über Personengruppen gemacht werden sollen, wofür aber personenbezogene Daten herangezogen werden müssen. Es gibt hier einen Übergang von einer individuellen Selbstbestimmung („Was passiert mit *meinen* Daten?“) zu einer gesellschaftlichen Selbstbestimmung („Was passiert mit *unseren* Daten?“), und damit auch zu neuen Formen der Wahrnehmung dieser Selbstbestimmung. Die große Herausforderung in diesem Sinne ist, entsprechend zu differenzieren und gesetzliche Verbote bestimmter Verarbeitungen bspw. mithilfe von Ethikkommissionen auszusprechen.

## 2.3 Datensparsamkeit und Zweckbindung

Es sind vor allem die folgenden Prinzipien des Bundesdatenschutzgesetzes (BDSG), welche häufig in Zusammenhang mit Big Data diskutiert werden: Datensparsamkeit, Zweckbindung, Einwilligung und Auskunftsrecht sowie Eingriffsrecht. Datensparsamkeit fordert, dass bei der Verarbeitung personenbezogener Daten so wenige Daten wie möglich gesammelt, gespeichert und genutzt werden sollen. Dies soll nach Möglichkeit auch anonymisiert oder pseudonymisiert geschehen, wenn der Aufwand dazu nicht unverhältnismäßig hoch ist. Die Zweckbindung besagt, dass personenbezogene Daten, die für einen Zweck erhoben wurden, nicht ohne Weiteres für einen anderen Zweck verwendet werden dürfen.

Hier zeigt sich auch bereits ein Reibungspunkt mit der Praxis von Big Data: Oft werden erst einmal möglichst viele Daten gesammelt und analysiert, um so neue Erkenntnisse zu gewinnen. Weder die notwendige Datenmenge noch der exakte Zweck sind hier zum Zeitpunkt des Sammelns präzise bestimmbar. Oft werden hier entsprechend weit gefasste Formulierungen bei der

Zweckbindung verwendet, um ein breites Spektrum an Optionen offen zu lassen, was de facto ein Aufweichen der Vorgaben darstellt.

## 2.4 Einwilligung

Um personenbezogene Daten erheben zu dürfen, bedarf es nach dem BDSG entweder einer gesetzlichen Erlaubnis oder einer Einwilligung (§ 4a BDSG) durch den betroffenen Bürger („Verbot mit Erlaubnisvorbehalt“). Diese ist nur wirksam, wenn sie auf der freien Entscheidung des Betroffenen beruht. Er ist auf den vorgesehenen konkreten Zweck der Erhebung, Verarbeitung oder Nutzung sowie, soweit nach den Umständen des Einzelfalles erforderlich oder auf Verlangen, auf die Folgen der Verweigerung der Einwilligung hinzuweisen. Die Einwilligung bedarf der Schriftform, soweit nicht wegen besonderer Umstände eine andere Form angemessen ist. Solche Umstände sind beispielsweise gegeben, wenn eine hohe Dringlichkeit in einem Not- oder Krankheitsfall besteht. Hier genügt eine mündliche Einwilligung. Auch wenn die Daten direkt bei der Erfassung anonymisiert werden, reicht dies aus. Soll die Einwilligung zusammen mit anderen Erklärungen schriftlich erteilt werden, ist sie besonders hervorzuheben.

In der Praxis von Big Data dürfte Dringlichkeit selten vorherrschen. Dementsprechend ist entweder die Schriftform der Einwilligung notwendig, oder es muss direkt ein Anonymisieren erfolgen.

## 2.5 Auskunftsrecht

Auch nachdem personenbezogene Daten erhoben wurden, hat der Bürger Rechte. Das Auskunftsrecht (§ 34 BDSG) besagt, dass der Bürger das Recht hat, zu erfahren, welche Daten über ihn gespeichert werden und wozu. Auf Verlangen muss die verantwortliche Stelle ihm Auskunft erteilen über Herkunft und Art der gespeicherten Daten, den Empfängern dieser Daten und den Zweck der Speicherung. Eine entsprechende Anfrage kann jährlich und kostenfrei angefordert werden. Allerdings kann eine solche Auskunft eingeschränkt werden, wenn die Wahrung von Geschäftsinteressen der Daten erhebenden Instanz wichtiger als die Auskunftspflicht angesehen wird. An dieser Stelle gehen die Ziele von Big Data und die Privatheit des Einzelnen wieder weit auseinander. Es ist fraglich, ob ein einzelner Bürger hier seine Interessen durchsetzen kann.

Einige dokumentierte Selbstversuche zeigen, dass die Praxis teilweise nicht nur die Auskunft einschränkt, sondern Anfragen in einigen Fällen schlicht ignoriert werden. Einen aktuellen Überblick bietet die Seite www.selbstauskunft.de unter dem Punkt ’Erfahrungen’. Hier werden mehrere Unternehmen genannt, die auf Anfragen nicht reagieren. Allerdings fällt auf, dass teilweise Unternehmen sehr unterschiedlich bewertet werden.

## 2.6 Eingriffsrechte

Der Bürger kann auch aktiv gegen über ihn gespeicherte Daten vorgehen: Die Eingriffsrechte (§ 35 BDSG) geben ihm das Recht, falsche Daten berichtigen und bestimmte Daten löschen bzw. sperren zu lassen. Ähnlich zum Thema Auskunft bedeutet dies für Big-Data-Anwender, dass sie theoretisch jederzeit über die Daten derart verfügen können müssen, dass eine Korrektur, Sperrung oder Löschung ohne Weiteres möglich ist. Auch hier kann der Sammelnde widersprechen, beispielsweise wenn eine Löschung nur mit unverhältnismäßig hohem Aufwand geschehen kann. Die Verhältnismäßigkeit muss dann wieder individuell geklärt werden. Dementsprechend schwer dürfte es in der Praxis fallen, entsprechende Rechte durchzusetzen.

Das Urteil des EuGH [Bund14] hat allerdings gezeigt, dass Unternehmen durchaus zu entsprechenden Schritten gezwungen werden können.

# 3 Stand der Technik

Um bei personenbezogenen Daten datenschutzrechtliche Bestimmungen umzusetzen, existieren eine Reihe von technischen Lösungen. Dabei ist zu unterscheiden zwischen Lösungen, die Dienstanbieter als Datensammler zum datenschutzfreundlichen Umgang mit Big Data einsetzen, und Lösungen, die Nutzer einsetzen können, um sich vor der Hergabe zu vieler personenbezogener Daten zu schützen (Selbstdatenschutz).

## 3.1 Systemsicht

Zum datenschutzfreundlichen Umgang mit Big Data sollten Anbieter wie auch bei anderen Anwendungen, die sensitive Daten verarbeiten, diese sowohl verschlüsselt abspeichern als auch übertragen, um ein Ausspähen der Daten durch Dritte zu erschweren. Da Big-Data-Lösungen oft von mehreren Anwendern parallel genutzt werden, ist es wichtig, dass die Anwender gegenseitig abgeschottet sind. So können Anwender nicht gegenseitig ihre Daten in einem laufenden Verarbeitungsprozess einsehen.

## 3.2 Anonymisieren et al.

Auch zur Verarbeitung der Daten selbst gibt es datenschutzfreundliche Sicherheitsansätze, die unter dem Begriff Privacy-Preserving Data Mining zusammengefasst werden. Beim Anonymisieren werden identifizierende Merkmale aus den Datensätzen gelöscht. Dieser Vorgang soll nach BDSG § 3 Abs. 6 nicht oder nur mit unverhältnismäßig hohem Aufwand umkehrbar sein. Oft bestehen an diese Anonymität bestimmte Vorgaben, die beschreiben, wie groß eine Gruppe von Personen mindestens sein muss, die mittels der vorhandenen Daten eingegrenzt werden kann. Hier spricht man von $k$-Anonymität ($k$-Anonymity) [Swee02], wobei $k$ die Größe der nicht unterscheidbaren Personengruppe bestimmt. .

Beim Pseudonymisieren werden die Namen oder andere identifizierende Merkmale nicht einfach gelöscht, sondern durch ein Pseudonym ersetzt. Wer dieses Pseudonym kennt, kann den zur Person gehörenden Datensatz weiterhin identifizieren. Ein anderer Weg ist die Datenaggregation: Hier werden mehrere Datensätze zusammengefasst. So könnte für das Beispiel oben der durchschnittliche Verbrauch von Trinkwasser für das Gebäude gespeichert werden, statt diesen pro Familie auszuweisen.

Eine wichtige Beobachtung bei den technischen Maßnahmen zur Sicherstellung der Privatheit ist, dass der oben genannte unverhältnismäßige Aufwand, der nach dem BDSG als Grenze der Umkehrbarkeit von Anonymität gilt, durch Big Data relativiert werden könnte. Denn die Verfügbarkeit von Big-Data-Verfahren, die komplexe Zusammenhänge viel effizienter ableiten können, kann in der Praxis zu höheren notwendigen Hürden bei der Umkehrbarkeit führen. Ein Beispiel zeigt die kürzlich veröffentlichte Arbeit von de Montjoye et al. [Mont15], die das Aufdecken von anonymisierten Kreditkarteninformationen adressiert.

# 4 Ergebnisse Onlinebefragung

Die öffentliche Wahrnehmung zu Big Data wurde durch eine Online-Befragung erfasst. Hieran nahmen 202 Personen teil. Davon entfielen 29 Prozent auf weibliche und 66 Prozent auf männliche Teilnehmer. Die verbliebenen Teilnehmer machten hierzu keine Angabe.

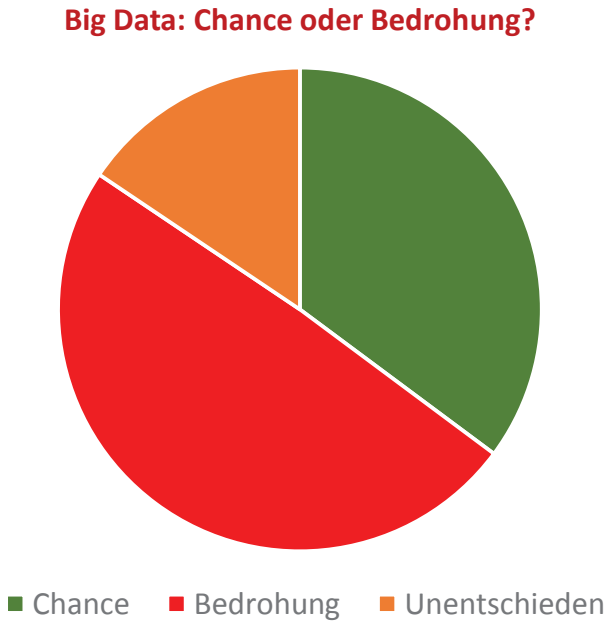


**Abb. 1:** Nehmen die Befragten Big Data als Chance oder Bedrohung wahr?

Die Altersverteilung ergab, dass der größte Anteil (42%) der Teilnehmer ein Alter von 31 bis 50 hatte. Im Alter von 18 bis 30 Jahren waren 23 Prozent der Teilnehmer und zwischen 51 und 65 Jahren waren es 27 Prozent. Immerhin 6 Prozent der Teilnehmer waren älter als 65 Jahre. Es nahm keine Person unter 18 Jahren teil. Berücksichtigt man die unterschiedliche Größe der Altersspannen, ergibt sich eine recht gleichmäßige Verteilung in dem Bereich von 18 bis 65 Jahren, wobei die Teilnehmer im Alter von 31 bis 50 Jahren pro Jahrgang am stärksten vertreten sind.

## 4.1 Chance oder Risiko?

Die zentrale Frage war, ob die Bürger Big Data eher als Chance oder als Bedrohung wahrnehmen. Beantwortet wurde diese Frage nicht (wie die meisten anderen Frage) durch Ankreuzen sondern mittels eines Schiebereglers. Eine Einstellung ganz links bedeutete „große Chance“, ganz rechts entsprechend „starke Bedrohung“. Die Antworten werden hier den Werten 0 (ganz links) bis 100 (ganz rechts) zugeordnet.

In Abbildung 1 is klar zu erkennen, dass mehr Teilnehmer eher eine Bedrohung in Big Data sehen als eine Chance. Zudem wurde auf der Seite der Bedrohung häufiger eine besonders ausgeprägte Bewertung vergeben als auf der Seite der Chance (siehe Abbildung 2). Einige Teilnehmer (13%) nahmen keine Einstellung an dem Schieberegler vor, so dass der voreingestellte Wert von 50 als Antwort übernommen wurde. Der Durchschnitt aller Antworten liegt bei 56, was einer leichten Tendenz zur Bedrohung entspricht.

## 4.2 Vertrauen in Branchen

Wem die Befragten hinsichtlich der Nutzung von Big Data vertrauen, zeigt Abbildung 3. Dabei ist insgesamt eine ausgeprägte Zurückhaltung zu sehen: Mehr als jeder Dritte vertraut keiner Branche. Forschung und Wissenschaft sind deutlicher Spitzenreiter, wobei hier über die Hälfte der Teilnehmer ihr Vertrauen bekunden. Knapp jeder Fünfte vertraut dem Staat und der

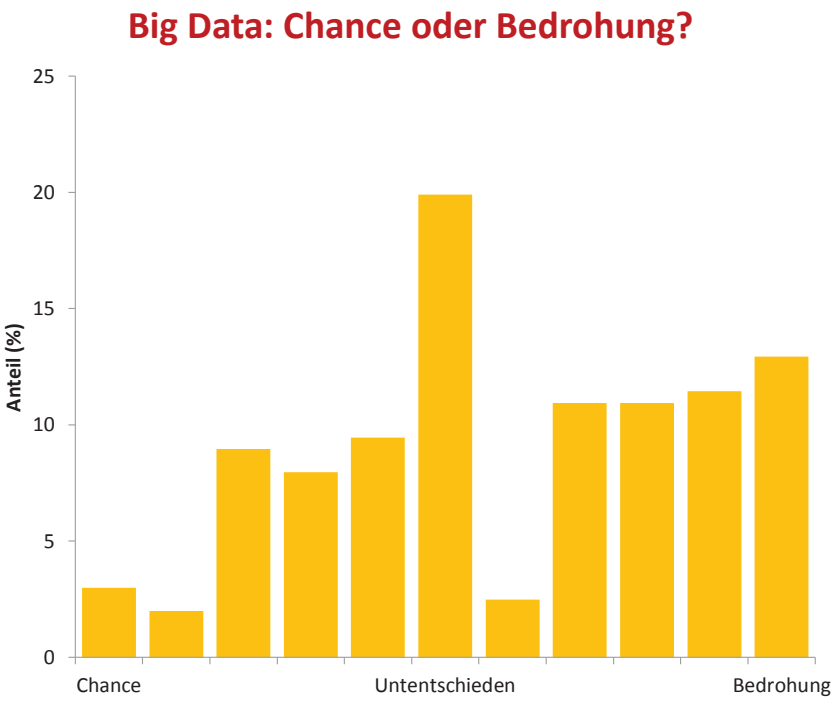


**Abb. 2:** Antworten auf die Frage nach 'Chance oder Bedrohung'

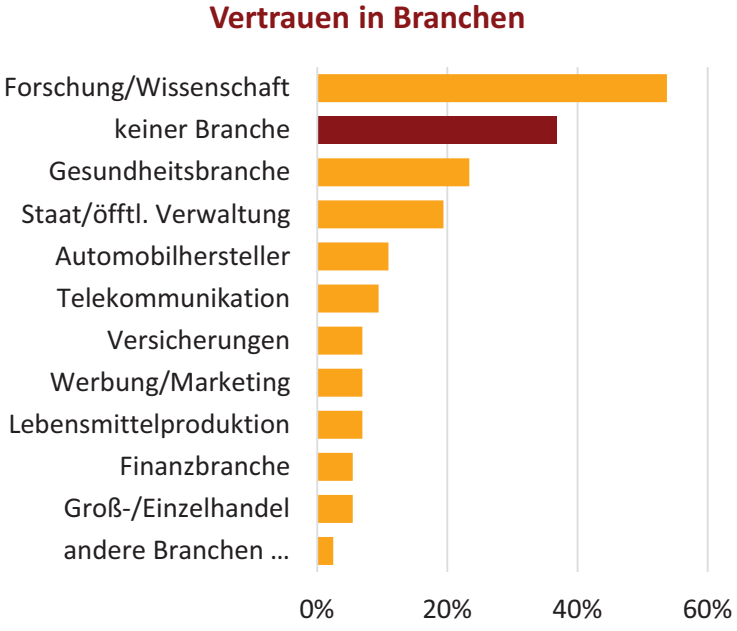


**Abb. 3:** Welchen Branchen vertrauen die Teilnehmer bezüglich des Einsatzes von Big Data?

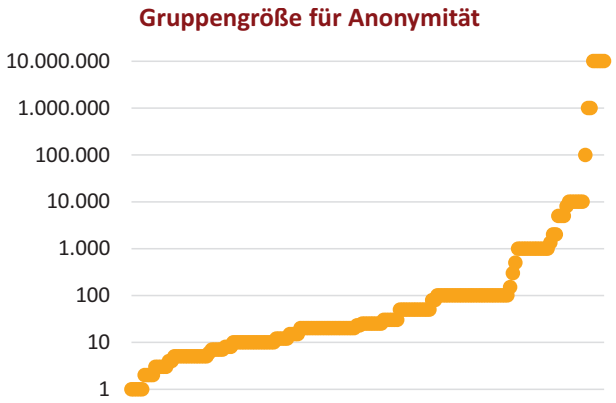


**Abb. 4:** Wie hoch soll k bei der k-Anonymität sein? (x/y = geforderte Größe/Nennung)

öffentlichen Verwaltung. Die Gesundheitsbranche hat unter den Industriezweigen den höchsten Vertrauenszuspruch von gut 20 Prozent. Den anderen aufgeführten Branchen wird beachtlich wenig Vertrauen ausgesprochen.

## 4.3 Anonymisieren

Eine Frage lautete wie folgt: 'In der deutschen Rechtsprechung gilt man bereits als anonym, wenn man innerhalb eines Kreises von mindestens 5 Personen nicht eindeutig identifizierbar ist, also bspw. innerhalb eines Fünf-Personen-Haushalts. Unter wie vielen Personen empfinden Sie sich als ausreichend anonym?' Übersetzen könnte man die Frage mit der Höhe von k bei Verfahren, die k-Anonymität herstellen. Die Antworten lassen sich am einfachsten mit dem Median zusammenfassen, da aufgrund der offenen Fragestellungen eine große Streuung mit Werten bis 10 Millionen genannt wurden. Der Median über alle Antworten beträgt 20. Einen Überblick bietet Abbildung 4.

Da unter anderem auch die Kompetenz im Umgang mit dem Internet abgefragt wurde, kann hier auch ein Bezug zwischen (empfundener) Kompetenz und Anforderung an die Anonymität hergestellt werden: Personen mit hoher und mittlerer Kompetenz wünschen sich im Median ein k von 25, die mit niedriger Kompetenz fordern nur ein k von 10.

## 4.4 Gesetzgebung

Eine häufige Kritik der Befragten liegt in der schwachen Position des Datenschutzbeauftragten. In mehreren Kommentaren zur Frage nach den Gesetzen zu Big Data wird gefordert, diesem mehr Befugnisse zu erteilen und Bußgelder beim Verstoß gegen das Datenschutzrecht deutlich zu erhöhen. Auf die Frage 'Wie sollte der Einsatz von Big-Data-Methoden in Deutschland gesetzlich reguliert werden?' (siehe auch Abbildung 5) fordert zwar die Mehrheit strengere Regeln. In der Diskussion wird allerdings schnell offensichtlich, dass schon eine strengere Umsetzung der bestehenden Regeln ausreichen würden. Auch die zahlreichen Kommentare zu dieser Frage gehen in diese Richtung.

Gesetze nutzen wenig, wenn sie nicht durchgesetzt werden. Diesen Aspekt betrachtet die Frage 'Für wie effektiv halten Sie gesetzliche Regelungen zum Datenschutz?', welche in Abbildung 6 zusammengefasst wiedergegeben wird. Über 70% der Befragten sind der Meinung, dass das Durchsetzen der Regeln derzeit scheitert, immerhin knapp 50% wünschen sich eine Änderung der Regeln, damit diese effektiver umgesetzt werden können.

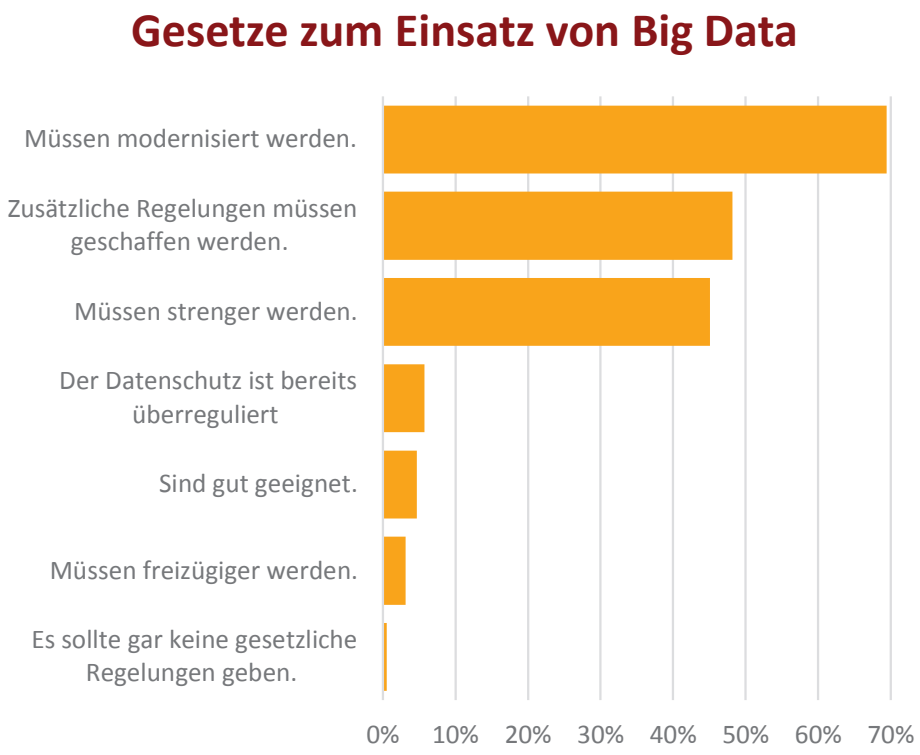


**Abb. 5:** Wie ist die Einstellung der Befragten zu den aktuellen Gesetzen im Datenschutz?


**Abb. 6:** Die Mehrheit der Befragten ist der Meinung, dass Datenschutz nicht effektiv durchgesetzt wird.

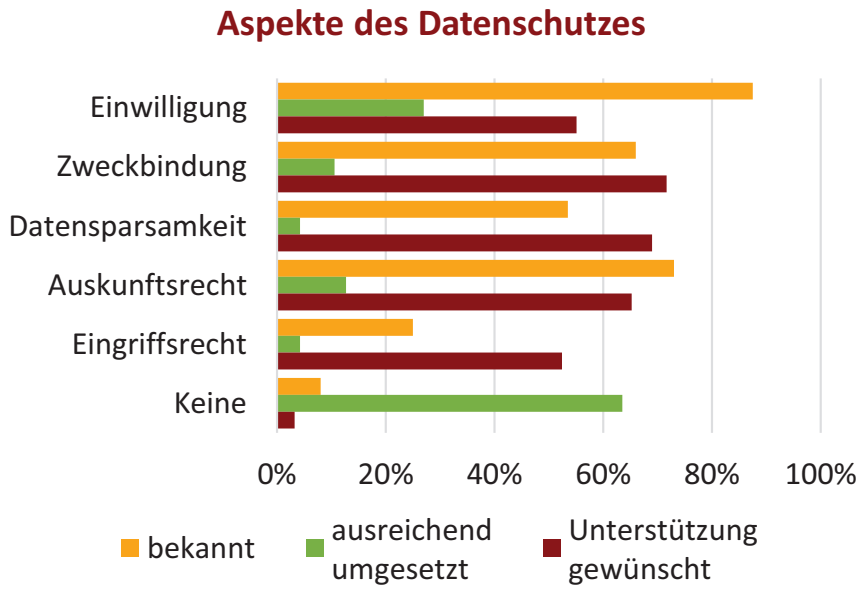


**Abb. 7:** Aspekte des Datenschutzes

Eine andere Sichtweise zeigt Abbildung 7. Hier werden die verschiedenen Aspekte des Datenschutzes aufgezählt und abgefragt, ob diese bekannt sind, ausreichend umgesetzt werden und ob bei ihrer Umsetzung Unterstützung erwünscht ist. Am bekanntesten ist die Einwilligung mit über 80% Bekanntheit. Dahingegen kennen nur gut 20% der Befragten das Eingriffsrecht. Auch hier kann eine Verbindung mit der Kompetenz der Befragten hergestellt werden, wie bereits in Kapitel 4.3: So wünschen sich im Schnitt weniger erfahrene Nutzer mehr Hilfe bei der Umsetzung wie erfahrenen Nutzer. Allerdings ist der Abstand von 4 genannten Aspekten zu 3 genannten Aspekten erstaunlich gering. Auch erfahrene Nutzer des Internets sind also durchaus offen für Hilfestellungen bei der Umsetzung des Datenschutzes.

Ebenfalls abgefragt wurde der Wunsch nach internationalen Regeln beim Schutz der Privatsphäre. Und bei aller Kritik am Datenschutz wünschen sich doch knapp 70% der Befragten, dass das hohe Datenschutzniveau nicht aufgegeben werden soll, wie in Abbildung 8 zu sehen ist. Eine Mehrheit ist ebenfalls für globale Datenschutzregelungen, während nur 2% internationale Abkommen für unnötig hält.

# 5 Diskussion

Um die Akzeptanz von Big Data in der Gesellschaft zu stärken, muss diskutiert werden, auf welcher Grundlage ein erfolgversprechender Kompromiss zwischen den durch Big Data entstehenden Chancen und Risiken erfolgen kann. Grundlage für die folgenden Punkten sind die Ergebnisse der Onlinebefragung, der Studie der Berichterstattung über Big Data und Gespräche mit Bürgern auf Bürgerdialogen.

Eine grundlegende Erkenntnis dabei ist, dass eine deutlich höhere Transparenz seitens der Betreiber zu einer höheren Akzeptanz führt: Big Data wird gerne als eine Quelle intransparenter Bewertungen gesehen, denen blind durch die Anwender gefolgt wird. Tatsächlich dient Big Data eher als Hilfe und Werkzeug bei der Entscheidungsfindung durch den Menschen. Dies deutlicher darzustellen erhöht bereits die Akzeptanz der Technologie deutlich.

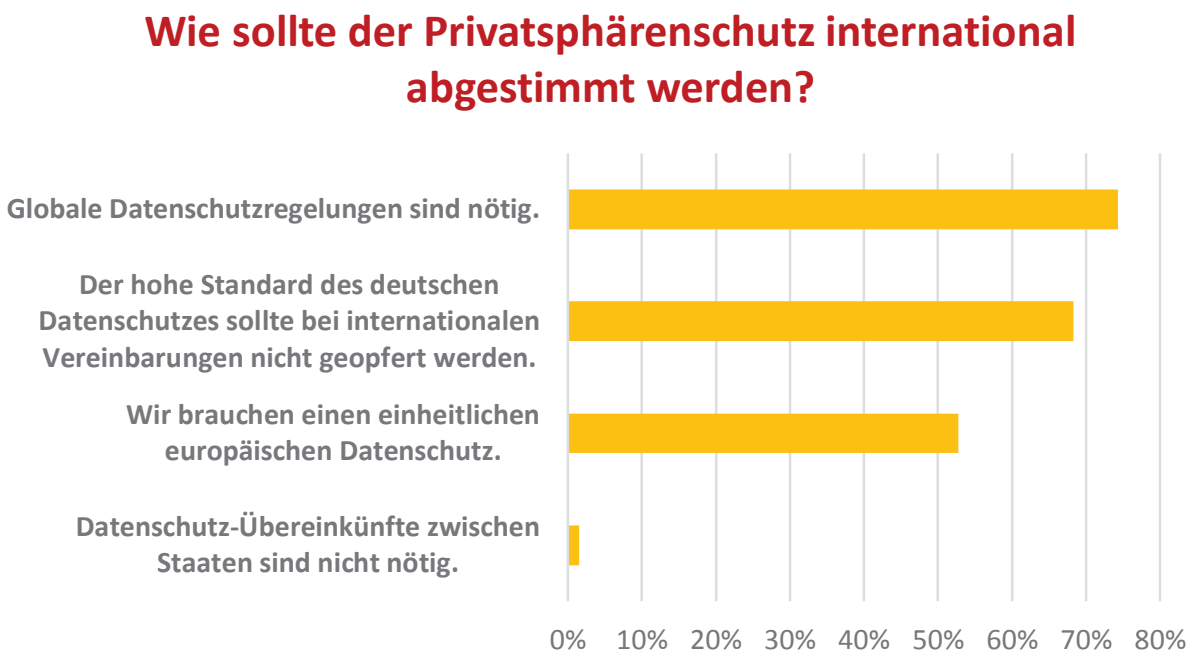


**Abb. 8:** Wie sehen die Wünsche nach einem internationalen Privatspährenschutz aus?

Interessant ist auch die Beobachtung, dass bei der Kritik an Big Data oft die Technologie als Synonym für Geschäftsmodelle gesehen wird. Bürger formulieren oft Skepsis gegen große Konzerne und deren Geschäftsgebaren, indem sie Big Data kritisieren. In einigen Fällen ist dabei aber Big Data nicht der ausschlaggebende Faktor. Insbesondere, wenn große Mengen strukturierter Daten erhoben werden, handelt es sich per Definition nicht um Big Data, wird aber häufig damit verwechselt. Soll Big Data als technologische Chance wahrgenommen werden, liegt eine Herausforderung dahin, zwischen Big Data, Nutzern von Big Data und allgemeinen Datenerhebungen zu unterscheiden. Sonst droht Big Data zu einem Synonym der Ängste von Bürgern bezüglich einem allgemeinen technologischem Wandel zu werden.

Eine wichtige Frage ist natürlich auch, ob ein Unternehmen Big Data mit personenbezogenen Daten betreiben und gleichzeitig datenschutzgerecht arbeiten kann. Soll diese technologische Chance genutzt werden, müssen die Aspekte des Datenschutzrechtes genau beachtet werden. Das bedeutet vor allem, dass durch die Zweckbindung eine genaue Planung der Nutzung der Daten im Vorhinein erfolgen muss. Einer explorativen Nutzung erhobener Daten sind so enge Grenzen gesetzt, wenn nicht vorher ein wirksames Anonymisieren erfolgt. So können die engen Vorgaben entkräftet werden und Daten werden frei nutzbar. Das Schaffen von Methoden, die effizient anonymisieren und trotzdem für Unternehmen relevante Erkenntnisse liefern, solle hier dementsprechend eine Priorität sein.

# 6 Zusammenfassung

Zum Thema Big Data herrschen vielen Vorurteile seitens der Bürger, die nicht zuletzt auf eine übersteigerte Darstellung der Möglichkeiten geschürt werden. Soll Big Data in der Zukunft zu einer akzeptierten Technologie werden, muss eine höhere Transparenz bezüglich der Technik, der erhobenen Daten sowie ihrer Nutzung erfolgen.

Es existieren umfangreiche rechtliche Rahmenbedingungen für den Einsatz von Big Data, einige von ihnen wie das Eingriffsrecht sind allerdings wenig bekannt. So entsteht eine Diskrepanz zwischen rechtlicher Situation und der Forderung der Öffentlichkeit nach einer besseren Regulierung.

### Danksagung

Diese Arbeit fasst das Dokument 'Big Data und Privatheit' (www.sit.fraunhofer.de/bigdata) zusammen, diskutiert es und fügt im Nachgang gewonnene neue Erkenntnisse hinzu. Möglich wurden beide Dokumente durch die Förderung des BMBF von EC SPRIDE, dem 'European Center for Security and Privacy by Design'. Mein Dank gilt meinen Co-Autoren des Originaldokuments Christian Winter, Oren Halvani, Marcel Schäfer und York Yannikos.

## Literatur

[BITK14] BITKOM: Potenziale und Einsatz von Big Data. Studienbericht, BITKOM (2014), http://www.bitkom.org/de/publikationen/38338_79283.aspx.

[Bund14] V. Bundesverband: EuGH: Google muss persoennliche Daten aus Suchergebnissen loeschen (2014), http://www.vzbv.de/meldung/eugh-google-muss-persoenliche-daten-aus-suchergebnissen-loeschen.

[Dreh15] S. Drehmel: Lassen wir Datenreichtum zu (2015), http://www.digitalewelt.org/content/lassen-wir-datenreichtum-zu.

[Lane01] D. Laney: 3D Data Management: Controlling Data Volume, Velocity and Variety. Research note, META Group (2001), http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

[Mont15] Y.-A. de Montjoye: Just four bits of credit card data can identify most anyone (Update) (2015), http://phys.org/news/2015-01-anonymous-credit-card-isnt.htm.

[Swee02] L. Sweeney: K-anonymity: A Model for Protecting Privacy. In: *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10, 5 (2002), 557–570, http://www.worldscientific.com/doi/10.1142/S0218488502001648.