International Working Group
on Data Protection
in Telecommunications

675.48.15

Arbeitspapier zu Big Data und Datenschutz
Bedrohung der Grundsätze des Datenschutzes in Zeiten von Big-Data-Analysen

55. Sitzung, 5.-6. Mai 2014, Skopje

- Übersetzung -

Einleitung[1]

1. Der Begriff „Big Data“ bezeichnet die enorme Zunahme von Zugriffen auf Informationen und deren automatisierte Nutzung[2]. Er bezieht sich auf die gigantischen Mengen digitaler Daten, über die Unternehmen, Behörden und andere große Organisationen verfügen und die sie mit Hilfe von Algorithmen umfassend analysieren[3].

2. Big Data gefährdet zentrale Datenschutzgrundsätze. Teilweise wird behauptet, dass eine Durchsetzung dieser Grundsätze in Zeiten von Big Data überhaupt nicht möglich sei[4]. Nach dieser Ansicht muss Datenschutz in erster Linie dadurch gewährleistet werden, dass Unternehmen eindeutig und umfassend über die Art und Weise des Umgangs mit personenbezogenen Daten informieren. Die Arbeitsgruppe ist jedoch der Meinung, dass der Schutz der Privatsphäre in Zeiten der Erfassung immer größerer Mengen personenbezogener Daten wichtiger denn je ist[5]. Die Datenschutzgrundsätze sind der Garant dafür, dass wir nicht einer umfas-

[1] Dieses Arbeitspapier enthält Hinweise zu gesetzlichen Bestimmungen, die möglicherweise nicht in allen in der Arbeitsgruppe repräsentierten Rechtssystemen enthalten sind.
[2] Vgl. White (2012): Big Data ist der Begriff für eine Datensammlung, die so groß und komplex ist, dass es schwierig wird, sie mit den vorhandenen Werkzeugen für die Verwaltung von Datenbanken oder traditionellen Datenverarbeitungsanwendungen zu verarbeiten.
[3] Vgl. Stellungnahme 03/2013 der Artikel 29-Datenschutzgruppe zur Zweckbindung, S. 35 [der engl. Fassung]
[4] Vgl. z. B.: Tene, Omer und Polonetsky, Jules (2012) Big Data for All: Privacy and User Control in the Age of Analytics, Northwestern Journal of Technology and Intellectual Property, im Erscheinen, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2149364; World Economic Forum (2013), Unlocking the Value of Personal Data: From Collection to Usage, http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf; Cate, Fred H. und Mayer-Schönberger, Viktor (2013), Tomorrow’s privacy. Notice and consent in a world of Big Data, International Data Privacy Law, 2013, Vol. 3, No. 2.
[5] Ähnliche Ansichten werden unter anderem von der Kommissarin der Federal Trade Commission Julie Brill (2013) geäußert: „Wir können das Potential von Big Data freilegen und seine Vorzüge genießen. Gleichzeitig können wir Datenschutzgrundsätze beachten, die den Konsumenten schützen.“; in: „Reclaim Your Name: Privacy in the Age of Big Data“, Sloan Cyber Security Lecture, Polytechnic Institute of NYU, October 23, 2013,

Secretariat
Berliner Beauftragter für
Datenschutz und Informationsfreiheit
An der Urania 4- 10
D-10787 Berlin
Phone +49 / 30 / 13889 0
Fax: +49 / 30 / 215 5050

E-Mail:
IWGDPT@datenschutz-berlin.de

Internet:
http://www.berlin-privacy-group.org

senden Profilbildung in einem ständig anwachsenden Gefüge neuer Zusammenhänge unterworfen werden. Eine Verwässerung zentraler Datenschutzgrundsätze in Verbindung mit einer immer umfangreicheren Nutzung von Big Data kann sich nachteilig auf den Schutz der Privatsphäre und auf andere wichtige gesellschaftliche Werte wie beispielsweise die Meinungsfreiheit und die Bedingungen für den Austausch von Ideen auswirken.

3. Die OECD und die Europäische Datenschutzrichtlinie haben in einigen Kernprinzipien festgelegt, wie personenbezogene Daten angemessen, korrekt und rechtmäßig verarbeitet werden dürfen[6]. Insbesondere die folgenden Grundsätze sind für Big Data von Relevanz: Zweckbeschränkung, Erforderlichkeit und Datenminimierung, Vollständigkeit und Qualität, Transparenz und das Recht auf Auskunft über personenbezogene Daten[7].

Geltungsbereich

4. Dieses Arbeitspapier stellt die mit Big Data einhergehenden Gefahren für den Datenschutz insbesondere auf dem Gebiet der Telekommunikation in den Mittelpunkt, damit diese von Datenschutzbehörden und anderen Interessengruppen berücksichtigt werden. Es richtet sich an Entscheidungsträger, Behörden, Wirtschaftsunternehmen und Bürger.

5. Big Data bringt ein breites Spektrum von Herausforderungen mit sich, von denen etliche, wie beispielsweise die Gefahr der Re-Identifizierung, bereits für sich allein genommen ein eigenes Thema umfangreicher Berichte darstellen könnten. Dieses Arbeitspapier befasst sich jedoch nicht im Detail mit einzelnen technischen Problemen, sondern mit den zentralen Gefahren für den Schutz der Privatsphäre.

Hintergrund

6. Daten sind allgegenwärtig. Weltweit nimmt die Datenmenge von Jahr zu Jahr um 50% zu. Allein in den letzten beiden Jahren wurden 90 % aller weltweit vorhandenen Daten erzeugt[8]; die meisten davon durch Verbraucher und deren Interaktion mit internetbasierten Diensten. Mit dem Aufkommen des „Internets der Dinge“[9] werden weitere Datenströme hinzukommen. Man schätzt, dass im Jahr 2015 über 50 Milliarden Sensoren existieren werden[10]. Diese werden Informationen darüber, wie Menschen mit den sie umgebenden Dingen interagieren, in Cloud-Computing-Dienste hochladen. Dies kann zu Veränderungen von Märkten und Geschäftsmodellen führen.

---

und von Ann Cavoukian, Alexander Dix und Khaled El Emam (2014), “The Unintended Consequences of Privacy Paternalism”, March 5, 2014, http://www.privacybydesign.ca/content/uploads/2014/03/pbd-privacy_paternalism.pdf.

[6] Vgl. die OECD Richtlinien über den Datenschutz und grenzüberschreitende Ströme personenbezogener Daten (2013) und Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr. Entsprechende Grundsätze sind auch in der Empfehlung CM/Rec(2010)13 des Europarates über den Schutz des Menschen bei der automatischen Verarbeitung personenbezogener Daten im Zusammenhang mit Profiling niedergelegt.

[7] Die entsprechenden Grundsätze in den OECD-Richtlinien sind: Der Grundsatz der Beschränkung der Datenerhebung, der Grundsatz der Zweckbestimmung, der Grundsatz der Datenqualität, der Grundsatz der Nutzungsbeschränkung und der Grundsatz der Beteiligung des Einzelnen.

[8] http://www-01.ibm.com/software/data/bigdata/

[9] „Internet der Dinge“ bezeichnet die Entwicklung einer steigenden Anzahl von Gegenständen und Personen, die mit Sensoren ausgestattet sind, die drahtlos miteinander in Netzwerken kommunizieren.

[10] The Internet of Things. How the Next Evolution of the Internet Is Changing Everything, Cisco White Paper, 2011, http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

7. Es steht außer Frage, dass die Fähigkeit zur Speicherung und Analyse enormer Datenmengen der Gesellschaft auf unterschiedlichste Weise nützen wird[11]. Big Data wird bereits heute in einigem Umfang zur Analyse von Daten mit dem Ziel der Bestimmung und Vorhersage von Trends und Korrelationen genutzt. Mit Hilfe von Big Data können beispielsweise die Ausbreitung von Epidemien vorhergesagt, schwere Nebenwirkungen von Medikamenten festgestellt und die Umweltbelastung in großen Städten bekämpft werden. Analysen dieser Art stellen per se keine Gefährdung der Privatsphäre dar, sofern die Daten hinreichend anonym sind (das Konzept der Anonymisierung wird in diesem Papier an anderer Stelle ausführlich behandelt). Darüber hinaus verwenden einige Big-Data-Analysen überhaupt keine personenbezogenen Daten, beispielsweise Wetterdatenanalysen oder Analysen der Sensordaten von Ölbohrinseln.

8. Big Data kann aber auch dergestalt eingesetzt werden, dass Einzelpersonen direkt betroffen sind. So gibt es Techniken zur Erstellung von Profilen und zur Vorhersage des Verhaltens von Personen und Personengruppen durch Zusammenstellung und Analyse von aus einer Vielzahl unterschiedlicher Quellen stammenden personenbezogenen Daten. Selbst wenn diese Informationen zusammengefasst und anonymisiert werden, kann das Ergebnis der Analyse immer noch Folgen für den Einzelnen haben.

9. „Personenbezogene Daten“ sind alle sich auf eine identifizierte bzw. identifizierbare Person beziehenden Informationen[12]. IP-Adressen, Mobiltelefonnummern, RFID-Tags und UDID-Nummern sind Beispiele für als personenbezogene Daten geltende eindeutige Kennungen[13]. Daten, die Informationen über Gewohnheiten und Interessen eindeutig identifizierter Personen geben, sind für Unternehmen und Regierungen von großem Interesse. Die Industrie entwickelt daher ständig neue, diesem Ziel dienende Techniken wie etwa das Device Fingerprinting. Dadurch wächst der Umfang von als personenbezogene Daten definierten eindeutigen Kennungen beständig.

10. Die „Wertschöpfungskette“ von Big Data umfasst mehrere Schritte, angefangen bei der Datenerhebung bis hin zu deren Speicherung und Verdichtung, ihrer Analyse sowie die Nutzung der Analyseergebnisse (siehe die Darstellung der Wertschöpfungskette am Ende des Dokuments). Auf diese einzelnen Schritte wird im Folgenden eingegangen.

11. Den ersten Schritt der Wertschöpfungskette bildet die *Datenerhebung.* Beispiele potenzieller Quellen personenbezogener Daten sind unter anderem Mobiltelefon-Apps, Smart-Grids, Straßenmaut-Transponder in Fahrzeugen, Patientenakten, Standortdaten, soziale Netzwerke, Flugzeugpassagierdaten, öffentliche Verzeichnisse, Kundenbindungsprogramme, Genomsequenzen, Einkaufshistorien etc. Aufgrund der zunehmenden Verbreitung der Sensortechnologie können Informationen von einer Vielzahl mobiler Geräte, darunter intelligente Zahnbürsten, Regenschirme, Kühlschränke, Schuhe, Fernsehgeräte etc. erhoben werden. Derartige Datenquellen können Informationen liefern, die potenziell sehr viel über die Lebensweise jedes Einzelnen preisgeben könnten.

12. Personenbezogene Daten können beispielsweise auf die folgende Weise *erhoben* werden:

---

[11] McKinsey Global Institute (2011),”Big Data: The next frontier for innovation, competition, and productivity” http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[12] So definiert in den OECD-Richtlinien über den Datenschutz und grenzüberschreitende Ströme personenbezogener Daten (2013), http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf, und in der Europäischen Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr.

[13] Stellungnahme 04/2007 der Artikel 29-Datenschutzgruppe zum Begriff „personenbezogene Daten“.

i. Personenbezogene Daten können vom Bürger selbst (beispielsweise durch Veröffentlichen persönlicher Angaben in sozialen Netzwerken) übermittelt werden.
ii. Personenbezogene Daten können als Voraussetzung für die Erbringung einer Dienstleistung erhoben werden.
iii. Personenbezogene Daten können aufgrund gesetzlicher Vorschriften erhoben werden.
iv. Personenbezogene Daten können in Verbindung mit der Inanspruchnahme spezifischer Dienstleistungen (beispielsweise Transaktions- und Standortdaten von Mautzahlstellen) *automatisiert* erhoben werden. Diese Datenerhebung kann auch *ohne Wissen des Betroffenen* erfolgen (beispielsweise bei der Erhebung von Daten aus Hot Spots an Flughäfen zur Verfolgung von Reisenden[14]).
v. Personenbezogene Daten können durch Verarbeitung und Analyse von für frühere und andere Zwecke erhobene Daten *abgeleitet* werden. Weiterhin können personenbezogene Daten aus verschiedenen, vermeintlich anonymen Datensätzen abgeleitet werden.
vi. Personenbezogene Daten (beispielsweise Kundendaten (Costumer Relationship Management)) können aus externen Quellen *hinzugefügt werden,* um (zuvor erhobene) Datenbestände zu erweitern.
vii. Personenbezogene Daten (beispielsweise (detaillierte) Kundendatensätze) können an externe Stellen *weitergegeben werden*, um (personenbezogene) Datenbestände von Partnerunternehmen anzureichern.

13. Im Zusammenhang mit Big Data sind über Internetnutzer gesammelte Informationen äußerst attraktiv, da sie detaillierte Informationen über deren Interessen, Netzwerke, Gewohnheiten und Verhaltensmuster enthalten können. Derartige Informationen können explizit (beispielsweise bei der Registrierung eines sozialen Profils im Internet) oder eher verdeckt durch den Einsatz verschiedener Tracking-Technologien erhoben werden[15].

14. Der zweite Schritt ist die *Verdichtung und Speicherung*[16] der Daten nach ihrer Erhebung. Einige Stellen verdichten und anonymisieren die Daten vor deren Speicherung; andere speichern Daten zusammen mit personenbezogenen Kennungen. Die enorme Steigerung der Speicher- und Analyseleistung bei immer niedrigeren Kosten bedeutet, dass Big Data nicht mehr einigen wenigen Branchenriesen vorbehalten ist. Big Data ist heute ein Werkzeug, das sowohl kleinen als auch großen Unternehmen aller Wirtschaftsbereiche zugänglich ist. Die Big Data-Technologie bedeutet eine Abkehr vom bisherigen Denken zur Datenspeicherung und -verarbeitung mithilfe von Großrechenanlagen. Dank neuer Technologien ist es möglich, neue und unstrukturierte Datenquellen zu verarbeiten und daraus Wert zu schöpfen.

15. Den dritten Schritt der Wertschöpfungskette stellen der Abgleich *und die Analyse* der erhobenen und gespeicherten Daten dar. Ein zentrales Element der Wertschöpfung dieser Stufe ist das Zusammenführen von Daten aus einer Vielzahl verschiedener Quellen zur Herstellung von Profilen sowie der Einsatz von Analysewerkzeugen zur Ableitung von ansonsten nicht verfügbaren Informationen. Die Nutzer von Big Data können entweder nur ihre eigenen internen Unternehmensdaten zusammentragen oder Daten von Dritten (oder aus öffentlich zugänglichen

[14] http://www.cbc.ca/news/politics/csec-used-airport-wi-fi-to-track-canadian-travellers-edward-snowden-documents-1.2517881

[15] Innerhalb der EU ist jetzt für den Einsatz bestimmter Cookies eine Einwilligung erforderlich, um das Sammeln von Daten für den Nutzer transparent zu machen und sicherzustellen, dass dieser mehr Kontrolle darüber erlangt.

[16] Aggregierung ist in diesem Zusammenhang zu verstehen als das Sich-Verschaffen von Erkenntnissen über eine Gruppe von Personen, nicht über Einzelpersonen. Aggregierung beinhaltet die Darstellung der Gesamtheit der Daten. Daten, die einer Einzelperson zugeordnet werden könnten oder diese identifizieren würden, werden nicht angezeigt. Abweichende Werte werden oft verborgen, indem sie als „unclear" dargestellt oder gelöscht werden. Ein Beispiel für Aggregierung ist das Verwenden von Durchschnittswerten.

Quellen) erwerben und diese mit eigenen Daten verbinden. Einige Beispiele von Analysetechniken in Verbindung mit Big Data sind Data Mining, maschinelles Lernen, soziale Netzwerkanalyse, prädiktive Analyse, „Sensemaking“, die Verarbeitung natürlicher Sprache und Visualisierung.

16. Der vierte Schritt der Wertschöpfungskette ist die *Nutzung* der Ergebnisse der Analyse. Big Data kann auf vielfältige Weise genutzt werden. Immer mehr Akteure, darunter beispielsweise Banken, Versicherungen, Ratingagenturen, Arbeitgeber sowie die Polizei sind im Interesse besserer und fundierterer Entscheidungen an einer Nutzung des durch die Analyse von Big Data erworbenen Wissens interessiert.

17. Eine Vielzahl von Interessengruppen ist an der gesamten Big-Data-Wertschöpfungskette beteiligt (siehe Abb.1 der Anlage). Einige Interessengruppen sind lediglich an ausgewählten Teilen der Wertschöpfungskette beteiligt. So nutzen beispielsweise Datenmakler personenbezogene Daten in der Regel nicht selbst, sondern verarbeiten und verkaufen sie lediglich weiter. Andere Interessengruppen können demgegenüber an sämtlichen Schritten der Wertschöpfungskette beteiligt sein. Ein Einzelhändler kann beispielsweise personenbezogene Daten mit Hilfe eines Kundenbindungsprogramms erheben, diese sodann speichern und verdichten und schließlich in seinem eigenen Geschäftsmodell verarbeiten und nutzen[17].

18. Personenbezogene Daten sind schon seit langer Zeit ein begehrtes Wirtschaftsgut und Anlass für die Entwicklung neuer, internetbasierter Dienstleistungen. Internetnutzer erhalten in der Regel Dienstleistungen kostenfrei, indem sie mit ihren personenbezogenen Daten dafür zahlen. Durch Big Data und die zunehmende Verbreitung des „Internets der Dinge“ wird der Markt für personenbezogene Daten an Volumen zunehmen und möglicherweise auch neue Wirtschaftsbereiche erschließen. So könnten beispielsweise intelligente Schuhe mit Sensoren gratis angeboten werden, wenn der Benutzer der Erfassung und Analyse der Daten seiner Laufgewohnheiten zustimmt. Ein Zahnarzt könnte seinen Patienten (vom Hersteller gratis zur Verfügung gestellte) intelligente Zahnbürsten kostenfrei überlassen, wenn die Patienten die von der Zahnbürste erhobenen Daten diversen interessierten Unternehmen zur Nutzung überlassen. Neue Unternehmen und Geschäftsmodelle werden entstehen, um den Mehrwert der gigantischen Mengen in einer ständig wachsenden Zahl von Situationen entstehenden personenbezogenen Daten abzuschöpfen.

Konsequenzen für den Schutz der Privatsphäre

19. Anhand der obigen Darstellung lassen sich die folgenden zentralen Herausforderungen für den Schutz der Privatsphäre durch die Nutzung von Big Data formulieren.

Datennutzung für neue Zwecke:

20. Big Data bedeutet zu weiten Teilen die Wiederverwendung von Daten. Dies stellt insoweit ein Problem für den Schutz personenbezogener Daten dar, als eine Nutzung erhobener Daten nicht für Zwecke zulässig ist, die *nicht* mit dem ursprünglichen Zweck der Erhebung *vereinbar* sind[18]. Das Potenzial von Big Data, durch die Aufbereitung immer größerer Datensätze wertvolles Wissen zu erschließen, stellt diesen Grundsatz der Zweckbindung infrage. Dieser Grundsatz besagt, dass ein Unternehmen, das erhobene personenbezogene Daten als Grundlage für vorausschauende Analysen nutzt, dafür Sorge zu tragen hat, dass die Analyse mit

[17] OECD (2013), “Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value”, OECD Digital Economy Papers, No. 220, OECD Publishing, http://dx.doi.org/10.1787/5k486qtxldmq-en

[18] Vgl. Artikel 6 Abs. 1 Bst. b der Richtlinie 95/46/EG.

dem ursprünglichen Zweck der Erhebung dieser Daten vereinbar ist. Eine Person, die Daten an Dritte weitergibt, stellt bestimmte natürliche Erwartungen an die Zwecke, für die diese Daten genutzt werden. Man überlässt einem Unternehmen oder dem Staat keine Informationen, wenn diese damit nach Belieben verfahren. Dies könnte eine erhebliche Herausforderung für die kommerzielle Nutzung von Big Data darstellen.

Datenmaximierung:

21. Big Data bedeutet Datenmaximierung. Big Data ist im Wesentlichen der absolute Gegenentwurf zu den Datenschutzgrundsätzen von Erforderlichkeit und Datenminimierung[19]. Diese Grundsätze sollen gewährleisten, dass nicht mehr personenbezogene Informationen erhoben und gespeichert werden, als für die Erreichung eindeutig definierter Zwecke erforderlich sind. Sobald die Daten nicht mehr für den ursprünglichen Zweck benötigt werden, sind sie zu löschen. Big Data bedeutet eine neue Betrachtungsweise von Daten, bei der diese einen Wert an sich erhalten. Der Wert von Daten wird mit ihren potenziellen *zukünftigen* Nutzungen verbunden. Diese Sicht auf Daten könnte den datenschutzrechtlichen Grundsatz infrage stellen, der besagt, dass die Verarbeitung von Daten für die zum Zeitpunkt ihrer Erfassung definierten und erklärten Zwecke angemessen, erforderlich und nicht übermäßig sein muss. Sie könnte auch den Wunsch und die Motivation der für die Datenverarbeitung verantwortlichen Stellen hinsichtlich des Löschens von Daten beeinflussen. Es ist denkbar, dass private Unternehmen und öffentliche Einrichtungen abgeneigt sind, Daten zu löschen, die sich irgendwann in der Zukunft als Quelle neuer Erkenntnisse und Einkünfte erweisen könnten. Die immer weiter verbreitete Nutzung von Big Data wird es den Datenschutzbehörden zunehmend erschweren, die Verpflichtung zur Löschung von Daten durchzusetzen.

Mangelnde Transparenz:

22. Das Recht auf Auskunft über die eigenen personenbezogenen Daten sowie über deren Verarbeitung ist ein wichtiger Grundsatz des Datenschutzes. Mangelnde Offenheit und fehlende Informationen hinsichtlich der Art der Erhebung und Nutzung von Daten können dazu führen, dass die Betroffenen Opfer von Entscheidungen werden, die für sie nicht nachvollziehbar sind und auf die sie keinen Einfluss haben. So weiß beispielsweise der durchschnittliche Internetnutzer nur sehr wenig darüber, wie der Online-Werbemarkt funktioniert und wie dort seine personenbezogenen Daten von einem breiten Spektrum kommerzieller Akteure gesammelt und genutzt werden[20]. Die meisten Bürger wissen nichts über etliche der auf diesem Markt tätigen Akteure, insbesondere über Datenmakler und Analysefirmen[21], wodurch die Wahrnehmung des Rechts des Einzelnen erschwert wird, Auskunft über seine Daten zu verlangen.

Aufdeckung sensitiver Informationen durch Kombination von Daten:

23. Ein bedenklicher Aspekt in Verbindung mit der Analyse von Big Data besteht darin, dass die Kombination erfasster Teilinformationen, die jeweils für sich genommen nicht notwendigerwei-

---

[19] Artikel 6 Abs. 1 Bst. c der Richtlinie 95/46/EG.

[20] Turow, Joseph (2011): "The Daily You. How the New Advertising Industry Is Defining Your Identity and Your Worth, Yale University Press", New Haven & London

[21] Acxiom ist einer der großen Datenmakler. Es handelt sich um ein US-Unternehmen, das für seine Klienten deren Kunden- und Firmendaten sammelt, analysiert und sie bei gezielten Werbekampagnen etc. unterstützt. Der Kundenstamm in den Vereinigten Staaten besteht vorwiegend aus Unternehmen aus den Bereichen Finanzwesen, Versicherungen, Direktmarketing, Medien, Vertrieb, Technologie, Gesundheit, Telekommunikation und Behörden. Das Unternehmen ist einer der weltweit größten Verarbeiter von Kundeninformationen. Es verfügt angeblich über 20 Milliarden Datensätze über Kunden und Informationen über 96 Prozent aller Haushalte in den Vereinigten Staaten.

se sensibel sind, zu einem sensitiven Ergebnis führen kann[22]. Mithilfe von Big-Data-Werkzeugen ist es möglich, Muster zu erkennen, die eine Vorhersage von Präferenzen Betroffener ermöglichen, beispielsweise in Bezug auf Gesundheit, politische Ansichten oder sexuelle Orientierung. Dies sind besonders schützenswerte Informationen. Die verarbeitenden Stellen müssen sich dieses Risikos bei der Kombination und Analyse von Daten bewusst sein[23].

Risiko der Re-Identifizierung:

24. Eines der größten Risiken in Verbindung mit der Analyse von Big Data ist das der Re-Identifizierung. Durch die Zusammenstellung von Daten aus verschiedenen Quellen besteht das Risiko, dass eine Person anhand von Datensätzen identifiziert werden kann, die auf den ersten Blick anonym zu sein scheinen. Dies beeinträchtigt die Effektivität der Anonymisierung als Methode zur Verhinderung von Problemen für die Privatsphäre in Verbindung mit Profilbildung und mit anderen Datenanalysen[2425]. Das Risiko der Re-Identifizierung lässt sich dadurch verringern, dass gewährleistet wird, dass zur Analyse ausschließlich anonymisierte Daten verwendet werden. Allerdings kann nicht immer ohne Weiteres festgestellt werden, ob ein Datensatz ausreichend und belastbar anonymisiert ist. Dies kann aus zwei Gründen schwierig sein:

- Erstens ist der Begriff „identifizieren“ – und damit auch „anonymisieren“ – komplex, weil eine Person auf viele verschiedene Arten identifiziert werden kann[26]. Hierzu gehören die direkte Identifizierung, bei der die Person anhand einer einzigen Datenquelle (beispielsweise einer Liste mit ihrem vollständigen Namen) eindeutig identifiziert wird, sowie die indirekte Identifizierung, bei der zwei oder mehr Datenquellen kombiniert werden müssen, um eine Identifizierung zu ermöglichen.

- Zweitens kann ein Unternehmen, das einen vermeintlich anonymisierten Datensatz verwendet, nicht mit letzter Sicherheit sagen, ob nicht noch weitere Datensätze existieren, aufgrund derer es einem Dritten möglich wird, Einzelpersonen in dem anonymisierten Datenbestand zu identifizieren. Selbst nach Löschung der identifizierenden Informationen ist es unter Umständen immer noch möglich, spezifische Informationen anhand von Verbindungen innerhalb verschiedener Sammlungen von Big Data einzelnen Personen zuzuordnen. Ein reales Beispiel hierfür enthält der Aufsatz „How to break anonymity of the Netflix Prize Dataset“[27].

---

[22] http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all

[23] Ein häufig zur Verdeutlichung dieses Problems herangezogenes Beispiel ist der sogenannte “Schwangerschafts-Algorithmus” der US-Kette Target. Target hat einen Algorithmus entwickelt, der auf der Grundlage der von ihnen gekauften Produkte bestimmen konnte, welche Kundinnen schwanger waren. Target sendete dann Gutscheine für „Schwangerschaftsprodukte“ an diese Kundinnen. In einem Fall führte das Versenden dieser Gutscheine dazu, dass ein Vater auf Schwangerschaft seiner Tochter aufmerksam wurde, bevor diese die Möglichkeit hatte, ihn darüber zu unterrichten.

[24] Anonymisierung entsteht durch eine die Identifizierung irreversibel verhindernde Verarbeitung personenbezogener Daten, vgl. Richtlinie 95/46/EG. Anonymisierung wird zudem in internationalen Regelungsstandards wie ISO 29100 definiert als “Prozess, bei dem Informationen, die einer Person zugeordnet werden können, so modifiziert werden, dass diese Informationen weder direkt noch indirekt von dem Halter der Information allein oder im Zusammenwirken mit einer anderen Stelle eine Person identifizieren können.“ (ISO 29100:2011).

[25] In dem Arbeitspapier „Webtracking und Privatsphäre: Die Beachtung von Kontext, Transparenz und Kontrolle bleibt unverzichtbar“ (15./16. April 2013, Prag (Tschechische Republik)), werden die Herausforderungen im Zusammenhang mit der Re-Identifizierung als „Game-Changer“ bezeichnet.

[26] Stellungnahme 05/2014 der Artikel 29-Datenschutzgruppe zu Anonymisierungstechniken

[27] http://arxiv.org/abs/cs/0610105v1

Konsequenzen für die Informationssicherheit

25. Big Data bringt auch Probleme für die Informationssicherheit und damit möglicherweise auch für den Schutz der Privatsphäre mit sich. Beispiele für derartige Sicherheitsprobleme sind die Nutzung mehrerer Infrastrukturebenen zur Verarbeitung von Big Data, neue Infrastrukturtypen zur Bewältigung des enormen Datenstroms sowie die nicht skalierbare Verschlüsselung großer Datensätze. Darüber hinaus kann eine Verletzung des Schutzes personenbezogener Daten schwerwiegende Folgen haben, wenn sehr große Datenbestände gespeichert sind. Ein Unternehmen, das eine große Menge personenbezogener Daten erwirbt und speichert, muss ein verantwortungsvoller Verwalter dieser Informationen sein.

Unrichtige Daten:

26. Ein wichtiger Grundsatz des Datenschutzes besagt, dass Personen betreffende Entscheidungen auf zutreffenden Informationen basieren müssen. Die Anwendung leistungsfähiger Data-Mining-Techniken ist beispielsweise in der Versicherungswirtschaft und bei Ratingagenturen zunehmend beliebt. Big Data erleichtert die Nutzung eines weitaus größeren Spektrums sowie neuer Arten von Datenquellen bei der Erstellung von Bonitätsbeurteilungen und Risikoprofilen. Neue, auf die Nutzung von Big Data spezialisierte Ratingagenturen sind auf den Markt getreten. Diese Agenturen erstellen Profile von Personen auf der Grundlage von ausschließlich aus Onlinequellen bezogenen Informationen.

27. Entscheidungen aufgrund von Informationen zu treffen, die beispielsweise aus sozialen Medien gewonnen und zusammengestellt wurden, beinhaltet jedoch die Gefahr, dass diesen Entscheidungen ungenaue Informationen zugrunde liegen. Auf derartigen Informationen basierende Entscheidungen sind weniger transparent und nachprüfbar als auf der Grundlage von Informationen aus offiziellen Registern getroffene Entscheidungen. Eine Schwäche von Big-Data-Analysen liegt darin, dass der Kontext oftmals unberücksichtigt bleibt[28]. Selbst bei richtigen Daten können sich Probleme für die Privatsphäre dadurch ergeben, dass die Daten außerhalb ihres ursprünglichen Zusammenhangs verwendet werden. Eine Entscheidungsfindung aufgrund von für andere Zwecke gesammelten und in anderen Zusammenhängen erzeugten Informationen kann zu Ergebnissen führen, die der tatsächlichen Situation nicht gerecht werden. Es ist wichtig zu betonen, dass die Nutzung von für andere Zwecke bestimmte Informationen unter Datenschutzgesichtspunkten per se rechtswidrig ist, es sei denn, diese anderen Zwecke sind mit den ursprünglichen Zwecken vereinbar oder die Daten sind anonymisiert.

28. Transparenz, beispielsweise in Form des Rechts des Betroffenen auf Auskunft zu den über ihn verarbeiteten Informationen, ist eine Voraussetzung dafür, dass der Betroffene in der Lage ist, seine eigenen Interessen wahrzunehmen. Es ist ein zentraler Grundsatz des Datenschutzes, dass Betroffene die Berichtigung oder Löschung von nachweislich unrichtigen Informationen, Beurteilungen und Behauptungen verlangen können.

Ungleichgewicht der Kräfte:

29. Der Einzelne hat grundsätzlich kaum Einflussmöglichkeiten auf das Verhalten von Großunternehmen. Die extensive Nutzung von Big-Data-Analysen kann das Ungleichgewicht zwischen Großunternehmen und Verbrauchern noch weiter verstärken[29]. Es sind doch die Unternehmen, die personenbezogene Daten sammeln und in den Genuss des der Analyse und Verarbeitung

[28] danah boyd und Kate Crawford sind zwei Forscherinnen, die die Wichtigkeit der Einbeziehung des Kontextes in Big Data-Analysen hervorgehoben haben.
boyd, danah u. Crawford, Kate (2012), "Critical Questions for Big Data", Information, Communication & Society 15:5, 662-679, http://dx.doi.org/10.1080/1369118X.2012.678878

[29] Vgl. Stellungnahme 03/2013 der Artikel 29-Datenschutzgruppe zur Zweckbindung

dieser Informationen innewohnenden, ständig wachsenden Werts kommen, und nicht der Einzelne, von dem diese Informationen stammen. Die Datenverarbeitung kann sich vielmehr sogar zum Nachteil des Verbrauchers auswirken, indem sie ihn dem Risiko zukünftiger potenzieller Benachteiligungen (beispielsweise im Hinblick auf Beschäftigungschancen, Bankkredite oder Wahlmöglichkeiten bei Krankenversicherungen) aussetzt[30].

Datendeterminismus und Diskriminierung:

30. Die „Big-Data-Haltung" basiert auf der Annahme, dass man, je mehr Daten man sammelt und auf je mehr Daten man Zugriff hat, desto bessere, fundiertere und genauere Entscheidungen treffen kann. Mehr Daten zu sammeln, bedeutet jedoch nicht notwendigerweise mehr Wissen. Mehr Daten können auch zu mehr Verwirrung und zu mehr „falsch positiven" Ergebnissen führen[31]. Ein übermäßiger Gebrauch von automatisierten Entscheidungen und Prädikatorenanalyse kann nachteilige Folgen für Betroffene haben. Algorithmen sind nicht neutral, sondern Ausdruck von Entscheidungen unter anderem in Bezug auf Daten, Verknüpfungen, Schlussfolgerungen, ihre Interpretation und Schwellenwerte für ihre Berücksichtigung, die einem spezifischen Zweck förderlich sind[32]. Big Data kann somit bestehende Vorurteile und Stereotypen bestätigen sowie soziale Ausgrenzung und Isolierung verstärken. Korrelationsanalysen können darüber hinaus im Einzelfall zu vollkommen falschen Ergebnissen führen. Korrelation wird oftmals mit Kausalität verwechselt. Wenn Analysen ergeben, dass Personen, die X mögen, mit einer Wahrscheinlichkeit von 80 % Y ausgesetzt werden, kann daraus unmöglich geschlossen werden, dass dies in 100 % der Fälle eintritt. Diskriminierungen auf der Grundlage statistischer Analysen kann daher zu einer Frage des Schutzes der Privatsphäre werden. Eine Entwicklung, bei der immer mehr gesellschaftliche Entscheidungen auf Algorithmen basieren, kann zu einer „Diktatur der Daten"[33] führen, in der wir nicht mehr anhand tatsächlicher Handlungen, sondern anhand dessen, was nach Datenlage die wahrscheinlichen Handlungen sein werden, beurteilt werden.

Der Einschüchterungseffekt („chilling effect"):

31. Wenn eine Entwicklung einsetzt, durch die Bonitätsbewertungen und Versicherungsprämien ausschließlich oder vorwiegend auf den Informationen basieren, die Nutzer in diversen Zusammenhängen im Internet und anderen Bereichen des Alltags hinterlassen, kann dies Folgen für den Schutz der Privatsphäre haben und für die Art und Weise, wie wir uns verhalten. So ist es möglich, dass unsere Kinder in zehn Jahren keinen Versicherungsschutz mehr bekommen, nur weil ihre Eltern beispielsweise in einem sozialen Netzwerk gepostet haben, eine Veranla-

---

[30] Die OECD hat diesem Thema ihre Aufmerksamkeit gewidmet und einen Bericht veröffentlicht, in dem auf die Methoden zur Schätzung des finanziellen Wertes personenbezogener Daten eingegangen wird. Nach diesem Bericht könnten die Methoden zur Bestimmung des Wertes personenbezogener Daten helfen, Transparenz zu gewährleiten und einen Einblick in den Markt für den Handel mit Daten zu erlangen. Zudem wird in dem Bericht argumentiert, dass ein gesteigertes Bewusstsein der Konsumenten über den Wert ihrer personenbezogenen Daten helfen könnte, das ökonomische Ungleichgewicht zwischen den Unternehmen und den Konsumenten auszugleichen. Dies könnte dem Konsumenten auch helfen, höhere Ansprüche und Erwartungen an den Umgang mit seinen personenbezogenen Daten zu stellen. OECD (2013), "Exploring the Economics of Personal Data: A survey of methodologies for measuring monetary value", OECD Digital Economy Papers, No. 220, OECD Publishing. http://dx.doi.org/10.1787/5k486qtxldmq-en

[31] Google „Grippe-Trends" war kürzlich Gegenstand einiger genauerer Untersuchungen; http://bits.blogs.nytimes.com/2014/03/28/google-flu-trends-the-limits-of-big-data/

[32] Dwork, Cynthia and Mulligan, Deirdre K. (2013), "It's not privacy, and it's not fair", 66 Stanford Law Review, Online 35, September 3, 2013, http://www.stanfordlawreview.org/online/privacy-and-big-data/its-not-privacy-and-its-not-fair

[33] Mayer-Schönberger, Viktor u. Cukier, Kenneth (2013), "Big Data. A Revolution That Will Transform How We Live, Work and Think", John Murray, London

gung für eine Erbkrankheit zu haben. Dies kann zu einer Selbstbeschränkung der Teilhabe am gesellschaftlichen Leben im Allgemeinen führen oder zu einer aktiven Anpassung des eigenen Verhaltens – sowohl online als auch im echten Leben. Wir könnte befürchten, dass sich die Spuren, die wir in den unterschiedlichsten Zusammenhängen hinterlassen, auf zukünftige Entscheidungen auswirken, wie beispielsweise auf Chancen auf dem Arbeitsmarkt, die Möglichkeit zum Erhalt von Krediten oder zum Abschluss von Versicherungen. Dies kann sogar so weit gehen, dass Nutzer davor zurückschrecken, sich online nach alternativen Ansichten umzusehen, aus Angst davor, identifiziert oder entdeckt zu werden oder um die Erstellung eines Nutzerprofils zu verhindern. Hinsichtlich der Nutzung von Big Data durch Behörden kann die Unsicherheit darüber, aus welchen Datenquellen Informationen erhoben und wie diese genutzt werden, das Vertrauen der Bürger in die Behörden selbst infrage stellen. Dies kann sich wiederum nachteilig auf die eigentlichen Grundlagen einer offenen und gesunden Demokratie auswirken. Ein unzureichender Schutz der Privatsphäre kann zu einer Schwächung der Demokratie führen, wenn Bürger ihre Beteiligung am offenen Meinungsaustausch einschränken. Im ungünstigsten Fall kann die übermäßige Nutzung von Big Data einen Einschüchterungseffekt in Bezug auf die Inanspruchnahme der Meinungsfreiheit haben, wenn die Voraussetzungen für die Nutzung von Big Data nicht offengelegt werden und nicht unabhängig überprüft werden können[34].

Echokammern:

32. Die Personalisierung im Internet durch individuell angepassten Medien und auf dem Internetverhalten des Einzelnen basierenden Nachrichtendiensten wird sich darüber hinaus auch auf die Rahmenbedingungen für öffentliche Debatten und den öffentlichen Gedankenaustausch auswirken und – wichtige Voraussetzungen für eine gesunde Demokratie. Dies ist nicht in erster Linie ein Problem des Datenschutzes, sondern stellt eine Gefahr für die Gesellschaft an sich dar. Die Gefahr aufgrund so genannter „Echokammern“ oder „Filterblasen“ liegt darin, dass der Nutzer nur solchen Inhalten ausgesetzt wird, die seinen ohnehin schon bestehenden Haltungen und Werten entsprechen. Der Austausch von Gedanken und Standpunkten kann gehemmt werden, wenn der Bürger seltener mit Ansichten konfrontiert wird, die von seinen bestehenden Meinungen abweichen.

Empfehlungen

33. Auch wenn Big Data in mehrfacher Hinsicht Gefahren für den Datenschutz mit sich bringt, ist dennoch eine Nutzung dieser Art von Analyse möglich, ohne gegen grundlegende Datenschutzprinzipien zu verstoßen. Die Arbeitsgruppe gibt folgende Empfehlungen für eine Nutzung von Big Data bei gleichzeitiger Wahrung der Privatsphäre jedes Einzelnen.

Einwilligung:

[34] Die norwegische Datenschutzbehörde hat 2013 unter Norwegern eine Umfrage zu datenschutzbezogenen Themen durchgeführt. Eines der untersuchten Themen war die Frage nach dem Bestehen einer generellen Tendenz hin zu einem „chilling effect“ in Norwegen. In der Umfrage wurden die Personen gefragt, ob sie sich entschlossen hätten, etwas nicht zu tun, weil sie sich nicht sicher waren, wie die Informationen darüber in der Zukunft verwendet würden. Die Ergebnisse deuten auf eine generelle Tendenz zu einem „chilling effect“ hin. Sie zeigen, dass ein signifikanter Teil der Bevölkerung bestimmte Handlungen vermieden hat, weil sie über die mögliche zukünftige Verwendung der Informationen darüber unsicher sind. Es ist erwähnenswert, dass nicht weniger als 26 Prozent sich entschieden, eine Petition nicht zu unterzeichnen und 16 Prozent bestimmte Internetsuchen unterließen. Norwegian Data Protection Authority (2014), “The Chilling Effect in Norway”, January, 2014. http://www.datatilsynet.no/Global/04_planer_rapporter/Nedkj%C3%B8ling%20i%20norge_eng_.pdf

34. Es wurde argumentiert, dass das Einwilligungserfordernis als rechtliche Grundlage für die Verarbeitung personenbezogener Informationen im Zeitalter von Big Data nur bedingt geeignet ist[35]. So wird gelegentlich vertreten, dass die ständige Forderung nach Einwilligung im Internet paradoxerweise sogar zu einem schlechteren Schutz des Einzelnen führen kann. Es lässt sich bereits jetzt eine Tendenz feststellen, dass Unternehmen ihre Kunden zur Abgabe sehr umfassender Einwilligungserklärungen auffordern, möglicherweise in der Hoffnung, dass derartige Erklärungen oftmals nicht gründlich gelesen werden und den Unternehmen damit „Ellbogenfreiheit" für die Nutzung der Informationen für zukünftige und andere Zwecke verschaffen. Eine solche Nutzung der Einwilligung ist rechtswidrig.

35. Auch wenn ohne Frage Schwierigkeiten bestehen, eine wirksame und aussagekräftige Einwilligung einzuholen, bleibt die Einwilligung gleichwohl der Eckpfeiler moderner Datenschutzgesetze. Eine Aufweichung der Einwilligung droht die Kontrolle des Einzelnen über die Nutzung seiner Daten ebenfalls zu schwächen[36]. Die Einwilligung ist nur eine von mehreren gesetzlichen Grundlagen für die Verarbeitung personenbezogener Daten. Auch wenn der Einwilligung eine wichtige Rolle zukommt, schließt dies je nach Kontext für die Verarbeitung jedoch nicht die Möglichkeit anderer rechtlicher Grundlagen aus, die unter Umständen aus Sicht sowohl der verantwortlichen Stelle als auch des Betroffenen geeigneter sind[37].

36. Im Zusammenhang mit der Nutzung personenbezogener Daten für Analyse- und Profilerstellungszwecke sollte eine wirksame Einwilligung des Betroffenen eingeholt werden[38].

37. Falls eine Einwilligung nicht eingeholt werden kann, könnte eine Verarbeitung der Daten innerhalb sorgsam abgewogener Grenzen trotzdem möglich sein[39]. So könnte die verantwortliche Stelle die Daten beispielsweise dann verarbeiten, wenn dies für ihre legitimen Zwecke erforderlich ist, so lange die Interessen des Betroffenen nicht überwiegen. Die verantwortliche Stelle muss zwei entgegenstehende Interessen gegeneinander abwägen, nämlich einerseits ihre legitimen Interessen und andererseits die Interessen des Einzelnen. Das Ergebnis dieser Interessenabwägung wird von Fall zu Fall verschieden sein und hängt von den jeweils in Frage stehenden Datenschutzinteressen des Einzelnen und von den legitimen Interessen der verantwortlichen Stelle ab[40]. Je stärker die Auswirkungen auf den Betroffenen, desto wichtiger ist die Beachtung entsprechender Schutzmechanismen[41].

---

[35] Vgl. Cate, Fred H. u. Mayer-Schönberger, Viktor (2013), "Data Protection Principles for the 21st Century: Revising the 1980 OECD Guidelines", December 2013, http://op.bna.com/pl.nsf/id/dapn-9gyjvw/$File/Data-Protection-Principles-for-the-21st-Century.pdf

[36] „Beim Entfernen der Einwilligung aus der Betrachtung riskiert man, fundamentale Grundrechte, Schutzfunktionen und Freiheiten jenseits des "Information-und-Einwilligungs"-Systems zu unterminieren. Anstatt die Einwilligung abzuschaffen, sollten wir daran arbeiten, die Transparenz und die individuellen Kontrollmechanismen zu verbessern, und die Herausforderungen direkt angehen" (Cavoukian et. Al (2014) „The Unintended Consequences of Privacy Paternalism").

[37] Stellungnahme 15/2011 der Artikel 29-Datenschutzgruppe zur Definition der Einwilligung.

[38] Eine zulässige Einwilligung muss ohne Zwang, für den konkreten Einzelfall und in Kenntnis der Sachlage erfolgen, vgl. Art. 2 Bst. h der Richtlinie 95/46/EG.

[39] Die Art. 29-Datenschutzgruppe hat Handlungsempfehlungen zu einer solchen Gratwanderung gegeben (Vgl. Arbeitspapier 217, S. 55-56 [der englischen Fassung]). Zudem betont der Bericht des Weißen Hauses zu Big Data die Wichtigkeit des Sachzusammenhangs in Situationen, in denen eine Einwilligung nicht praktikabel ist. (Executive Office of the President, White House (2014): "Big Data: Seizing opportunities, preserving values").

[40] Obwohl sich nicht ausschließen lässt, dass eine Interessenabwägung in einigen Fällen die Verarbeitung personenbezogener Daten rechtfertigt, muss stets der Einzelfall berücksichtigt werden. Diese Möglichkeit kann deshalb kaum als allgemeine Rechtsgrundlage für die Erhebung und Analyse von Big Data dienen. In jedem Fall hat die verarbeitende Stelle die Beweislast, dass die Bedingun-

38. Verantwortliche Stellen, die die erhobenen Daten für einen anderen als den ursprünglichen Zweck verwenden wollen, sind verpflichtet, die Vereinbarkeit zwischen den ursprünglichen und den neuen Zwecken einzelfallabhängig zu prüfen[42]. Solange diese Vereinbarkeit nicht gewährleistet ist, dürfen keine Daten verarbeitet werden, die eine persönliche Identifizierung ermöglichen.

Verfahren für eine wirksame Anonymisierung:

39. Die verantwortliche Stelle muss entscheiden, ob die in der Big-Data-Analyse zu verwendenden personenbezogenen Daten anonymisiert oder pseudonymisiert werden müssen oder ob sie identifizierbar bleiben können. Diese Entscheidung bestimmt, wie sich die Rechtsvorschriften zum Datenschutz auf die weitere Verarbeitung der Informationen durch das Unternehmen auswirken. Anonymisierte Daten sind vom Anwendungsbereich der Datenschutzrechtvorschriften nicht erfasst.

40. Anonymisierung kann bei der Reduzierung oder Beseitigung von Datenschutzrisiken im Zusammenhang mit Big-Data-Analysen hilfreich sein, jedoch nur, wenn die Anonymisierung technisch einwandfrei durchgeführt wird[43].

    Anonymisierung ist das Ergebnis einer Verarbeitung personenbezogener Daten mit dem Ziel, die Identifizierung des Einzelnen unumkehrbar zu verhindern. Dabei hat die verantwortliche Stelle verschiedene Aspekte zu bedenken, insbesondere sämtliche (entweder durch die verantwortliche Stelle selbst oder durch Dritte) „mit einiger Wahrscheinlichkeit“ zu einer Identifizierung eingesetzten Mittel. Wichtig ist eine Prüfung anonymisierter Daten hinsichtlich eines ak-

---

gen für eine Interessenabwägung zu ihren Gunsten erfüllt sind. Diese Abwägung und ihre zahlreichen Ermessensbestandteile und Unklarheiten können in dieser Hinsicht Schwierigkeiten für die verarbeitende Stelle mit sich bringen.

[41] Die Art. 29-Datenschutzgruppe hat in ihrer Stellungnahme 06/2014 zu „berechtigten Interessen verarbeitender Stellen im Hinblick auf Art. 7 der Richtlinie 95/46/EG“ Handlungsanweisungen zu Faktoren gegeben, die bei einer Interessenabwägung zu berücksichtigen sind. Besondere Aufmerksamkeit wird dabei der Rolle gewidmet, die Schutzmaßnahmen bei der Reduzierung unzulässiger Auswirkungen für von der Datenverarbeitung Betroffene spielen könnten und dadurch die Gewichtung von Rechten und Interessen so verändern würden, dass die berechtigten Interessen der verarbeitenden Stelle zum Zuge kommen. Solche Schutzmaßnahmen könnten unter anderem Begrenzungen bezüglich der Datenmenge, eine unverzügliche Löschung der Daten nach ihrer Nutzung, technische und organisatorische Maßnahmen zur Sicherstellung von funktionaler Trennung, geeignete Anonymisierungstechniken, die Aggregierung von Daten sowie der Einsatz datenschutzfreundlicher Technologien, aber auch gesteigerte Transparenz, Verantwortlichkeit und die Möglichkeit zum Widerspruch gegen die Verarbeitung sein.

[42] Die Art. 29 Datenschutzgruppe schlägt in ihrer Stellungnahme 15/2011 zur Zweckbestimmung eine Beurteilung sämtlicher relevanter Umstände vor, insbesondere der folgenden Schlüsselfaktoren:
- das Verhältnis zwischen dem Zweck, für den die Daten erhoben wurden und den Zwecken weiterer Verarbeitung;
- der Sachzusammenhang, in dem die personenbezogenen Daten erhoben wurden und die berechtigten Erwartungen des Betroffenen der Datenverarbeitung bezüglich ihrer weiteren Nutzung;
- die Beschaffenheit der personenbezogenen Daten und die Auswirkungen der weiteren Verarbeitung auf den Betroffenen.
-die von der verarbeitenden Stelle zur Sicherstellung einer fairen Verarbeitung und zur Verhinderung unzulässiger Auswirkungen auf den Betroffenen angewandten Schutzmaßnahmen

[43] Die Art. 29 Datenschutzgruppe betont in ihrer Stellungnahme 05/2014 zu Anonymisierungstechniken, dass Anonymisierungstechniken Garantien für den Schutz der Privatsphäre bilden können, aber nur, wenn sie sachgemäß angewendet werden, d.h. dass die Voraussetzungen (Sachzusammenhang) und der Zweck (bzw. die Zwecke) des Anonymisierungsprozesses klar abgesteckt werden müssen, um den beabsichtigten Anonymisierungsgrad zu erreichen.

zeptablen Risikoniveaus[44]. Diese Prüfung ist beispielsweise im Rahmen einer Datenschutz-Folgenabschätzung zu dokumentieren.

41. Über die optimale Lösung zur Anonymisierung der Daten ist einzelfallabhängig zu entscheiden, ggf. auch mithilfe einer Kombination verschiedener Verfahren. Hierbei kommen verschiedene Anonymisierungsverfahren in Betracht; in erster Linie die Randomisierung und die Generalisierung der Daten[45]. Die Kenntnis der wesentlichen Stärken und Schwächen der einzelnen Verfahren kann bei der Gestaltung des geeigneten Verfahrens zur Anonymisierung hilfreich sein. Die Robustheit der einzelnen Verfahren sollte anhand von drei Kriterien ermittelt werden[46]:

    i. Ist es immer noch möglich, eine Einzelperson zu identifizieren?
    ii. Ist es immer noch möglich, Datensätze einer Einzelperson zuzuordnen?
    iii. Können Informationen über eine bestimmte Person abgeleitet werden?

42. Pseudonymisierte Daten sind nicht dasselbe wie anonymisierte Daten. Eine verantwortliche Stelle, die sich statt für eine Anonymisierung für eine Pseudonymisierung von Daten entscheidet, muss sich der Tatsache bewusst sein, dass diese Informationen nach wie vor als personenbezogene Daten gelten und daher zu schützen sind.

43. Es ist äußerste Vorsicht geboten, bevor pseudonymisierte oder anderweitig identifizierbare Datenbestände weitergegeben oder veröffentlicht werden. Falls die Daten detailliert sind, mit anderen Datensätzen verknüpft werden können[47] und personenbezogene Daten beinhalten, ist der Zugang zu beschränken und sorgfältig zu kontrollieren. Bei aggregierten Daten, bei denen ein geringeres Risiko einer Verknüpfung mit anderen Datensätzen besteht, ist die Wahrscheinlichkeit größer, dass diese Daten ohne erhebliche Risiken zugänglich gemacht werden können.

44. Für den Fall, dass verantwortliche Stellen anderen Einrichtungen pseudonymisierte oder auf sonstige Weise identifizierbare Daten zur Verfügung stellen, sollte vertraglich untersagt werden, Versuche zur Reidentifizierung der Daten zu unternehmen[48]. Dies sollte auch für frei verfügbare und nutzbare Daten („open data“) gelten[49].

---

[44] Die Art. 29 Datenschutzgruppe hebt in ihrer Stellungnahme 05/2014 zu Anonymisierungstechniken hervor, dass es weder möglich noch hilfreich ist, eine abschließende Aufzählung von Umständen vorzugeben, bei deren Vorliegen eine Identifikation nicht mehr möglich ist. Dennoch gibt das Dokument einige generelle Handlungsempfehlungen zur Beurteilung des Identifizierungspotentials eines bestimmten Datensatzes, der einer Anonymisierung nach den unterschiedlichen verfügbaren Techniken unterzogen wird.

[45] Randomisierung und Generalisierung sind zwei Oberbegriffe für Anonymisierungstechniken, die zum Beispiel „Hinzufügen von Rauschen“ (noise addition), Permutation, „differentielle Privatheit“ (differential privacy), Aggregierung, k-Anonymität, l-Diversität und t-Nähe („t-closeness“) abdecken.

[46] Die Art. 29 Datenschutzgruppe gibt in ihrer Stellungnahme 05/2014 zu Anonymisierungstechniken eine Übersicht zu den Stärken und Schwächen der Techniken, die die drei grundlegenden Kriterien berücksichtigen.

[47] Pseudonymisierte Informationen in einem Datensatz könnten mit Informationen in einem anderen Datensatz verbunden werden, z.B. bei Verwendung desselben eindeutigen Identifikators für jede Person.

[48] Diese Empfehlung wird auch von der Federal Trade Commission in dem Bericht “Protecting Consumer Privacy in an Era of Rapid Change” vorgebracht, FTC Report, Federal Trade Commission, March 2012

[49] Art. 29 Datenschutzgruppe, Stellungnahme 6/2013 zu den Offenen Daten ('Open Data') und der

45. Die Arbeitsgruppe empfiehlt die Einrichtung eines Gremiums oder Netzwerks, durch das jeder zur Anonymisierung oder Pseudonymisierung von Daten verpflichtete Akteur die Möglichkeit zum Austausch über die mit der Anonymisierung verbundenen Schwierigkeiten sowie von Erfahrungen erhält. Ein solches Netzwerk existiert bereits in Großbritannien (das UK Anonymisation Network (UKAN)). Es wird von den Universitäten Manchester und Southampton, dem Open Data Institute sowie dem Office for National Statistics koordiniert[50].

Mehr Transparenz und Kontrolle von der Erhebung bis hin zur Nutzung von Daten:

46. Jeder Betroffene sollte darüber informiert werden, welche Daten gesammelt werden, wie mit diesen umgegangen wird, für welche Zwecke sie genutzt werden und ob die Daten an Dritte weitergegeben werden oder nicht[51].

47. Jeder Betroffene sollte Zugang zu seinem Profil sowie zu sämtlichen seiner Informationen erhalten, sich bei der verantwortliche Stelle befinden. Jeder Betroffene sollte darüber hinaus auch über die Quellen der diversen personenbezogenen Daten informiert werden. Er sollte zudem nach Maßgabe der einschlägigen gesetzlichen Bestimmungen[52] in der Lage sein, Informationen über sich zu berichtigen und deren Nutzung in Programmen zur Erstellung von (Verhaltens-)Profilen zu widersprechen bzw. darin einzuwilligen[53].

48. Klassifizierungssysteme können nachteilige Folgen für den Einzelnen haben. Jeder Bürger sollte daher Zugang zu Informationen darüber haben, welche Algorithmen als Grundlage für eine Profilerstellung oder für Entscheidungen verwendet worden sind. Diese Informationen sollten klar und verständlich gehalten sein, um unangemessene Diskriminierungen zu verhindern und für den Betroffenen bedeutsame Entscheidungen auf falscher Tatsachengrundlage zu vermeiden[54].

49. Jedem Einzelnen sollten auf Verlangen sämtliche Daten im Besitz der verantwortlichen Stelle auf benutzerfreundliche Weise und – wo dies angemessen ist – in maschinenlesbaren, portablen Formaten zur Verfügung gestellt werden. Dies erleichtert den Wechsel zu einem anderen Diensteanbieter mit den bestmöglichen Bedingungen einschließlich des besten Schutzes der

---

Weiterverwendung von Informationen des öffentlichen Sektors ('PSI'), S. 17f.

[50] http://www.ukanon.net/

[51] Zum Beispiel bieten einige Unternehmen ihren Kunden sogenannte „Übersichtsseiten für personenbezogene Daten“ („personal data dashboards”) an, die dem Kunden einen Überblick über die Verarbeitung ihrer personenbezogenen Daten geben.

[52] Für den öffentlichen und den privaten Bereich gelten ggf. unterschiedliche Bestimmungen.

[53] Die Kommissarin der FTC, Julie Brill, hat in ihrer „Reclaim Your Name“-Initiative ähnliche Empfehlungen abgegeben. Diese Initiative zielt darauf ab, Betroffene dergestalt zu stärken, dass sie herausfinden können, wie Datenhändler ihre Daten erheben und verwenden. „Reclaim Your Name“ verschafft Kunden Zugang zu Informationen, die Datenhändler über sie angesammelt haben, und eröffnet ihnen die Möglichkeit, sich im Wege eines Opt-out-Verfahrens dieser Datenverarbeitung zu widersetzen, wenn die Betroffenen herausfinden, dass ein Datenhändler die Informationen für Marketingzwecke verkauft. Ebenso können Betroffene Informationen berichtigen lassen, die wesentlichen Entscheidungen zugrunde liegen. (Reclaim Your Name, 23rd Computers, Freedom and Privacy Conference, Grundsatzreferat v. Julie Brill, FTC, Washington, DC, 26. Juni 2013)

[54] Es ist wichtig, Transparenz bezüglich der bei der Profilbildung verwendeten Algorithmen zu schaffen. Angesichts der Komplexität von Algorithmen ist es jedoch unangemessen, zu erwarten, dass allein durch Transparenz möglichen inhärenten Verzerrungen abgeholfen wird. Automatisierte Systeme zur Entscheidungsfindung sollten ebenso Gegenstand ethischer und rechenschaftspflichtiger Aufsicht sein, wie bereits in Punkt 53 dargestellt.

Privatsphäre. Die Portabilität von Daten verhindert, dass der Kunde untrennbar an Dienstleistungen zu nicht akzeptablen Geschäftsbedingungen gebunden ist. Diese Forderung kann im Laufe der Zeit zur Entwicklung datenschutzfreundlicherer Dienstleistungen führen und die Betroffenen darin unterstützen, ihr Verständnis der sie betreffenden Daten zu verbessern.

Privacy by Design und Rechenschaftspflicht:

50. Robustere Anonymisierungsverfahren allein sind noch keine Lösung der Herausforderungen, die Big Data an den Schutz der Privatsphäre stellt. Zusätzliche Lösungen sind erforderlich. „Privacy by Design“ und Rechenschaftspflicht sind weitere wichtige Elemente zur Entschärfung der datenschutzrechtlichen Herausforderungen.

51. Die Anwendung von Big-Data-Technologien sollte auf den sieben Grundsätzen des “Privacy by Design“ aufbauen[55]. “Privacy by Design“ beinhaltet die Berücksichtigung des Schutzes der Privatsphäre in allen Stadien der Systementwicklung, in Verfahren und Geschäftspraktiken.

52. Um das Vertrauen derjenigen zu erhalten, deren personenbezogene Daten erhoben, verarbeitet und analysiert werden, bedarf es einer frühestmöglichen Abschätzung der Herausforderungen für den Schutz der Privatsphäre; auf jeden Fall vor Beginn der Verarbeitung von Big Data. Eine Möglichkeit hierfür stellt die Datenschutz-Folgenabschätzung dar. Diese sollte eine Beurteilung jeder rechtlichen Grundlage für die Verteilung und Wiederverwendung personenbezogener Daten umfassen, die Grundsätze der Zweckbegrenzung, Verhältnismäßigkeit und Datenminimierung ebenso wie Datenschutz- und Datensicherheitsmechanismen abschätzen. Bei einer solchen Folgenabschätzung sollten auch alle potenziellen Folgen für die Betroffenen sorgfältig untersucht werden[5657].

53. Rechenschaftspflicht ist ein weiterer wichtiger Grundsatz des Datenschutzes und schafft Vertrauen zwischen Betroffenen und verantwortlichen Stellen. Letztere müssen den Nachweis erbringen, dass sie ihre Rechenschaftspflicht wahrnehmen und in der Lage sind, verantwortungsvolle und ethische Entscheidungen bezüglich ihrer Nutzung von Big Data zu treffen. So müssen sich verantwortliche Stellen der Tatsache bewusst sein, dass auch ein anonymisierter

---

[55] Die sieben Prinzipien des eingebauten Datenschutzes sind: 1. Proaktiv, nicht reaktiv; als Vorbeugung und nicht als Abhilfe, 2. Datenschutz als Standardeinstellung, 3. Der Datenschutz ist in das Design eingebettet, 4. Volle Funktionalität – eine Positivsumme, keine Nullsumme, 5. Durchgängige Sicherheit – Schutz während des gesamten Lebenszyklus, 6. Sichtbarkeit und Transparenz – Für Offenheit sorgen, 7. Die Wahrung der Privatsphäre der Nutzer – Für eine nutzerzentrierte Gestaltung sorgen, https://www.privacybydesign.ca/content/uploads/2009/08/7foundationalprinciples-german.pdf

[56] Art. 29 Datenschutzgruppe, Stellungnahme 06/2013 zu Open Data sowie Stellungnahme 06/2014 zum Begriff des berechtigten Interesses der verantwortlichen Stelle gemäß Artikel 7 der RL 95/46/EG.

[57] Die EU hat ein Rahmenkonzept für Datenschutz-Folgeabschätzungen bei RFID-Anwendungen geschaffen, das dabei helfen soll, die datenschutzrechtlichen Konsequenzen des Einsatzes der RFID-Technologie zu identifizieren. Dieses Rahmenkonzept ist ebenso für Unternehmen von Interesse, die Big Data im Zusammenhang mit der Entstehung des Internet der Dinge verwenden. Das Rahmenkonzept wurde von der RFID-Industrie entwickelt und von den Datenschutzaufsichtsbehörden in der EU als mit der Datenschutzgesetzgebung konform erachtet. Die Artikel-29-Datenschutzgruppe regt die Schaffung eines gleichartigen Rahmenkonzepts für die Nutzung von Big-Data-Technologien durch auf diesen Bereich spezialisierte Unternehmen an (Europäische Kommission (2011), „Privacy and Data Protection Impact Assessment Framework for RFID Applications, 12. Januar 2011, http://ec.europa.eu/justice/policies/policy/docs/wpdocs/2011/wp/wp180_annex_en.pdf.)

Datensatz immer noch Folgen für den Einzelnen haben kann. Anonymisierte Datensätze können zur Vervollständigung existierender Profile von Personen genutzt werden, was neue Fragen für den Schutz der Privatsphäre aufwirft. Sowohl die Profile als auch die zugrunde liegenden Algorithmen bedürfen einer kontinuierlichen Begutachtung. Dies erfordert regelmäßige Kontrollen, um zu gewährleisten, dass auf der Profilerstellung aufbauende Entscheidungen verantwortungsbewusst, gerecht, ethisch und mit dem Zweck, für den die Profile genutzt werden, vereinbar sind. Eine ungerechte Behandlung Einzelner aufgrund automatisierter falsch positiver oder falsch negativer Ergebnisse gilt es zu vermeiden[58].

Verbesserung von Wissen und Bewusstsein:

54. Das Wissen um und Bewusstsein von mit Big Data verbundene Herausforderungen für den Datenschutz sind bedeutsam für verantwortliche Stellen, die diese Technologie einsetzen. Die Industrie muss sich dieser Herausforderungen stellen und Schulungen hinsichtlich Maßnahmen zu ihrer Lösung anbieten, beispielsweise in Form von “Privacy by Design“.

55. Der Schutz der Privatsphäre sowie die Herausforderungen für den Datenschutz im Zusammenhang mit Big Data sollten an Universitäten und sonstigen Hochschulen behandelt werden, an denen Informatik oder Informationswissenschaften gelehrt werden.

56. Behörden müssen über das notwendige Wissen und Bewusstsein hinsichtlich des Potenzials von Big Data verfügen. Dies ist insbesondere in Verbindung mit der Formulierung neuer Gesetze und Bestimmungen von Bedeutung. Weiterhin ist Problembewusstsein nötig, damit Behörden in der Lage sind, ihre Aufgabe zum Schutz einer Reihe zentraler Werte der Gesellschaft wahrnehmen können.

[58] „Uruguay Erklärung zum Profiling“ (2012), 34. Internationale Konferenz der Datenschutzbeauftragten, 25.–26. Oktober 2012, http://privacyconference2012.org/wps/wcm/connect/7b10b0804d5dc38db944fbfd6066fd91/Uruguay_Declaration_final.pdf?MOD=AJPERES .

Die Big-Data-Wertschöpfungskette