Alexander Dix

Datenschutz im Zeitalter von Big Data. Wie steht es um den Schutz der Privatsphäre?

Big Data wird in der aktuellen Diskussion oft als unvereinbar mit dem Datenschutz und dem Schutz der Privatsphäre dargestellt. Dabei ist eine differenzierte Analyse nötig, will man die möglichen positiven Auswirkungen von Big-Data-Analysen grundrechtskonform und datenschutzgerecht erreichen. Der neue europäische Rechtsrahmen für den Datenschutz wird dabei eine wesentliche Rolle spielen. Allerdings sind darüber hinaus auch ethische Überlegungen notwendig, soll der Einsatz von Big Data gesellschaftlich akzeptabel gestaltet werden.

Dr. Alexander Dix
LL.M., von 2005 bis 2016 Berliner Beauftragter für Datenschutz und Informationsfreiheit sowie Vorsitzender der Internationalen Arbeitsgruppe zum Datenschutz in der Telekommunikation; Stellv. Vorsitzender der Europäischen Akademie für Informationsfreiheit und Datenschutz (www.eaid-berlin.de)
dix@eaid-berlin.de

Schlüsselwörter:
Algorithmen – Anonymisierung – Datenschutz-Grundverordnung – Predictive Policing – Smart City

Einleitende Begriffsklärungen

Datenschutz und Privatsphäre sind nicht, wie es der Titel dieses Beitrags nahelegt, synonyme Begriffe. Das ergibt sich schon daraus, dass die Charta der Grundrechte der Europäischen Union die Achtung des Privatlebens einerseits und den Datenschutz andererseits in zwei getrennten, allerdings unmittelbar zusammenhängenden Artikeln (Art. 7 und 8) garantiert. Das Recht des Einzelnen, über seine Daten grundsätzlich selbst zu entscheiden, ist nicht deckungsgleich mit seinem Anspruch, vor Eingriffen in seine private und familiäre Sphäre geschützt zu sein.

Das übersehen diejenigen, die den Datenschutz als überholt ansehen und sich mit dem Schutz der Privatsphäre begnügen wollen (so Härting/Schneider). Auch im öffentlichen, städtischen Raum, also jenseits der privaten, häuslichen Sphäre hat der Einzelne ein grundrechtlich verbürgtes Recht, nicht auf Schritt und Tritt z. B. durch Kameras beobachtet zu werden (BVerfG, Beschluss v. 20. März 2007 [1 BvR 2368/06]). Das systematische Fotografieren von Hausfassaden durch Google Street View war zwar kein Eingriff in die Privatsphäre. Wohl aber verletzte das Einstellen dieser Bilder in eine weltweit zugängliche Datenbank solange den Datenschutz, wie die Hausbewohner dem nicht vorab widersprechen konnten. Die Privatsphäre war allerdings durch das heimliche Abhören der Kommunikation über häusliche WLAN-Router mithilfe der Street View-Fahrzeuge verletzt, deretwegen Google weltweit in mehreren Staaten strafrechtlich belangt worden ist.

Es gibt also Überschneidungen zwischen dem Datenschutz und dem Schutz der Privatsphäre. Das Bundesverfassungsgericht hat das Recht auf informationelle Selbstbestimmung aus der Garantie der Menschenwürde (Art. 1 GG) und dem allgemeinen Persönlichkeitsrecht (Art. 2 Abs. 1 GG) abgeleitet. Der Europäische Gerichtshof rekurriert in seiner neueren Rechtsprechung zu Fragen des Datenschutzes (Google, Vorratsdatenspeicherung, Facebook) sowohl auf die Art. 7 und 8 der Europäischen Grundrechte-Charta. Deshalb erscheint es gerechtfertigt, die Auswirkungen von Big Data auf den Datenschutz und die Privatsphäre gemeinsam zu betrachten. Das ist auch vor dem Hintergrund der transatlantischen Debatte naheliegend. Bei allen Unterschieden zwischen dem europäischen Datenschutz und „privacy" im US-amerikanischen Verständnis wird doch auf beiden Seiten des Atlantiks zunehmend erkannt, dass sich die Freiheit des Einzelnen nicht auf den Schutz seiner häuslichen, „privaten" Sphäre beschränkt.

Allerdings herrschen in den USA andere Vorstellungen darüber, wo die Freiheit des Einzelnen beginnt und welche Einschränkungen er hinnehmen muss.

Wenn vom Zeitalter von Big Data die Rede ist, wird eine grundlegend neue Form der Analyse unstrukturierter Datenmengen in den Blick genommen, die möglicherweise eine Reihe neuer Fragen aufwirft. Allerdings ist daran zu erinnern, dass sich das Bundesverfassungsgericht schon sehr früh mit Risiken auseinandergesetzt hat, die mit der Sammlung und Auswertung großer Datenmengen – freilich noch in strukturierter Form – einhergehen, nämlich der amtlichen Statistik. So hat das Gericht bereits im Mikrozensusbeschluss von 1969 betont, dass der Einzelne von Verfassungs wegen vor einer umfassenden zwangsweisen Registrierung und Katalogisierung durch den Staat zu schützen ist (BVerfGE 27, 1, 6).

Darüber hinaus ist vor allem das Volkszählungsurteil von 1983 als weitsichtige Auseinandersetzung mit den Problemen zu verstehen, wie sie durch Big Data-Anwendungen qualitativ noch verschärft werden (Sädtler 2015, 69, 80 f.). Das Bundesverfassungsgericht hat vor mehr als 30 Jahren mit Blick auf die amtliche Statistik festgestellt:

„Ist die Vielfalt der Verwendungsmöglichkeiten und Verknüpfungsmöglichkeiten damit bei der Statistik von der Natur der Sache her nicht im Voraus bestimmbar, müssen der Informationserhebung und Informationsverarbeitung innerhalb des Informationssystems zum Ausgleich entsprechende Schranken gegenüberstehen. Es müssen klare Verarbeitungsvoraussetzungen geschaffen werden, die sicherstellen, dass der Einzelne unter den Bedingungen einer automatischen Erhebung und Verarbeitung der seine Person betreffenden Angaben nicht zum bloßen Informationsobjekt wird." (BVerfGE 65, 1, 48)

Diese Aussagen betrafen seinerzeit und betreffen heute unmittelbar die amtliche Statistik und die auf ihr aufbauende Stadtforschung. Ihre Bedeutung geht aber weit über das Staat-Bürger-Verhältnis hinaus. Sie betreffen in gleicher Weise auch die Datenverarbeitung durch Unternehmen, die Daten gern als „das Öl des 21. Jahrhunderts" bezeichnen und ihre Geschäftsmodelle auf Big Data-Analysen aufbauen. Zwar gibt es einen wesentlichen Unterschied zwischen der staatliche Datenverarbeitung – etwa durch die amtliche Statistik – einerseits und der Datenerhebung durch Internetanbieter, Versicherungsunternehmen und andere private Datenverarbeiter andererseits, weil nur der Staat im Gegensatz zur Wirtschaft den Bürger zur Offenbarung seiner Daten verpflichten kann. Aber für den betroffenen Bürger macht es keinen Unterschied, ob eine Behörde oder ein Unternehmen ihn durch die automatische Verarbeitung seiner Daten zum „bloßen Informationsobjekt" macht. Sein Grundrecht auf Schutz vor informationeller Fremdbestimmung richtet sich nicht nur gegen den Staat bei Volkszählungen oder Mikrozensen, sondern es verpflichtet den Gesetzgeber auch dazu, die Menschen vor unbegrenzter Erfassung und Auswertung ihrer Daten durch privatwirtschaftliche Akteure zu schützen.

Zunächst sollte aber eine Klarstellung getroffen werden, bevor die von Big Data ausgehenden Risiken für Datenschutz und Privatsphäre untersucht werden: Big Data heißt nicht notwendig Big *Personal* Data. Wo große Datenmengen von vornherein ohne jeden denkbaren Personenbezug analysiert werden (z. B. in der Klima- oder Weltraumforschung), entstehen auch keine Gefahren für den Datenschutz, geschweige denn für die Privatsphäre. Daraus folgt zugleich, dass Big-Data-Analysen nur insoweit und solange Datenschutzrisiken auslösen, wie sie mit personenbezogenen Daten durchgeführt werden. Diese Risiken lassen sich beherrschen, wenn man die erhobenen Datenmengen vor der Analyse frühzeitig anonymisiert oder den Personenbezug zumindest (z. B. durch Pseudonymisierung) stark reduziert. So ist auch der Grundsatz der Datenvermeidung oder Datensparsamkeit zu verstehen, wie er in § 3a des Bundesdatenschutzgesetzes Ausdruck gefunden hat. Er ist seinerseits Konsequenz des Prinzips der Erforderlichkeit, dem jede Verarbeitung personenbezogener Daten unterliegt. Das wird in der Diskussion über Big Data häufig verkannt, wenn im Gegensatz zur Datensparsamkeit der „Datenreichtum" oder die Datenmaximierung als erstrebenswertes Ziel deklariert werden. Datensparsamkeit bedeutet nicht die Begrenzung jeglicher Datenverarbeitung, sondern allein der weitgehende Verzicht auf den Personenbezug.

Statistikgesetze als frühe „Big-Data-Gesetze"

In diesem Bereich verfügt gerade die amtliche Statistik spätestens seit dem Volkszählungsurteil von 1983 über wertvolle Erfahrungen. Man kann die in der Folge dieser Rechtsprechung ergangenen Statistikgesetze des Bundes und der Länder insofern als frühe „Big-Data-Gesetze" für den öffentlichen Bereich verstehen. Insbesondere durch das Bundestatistikgesetz von 1987 (§ 12) wurden der Grundsatz der frühestmöglichen Reduzierung des Personenbezugs durch Trennung von Hilfs- und Erhebungsmerkmalen und alsbaldiger Vernichtung der Hilfsmerkmale, die Begrenzung der Verknüpfungsmöglichkeiten innerhalb der Statistikämter und das Statistikgeheimnis zum festen Bestandteil des gesetzlichen Rahmens für die amtliche Statistik in Deutschland. Zwar gibt es bestimmte Bereiche der Forschung wie etwa medizinische Kohortenstudien (z. B. die Nationale Kohorte), wo der Personenbezug bei Big-Data-Analysen Teil des legitimen Erkenntnisinteresses ist (z. B. um einen Probanden über Untersuchungsergebnisse zu informieren, die für ihn lebenswichtig sein können). Für den hier interessierenden Bereich der Stadtforschung und Städtestatistik wie für die meisten anderen Felder empirischer Forschung gilt aber, dass der Einzelne letztlich nicht als Person, sondern nur als Träger bestimmter Eigenschaften von Bedeutung ist, die in aggregierten Ergebnissen zusammengefasst werden.

Anonymisierung

Die Konzepte der Anonymität und der Anonymisierung sind von zentraler Bedeutung für den modernen Datenschutz insbesondere angesichts der Herausforderungen durch Big Data. Methoden der Anonymisierung oder De-Identifizierung spielen im Übrigen auch eine zunehmende Rolle bei der Überbrückung transatlantischer Differenzen beim Datenschutz. So haben Experten aus Europa und den USA im Rahmen des „Privacy-Bridges"-Projekts 2015 die Anonymisierung als eine von zehn möglichen Brücken zur Überwindung von Schwierigkeiten aufgrund des ungleichen Datenschutzniveaus auf

beiden Seiten des Atlantik identifiziert (Privacy Bridges, Bridge 6, De-identification of personal data). Anonymität war allerdings nie statisch zu verstehen. Das Bundesdatenschutzgesetz (§ 3 Abs. 6) definiert „Anonymisieren" als das „Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können." Welcher Aufwand an Zeit, Kosten und Arbeitskraft „unverhältnismäßig groß" ist, hängt vom Stand der Technik und der Menge der zur Verfügung stehenden Daten ab. Gerade das zuletzt genannte Kriterium bezeichnet die entscheidende durch Big Data eingetretene Veränderung: Je mehr Daten für die Analyse durch hochentwickelte Werkzeuge und Algorithmen zur Verfügung stehen, desto höher ist das Risiko der Reidentifizierung, also der Aufhebung oder Vereitelung von Anonymität.

Ob ein Datenbestand von vornherein anonym ist oder später anonymisiert worden ist, bedarf deshalb einer eingehenden Analyse, die nicht ein für allemal vorgenommen werden kann, sondern vor dem Hintergrund der technischen Entwicklung und der sich verändernden Datenmengen periodisch wiederholt werden muss. Während es bei der Volkszählung noch ausreichte, nach Abschluss der Plausibilität die abgetrennten Hilfsmerkmale zu vernichten, kann – zumindest faktische – Anonymität im Sinne des Bundesdatenschutzgesetzes wie auch der Statistikgesetze bei der Auswertung von Datenströmen, wie sie im städtischen Umfeld Sensoren, autonome Fahrzeuge oder „smarte" Stromzähler generieren, nur mit sehr viel größerem Aufwand hergestellt werden.

Allerdings wäre es verfehlt anzunehmen, im Zeitalter von Big Data wäre angesichts der modernen Analysetools eine im datenschutzrechtlichen Sinne ausreichende Anonymität überhaupt nicht mehr zu erreichen. Richtig ist aber, dass sich der Aufwand, den die Reidentifizierung erfordert, in Zukunft minimieren wird (Sädtler, 81). Deshalb ist die permamente Bewertung des Reidentifizierungsrisikos von so großer Bedeutung für die Privatsphäre und den Datenschutz bei Big-Data-Analysen. Sowohl die Internationale Arbeitsgruppe zum Datenschutz in der Telekommunikation als auch die Art. 29-Gruppe der europäischen Datenschutzbehörden haben dies unterstrichen (Berlin Group [2014], Art. 29-Datenschutzgruppe, [2014]). Insbesondere die Art. 29-Gruppe hat die verschiedenen komplexen Techniken der Reduktion oder des Ausschlusses des Personenbezugs (Randomisierung durch stochastische Überlagerung, Ersetzung oder Differential Privacy, Generalisierung durch Aggregation, k-Anonymität oder L-Diversität, Pseudonymisierung durch Verschlüsselung, Hashen oder Tokenisierung) eingehend auf ihre jeweiligen Stärken und Schwächen untersucht und einen Leitfaden zu Anonymisierungstechniken veröffentlicht. Das ist deshalb von besonderer Bedeutung, weil Anonymität in den Mitgliedstaaten der Europäischen Union nicht einheitlich verstanden wird. Da keine der beschriebenen Techniken für sich genommen zuverlässig die Kriterien einer wirksamen Anonymität erfüllt, ist nach den Vorgaben der Art. 29-Gruppe stets eine Einzelfallbewertung notwendig, deren Ergebnis der Datenschutzbehörde vorgelegt werden sollte, um das verbleibende Reidentifizierungsrisiko im Zusammenhang mit dem Verwendungszweck der Daten zu beurteilen.

Zweckbindung

Das führt zum nächsten Datenschutzgrundsatz, der durch Big Data in besonderer Weise in Frage gestellt wird: dem Grundsatz der Zweckbindung. Mit den Worten der Internationalen Arbeitsgruppe zum Datenschutz in der Telekommunikation: „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 (Berlin Group, Rn. 20)." Je größer das Risiko der De-Anonymisierung, des Herausgreifens Einzelner (singling out) bei Datenbeständen ist, desto wichtiger wird die Beachtung des Grundsatzes, dass personenbezogene Daten nur für bestimmte, bei der Erhebung klar definierte Zwecke verwendet werden dürfen. Jede darüber hinausgehende, zweckfremde Nutzung bedarf der informierten Einwilligung der betroffenen Personen. Der Grundsatz der Zweckbindung steht in einem engen Zusammenhang mit dem Erforderlichkeitsprinzip. Die Erforderlichkeit der personenbezogenen Datenverarbeitung lässt sich immer nur in Bezug auf einen hinreichend bestimmten Zweck ermitteln. Wird die Bindung der weiteren Verwendung der Daten an den ursprünglichen Zweck aufgehoben, verliert auch der Erforderlichkeitsgrundsatz seine eingriffsbeschränkende Funktion (Richter 2015, 735). Eine vollkommen zweckfreie Verarbeitung personenbezogener Daten war auch schon vor den Zeiten von Big Data unzulässig. Aber auch pauschale Zwecksetzungen z. B. in der Forschung (Nationale Kohorte) können nur dann Grundlage einer die Datenverarbeitung legitimierenden informierten Einwilligung sein, wenn diese Einwilligung nicht unbefristet gilt, sondern jederzeit zumindest mit Wirkung für die Zukunft widerrufen werden kann. Die auch vom Bundesverfassungsgericht anerkannte Besonderheit der amtlichen Statistik, dass dort keine enge und konkrete Zweckbindung verlangt werden kann und auch ein legitimes Bedürfnis nach Vorratsspeicherung besteht (BVerfGE 65, 1, 47) hat das Gericht gerade dazu veranlasst, zum Ausgleich bestimmte Schranken in Form von Verarbeitungsvoraussetzungen für die amtliche Statistik vorzuschreiben (s. o. S.). Allerdings gefährden im privatwirtschaftlichen Bereich auf Big Data basierende Geschäftsmodelle ebenso wie ihre Vorläufer des *Datamining* und des *Data Warehousing* die Zweckbindung in besonderer Weise (Simitis, Einleitung, Rz.111).

Transparenz

Ein weiteres der durch Big-Data-Analysen verursachten Hauptprobleme besteht in der Profilbildung unter Verwendung komplexer und intransparenter Algorithmen. Gerade die Intransparenz der Datenverarbeitung birgt verstärkte Risiken der informationellen Autonomie des Einzelnen, weil sie ihm die Möglichkeit nimmt, seine Rechte z. B. auf Auskunft, Korrektur oder Löschung wahrzunehmen (Dix in: Simitis, § 6 Rz. 10). Zu unterstützen ist deshalb die Forderung des Bundesjustizministers, der kürzlich gefordert hat, dass kein Mensch zum Objekt eines Algorithmus werden dürfe (Maas, Unsere digitalen Grundrechte (2015). Diese Forderung ist eine andere Formulierung für die Vorgabe des Bundesverfassungsgerichts, dass

der Einzelne nicht zum Objekt der Informationsverarbeitung gemacht werden dürfe. Angesichts der wachsenden, mit Hilfe von Algorithmen ausgeübten Informationsmacht wird die Forderung nach Offenlegung und Kontrolle dieser Algorithmen immer wichtiger. Die gegenwärtige Praxis des Kreditscoring ist für viele Menschen mitentscheidend für die Wahrnehmung von Entfaltungschancen z. B. bei der Wohnungsvergabe oder in finanzieller Hinsicht. Schon dieses Kreditscoring ist von einer nicht akzeptablen Intransparenz gekennzeichnet, die der Bundesgerichtshof unter Berufung auf Betriebs- und Geschäftsgeheimnisse der Auskunfteibranche rechtfertigt (vgl. Dix in Simitis, § 34, Rn. 33). Denn die Betroffenen können nicht einmal Auskunft über die bei der Ermittlung ihres Scorewerts zugrundegelegten Faktoren oder die Vergleichsgruppen verlangen, denen sie zugeordnet werden. Die Gewichtung der einzelnen Faktoren muss die Auskunftei nur dann offenlegen, wenn dem Betroffenen eine automatisierte Einzelentscheidung droht. Diese wurde bereits in der Europäischen Datenschutzrichtlinie von 1995 an enge Voraussetzungen geknüpft.

Herrschaft durch Algorithmen

Das Kreditscoring ist jedoch nur ein Vorgeschmack dessen, was künftig möglich sein wird. Algorithmen können immer besser Schrift, Sprache und Muster erkennen, sie steuern 70 Prozent aller Finanztransaktionen und erzeugen zum Teil Zeitungsinhalte automatisch (Helbing u. a., Digital-Manifest, 51). In der Volksrepublik China wird an der Einführung eines Citizen Score gearbeitet, mit dessen Hilfe Bürger aufgrund ihres Surfverhaltens im Internet und ihres sonstigen politischen Wohlverhaltens durch die Vergabe von Punkten auf einer Ranking-Scala eingeordnet werden sollen. Dieses Punktekonto entscheidet nicht nur über die Vergabe von Krediten, sondern auch darüber, ob der Betroffene einen bestimmten Beruf ergreifen oder nach Europa reisen darf. Die Verwirklichung solcher Orwell-Huxley'scher Szenarien ist keineswegs auf China begrenzt. In Singapur, für manche das Vorbild einer datengesteuerten Gesellschaft, für andere das Paradebeispiel einer fürsorglichen Diktatur, wird ebenfalls an Konzepten der staatlichen Verhaltenssteuerung durch Sammlung von Bürgerdaten gearbeitet.

Ein besonderes Risiko von Big-Data-Analysen sieht die Internationale Arbeitsgruppe zum Datenschutz in der Telekommunikation darin, dass die „Big-Data-Haltung" häufig von der irrigen Annahme ausgeht, je mehr Daten man sammele und auf je mehr Daten man zugreifen könne, desto fundiertere, genauere und bessere Entscheidungen können man treffen (Berlin Group, Rn. 30). Stattdessen sollte die Erkenntnis selbstverständlich sein, dass mehr Daten nicht notwendig mehr Wissen bedeuten. Mehr Daten können auch zu Verwirrung und zu mehr „falsch-positiven" Ergebnissen führen, wie das Beispiel der Google-„Grippe-Trends" gezeigt hat. Big Data kann bestehende Vorurteile verfestigen und zu sozialer Ausgrenzung und informationeller Diskriminierung ganzer Bevölkerungsgruppen (social sorting) führen. Korrelationsanalysen können im Einzelfall zu völlig falschen Ergebnissen führen, insbesondere wenn Korrelation mit Kausalität verwechselt wird. Insofern können Diskriminierungen aufgrund statistischer Analysen auch zu Verletzungen des Datenschutzes im Einzelfall gegenüber den Angehörigen ausgegrenzter Gruppe führen. Generell kann eine Entwicklung, bei der immer mehr gesellschaftliche Entscheidungen auf Algorithmen basieren, zu einer „Diktatur der Daten" führen, in der wir nicht mehr anhand tatsächlicher Handlungen, sondern anhand dessen, was nach Datenlage die wahrscheinlichen Handlungen sein werden, beurteilt werden (Mayer-Schönberger/Cukier).

Sollten zukünftig – und erste Anzeichen dafür gibt es bereits – Bonitätswerte und Versicherungswerte von Informationen abhängen, die Nutzer im Internet oder in sozialen Netzen hinterlassen (oder sogar davon, dass soziale Netze gemieden werden), dann wäre ein massiver Einschüchterungseffekt (chilling effect) die Folge. Denn beispielweise könnte die Veröffentlichung der Erbkrankheit eines Kindes durch seine Eltern in sozialen Netzen zum späteren Verlust des Versicherungsschutzes für das Kind führen (vgl. das Beispiel der Berlin Group, Rn. 31). Dann aber würden Nutzer ihr Verhalten sowohl online wie offline möglicherweise danach ausrichten, was dies bei künftigen Auswertungen ihrer Datenspuren für negative Konsequenzen haben könnte. Der Verzicht auf die Wahrnehmung von Teilhabmöglichkeiten und Grundrechten wäre die Folge. Das hätte nicht nur Auswirkungen auf die Privatsphäre des Einzelnen, sondern auf das demokratische Gemeinwesen als Ganzes.

Big Data zur Kriminalitätsbekämpfung

Der britische Geheimdienst GCHQ soll bereits jetzt im Rahmen seines Programms „Karma Police" sämtliche Website-Aufrufe überwachen (https://tarnkappe.info/karma-police-gchq-ueberwacht-alle-website-besuche/). Während das im Dezember 2015 in Kraft getretene deutsche Gesetz zur Einführung einer Speicherpflicht und einer Höchstspeicherfristen die Vorratsspeicherung von Inhaltsdaten noch ausdrücklich ausschließt (§ 113b Abs. 5 TKG, BGBl.I 2015, 2218, 2222), will die britische Regierung durch einen neuen Gesetzentwurf (Investigatory Powers Bill) auch die Provider zur anlasslosen und flächendeckenden personenbezogenen Speicherung der Namen besuchter Internet-Domains verpflichten. Diese sollen im Eilfall von der Polizei auch ohne richterliche Anordnung für die Sicherheitsbehörden zugänglich sein, um jede Form der Kriminalität (nicht nur den Terrorismus) bekämpfen zu können (http://www.heise.de/newsticker/meldung/Grossbritannien-Entwurf-fuer-Netzueberwachungsgesetz-vorgestellt-2877521.html). Das Gesetz soll im Frühjahr 2016 verabschiedet werden. Diese Entwicklung macht deutlich, dass sich mit dem Argument „Kriminalitätsbekämpfung" jede, auch eine inhaltliche Form der Vorratsdatenspeicherung rechtfertigen lässt (ohne dass deren Nutzen bisher belegt worden wäre). Aus Sicht der Sicherheitsbehörden wäre es deshalb wünschenswert, jede Form der Kommunikation (Sprachtelefonie, E-Mail- und sonstiger Datenverkehr) auf Vorrat zu speichern. Dies praktizieren die National Security Agency und die übrigen Geheimdienste der „Five Eyes" offenbar bereits seit Jahren. Der Koordinator der US-Geheimdienste, Clapper, soll nach den von Edward Snowden initiierten Veröffentlichungen lediglich eingeräumt haben, er bedauere es, die Geheimdienstaktivitäten nicht selbst früher öffentlich gemacht zu ha-

ben. Auch beim Bundesnachrichtendienst spielen rechtliche Schranken offenbar eine nachgeordnete Rolle. Wo sie existieren, wird ihre Einhaltung nur unzureichend kontrolliert. Inwieweit das deutsche Gesetz zu Höchstspeicherfristen und die britischen Befugnisse zur Netzüberwachung Bestand haben werden, müssen die Gerichte insbesondere auf europäischer Ebene entscheiden. Dabei wird es um mehr gehen als um Transparenz: Entscheidend wird sein, ob derartig pauschale Überwachungsmaßnahmen in einer demokratischen Gesellschaft zwingend notwendig sind. Nur dann rechtfertigen sie nach der Europäischen Menschenrechtskonvention und der Europäischen Grundrechte-Charta eine derart weitreichende Einschränkung von Grundrechten.

Dass Sicherheitsbehörden sich verstärkt der Big-Data-Methoden bedienen wollen, zeigt auch das Beispiel des Predictive Policing. Polizeibehörden in Zürich und München setzen Analyse-Software mit selbst lernenden Algorithmen ein, um anhand von Daten aus den verschiedensten Quellen (z. B. Kriminalstatistik, soziale Netzwerke) die Wahrscheinlichkeit künftiger Straftaten oder Gefahren in bestimmten Stadtteilen vorherzusagen und die Polizeipräsenz dort entsprechend zu erhöhen. Während Nordrhein-Westfalen eine entsprechende Software testen will, hat Niedersachsen einen entsprechenden „Orakel-Versuch" mittlerweile gestoppt. Wissenschaftliche Untersuchungen darüber, was eine solche Software tatsächlich leistet, gibt es bisher nicht. Die Konferenz der Datenschutzbeauftragten des Bundes und der Länder hat darauf hingewiesen, dass durch solchen Vorfeldanalysen stets die Gefahr von fehlerhaften Prognosen mit entsprechenden Auswirkungen auf die dabei in Verdacht geratenen Personen innewohnt. Zwar bestünden derartige Gefahren bei rein orts- und nicht personenbezogenen Auswertungen noch nicht. Diese Bewertung könne sich jedoch bereits bei geringfügigen Modifikationen verändern, etwa wenn allgemein zugängliche Daten aus sozialen Netzen in die Analyse miteinbezogen werden. Zudem weist die Konferenz darauf hin, dass „die beständig weiterentwickelten technischen Auswertungsmöglichkeiten schon heute das Potential dafür bergen, dass die Bürgerinnen und Bürger die Kontrolle über ihre Daten in einem Umfang und auf eine Art und Weise verlieren könnten, die in der Vergangenheit nicht vorstellbar gewesen ist." (Entschließung v. 19.03.2015)

Informationssicherheit

Je größer die Menge der erhobenen und analysierten Daten ist, desto mehr wächst das Risiko von Datenlecks. Das mag daran liegen, dass große Datenbanken wie „Honigtöpfe" auf kriminelle Angreifer wirken. Aber auch die Delegation von IT-Dienstleistungen durch den Staat an Dritte birgt Risiken für die Datensicherheit, wie das Beispiel Edward Snowdens auch gezeigt hat (auch wenn die von ihm veranlassten Veröffentlichungen im öffentlichen Interesse lagen). Angriffe auf oder Sicherheitslücken in großen personenbezogenen Datenbanken können weitreichende Konsequenzen für den Schutz der Privatsphäre und den Datenschutz haben. Deshalb gelten gesteigerte Anforderungen an die Informationssicherheit bei Big-Data-Anwendungen, wie dies z. B. im Gesetz zur Einführung einer Speicherpflicht und einer Höchstspeicherfrist für Verkehrsdaten (§§ 113d – 113g TKG, BGBl. I, 2015, 2218, 2223 f.) zum Ausdruck kommt. Eine mögliche Maßnahme zur Begrenzung der Informationssicherheitsrisiken könnte die konsequente Dezentralisierung der Datenhaltung sein, wie sie die Autoren des Digital-Manifests fordern (Helbing u. a., Digital-Manifest, 58).

Europäische Datenschutz-Grundverordnung

Mit der Anfang Dezember 2015 erzielten politischen Einigung über den künftigen Rechtsrahmen für den Datenschutz in Europa, insbesondere über die hier interessierende Datenschutz-Grundverordnung, kann die Frage beantwortet werden, ob und wenn ja welche Anforderungen der europäische Gesetzgeber für Big Data formuliert hat. Im Vorfeld war einerseits zu befürchten, dass vor allem die Vorschläge des Rats für den Trilog zu einer problematischen Aufweichung des Zweckbindungsgrundsatzes führen würden, falls sich das Europäische Parlament und die Kommission dem angeschlossen hätten. Das hätte nicht nur Auswirkungen auf die Verarbeitung von personenbezogenen Daten für kommerzielle Zwecke, sondern insbesondere auch für Forschungszwecke gehabt. Andererseits äußerte die Vorsitzende des Rats für Sozial- und Wirtschaftsdaten, Regina Riphahn, mehrfach öffentlich die Befürchtung, dass eine Umsetzung der Vorschläge des Europäischen Parlaments für die Datenschutz-Grundverordnung zur einer Blockade der empirischen Sozialforschung in Deutschland führen würden (vgl. zuletzt Der Tagesspiegel v. 08.12.2015 – Der Wissenschaft droht Blockade durch Datenschutz). Ob diese Befürchtungen begründet waren, kann dahinstehen. Denn die nun vorgesehenen Regelungen, die im Frühjahr 2018 in Kraft treten werden, verwässern weder die Zweckbindung noch erschweren sie die Forschung. Zum einen hat sich der Rat mit seinem Vorschlag, die Nutzung personenbezogener Daten für inkompatible Zwecke unter vagen Voraussetzungen zuzulassen, nicht durchsetzen können. Das ist eine wichtige Bekräftigung des Zweckbindungsgrundsatzes gerade auch im Zusammenhang mit Big Data. Die Grundverordnung sieht zugleich vor, dass die Verwendung personenbezogener Daten für Zwecke der Forschung und Statistik keine unzulässige Zweckentfremdung dieser Daten darstellt (Art. 5 [1] [b], 83 [1] der Grundverordnung). Der in Art. 83 der Grundverordnung gefundene Kompromiss betont zum einen den Grundsatz der Datensparsamkeit als wesentliches Element von „Privacy by Design", zum anderen macht er deutlich, dass diesem Grundsatz auch durch Pseudonymisierung Rechnung getragen werden kann. Dass personenbezogene Daten für bestimmte Forschungszweige (z. B. Zeitgeschichte) erforderlich sind, war nie streitig. Dem trägt die Grundverordnung jetzt durch entsprechende Regelungen für das Archivwesen und die historische Forschung Rechnung. Aber auch die empirische Sozial- und Wirtschaftsforschung wird selbst nach Auffassung des Rates für Wirtschafts- und Sozialdaten (vgl. dessen Presseerklärung vom 23.12.2015) nicht gefährdet. Der deutsche Gesetzgeber wird die den Mitgliedstaaten in der Verordnung eingeräumte Konkretisierungbefugnis sicher auch nicht dazu nutzen, um die Möglichkeiten der Forschung weiter einzuschränken. Was man an der Datenschutz-Grundverordnung allerdings

kritisieren muss, ist die Tatsache, dass sie keine ausreichende Regelung zur Profilbildung enthält. Entsprechende Vorschläge aus Deutschland (die allerdings von Seiten der deutschen Wirtschaft stets bekämpft wurden) fanden weder im Rat noch im Trilog genügend Fürsprecher. Bei der in die Verordnung aufgenommenen Regelung (Art. 20) bleibt unklar, inwieweit sie über das bereits bestehende Verbot der automatisierten Einzelentscheidung hinaus bestimmte Formen der Profilbildung gerade durch Big-Data-Analysen unterbindet.

Smart City

Zunehmend entwickeln Großstädte in aller Welt Konzepte für ein „Smart City", in der sowohl Umweltprobleme gelöst (z. B. durch intelligente Mülltonnen oder Straßenlaternen, die sensorgesteuert nur dann Licht geben, wenn Menschen die Straße benutzen), als auch Sicherheitsfragen bis hin zur Terrorismusbekämpfung beantwortet werden können (vgl. den Bericht von H. Rauterberg in DIE ZEIT v. 26.11.2015, 49 f.). Die Verfügbarkeit und der Einsatz von Detektoren und Sensoren bis hin zum Angebot offener WLANs eröffnet tatsächlich Möglichkeiten, um Gefahren im öffentlichen Raum zu reduzieren. Allerdings darf man nicht – wie schon bei der Debatte über die Videoüberwachung – dem Fehlschluss erliegen, mehr Überwachung und mehr Datenquellen für Big Data-Analysen könnten auch nur annähernd eine vollständige Sicherheit garantieren. Im übrigen ist sehr zweifelhaft, ob die Bürgerinnen und Bürger sich in einer Smart City wohlfühlen würden, in der sie auf Schritt und Tritt von Sensoren begleitet und beobachtet werden, auch wenn manche die Stadt mithilfe der Technik fast zu ihrem „zweiten Wohnzimmer" machen möchten (Rauterberg, ebda.). Max Weber hat 1920 die bürokratische Organisation als „lebende Maschine" bezeichnet, die im Verein mit der toten Maschine an der Arbeit ist, das „Gehäuse jener Hörigkeit der Zukunft herzustellen, in welche dereinst die Menschen sich, wie die Fellachen im altägyptischen Staat, ohnmächtig zu fügen gezwungen sein werden …" (Weber, 151). Die Smart City darf nicht zu einem solchen Gehäuse der Hörigkeit werden, sondern sie muss den Menschen stets auch im öffentlichen Raum die Möglichkeit bieten, sich unbeobachtet und damit frei zu bewegen. Nur in einer solchen wirklich „smarten" Stadt wollen die Menschen leben (Dix, 7069).

Fazit

„Big Data" ist der Oberbegriff für eine Vielzahl von Methoden zur Auswertung großer strukturierter und unstrukturierter Datenmengen. Die Auswirkungen von Big Data auf den Datenschutz und den Schutz der Privatsphäre lassen sich nicht pauschal, sondern nur bezogen auf konkrete Anwendungsszenarien beurteilen. Insbesondere ist „Big Data" nicht mit „Big Personal Data" gleichzusetzen. Zahlreiche wichtige Big-Data-Anwendungen sind ohne die Nutzung personenbezogener Daten möglich. Bei der Nutzung großer Mengen personenbezogener Ausgangsdaten verfügt die amtliche Statistik über eine lange Erfahrung. Die Statistikgesetze können als frühe Big-Data-Gesetze angesehen werden. Auch bei neueren Big-Data-Analysen außerhalb der Statistik, z. B. im Bereich der Forschung, sind die Grundsätze der frühzeitigen Anonymisierung oder der Reduzierung des Personenbezugs durch Pseudonyme zu beachten. Dies hat ebenso Eingang in die Europäische Datenschutz-Grundverordnung gefunden wie die Grundsätze der Zweckbindung und Transparenz. „Privacy by Design" verlangt die datenschutzgerechte Konzeption von Big-Data-Analysen. Allerdings ist die Rechtmäßigkeit kein alleiniger Garant für einen gesellschaftlich akzeptablen Einsatz von Big Data. Denn auch die Verwendung anonymer Datensätze können Auswirkungen auf den Einzelnen haben; deshalb müssen darüber hinaus außerrechtliche, ethische und politische Grundsätze beachtet werden, um zu verhindern, dass der Mensch zum bloßen Objekt der Informationsverarbeitung wird (Berlin Group, Empfehlung 53; Helbing u. a., Digital-Manifest).

Literatur

Art. 29 Datenschutzgruppe, Opinion 03/2013 on purpose limitation v. 2.4.2013, WP 203

Dies., Stellungnahme 5/2014 zu Anonymisierungstechniken v. 10.4.2014, WP 216

Dies., Stellungnahme 06/2014 zum Begriff des berechtigten Interesses des für die Verarbeitung Verantwortlichen gemäß Artikel 7 der Richtlinie 95/46/EG v. 9.4.2014, WP 217

Dies., Statement on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU v. 19.9.2014, WP 221

Berlin Group: International Working Group on Data Protection in Telecommunications, Arbeitspapier zu Big Data und Datenschutz – Bedrohung der Grundsätze des Datenschutzes in Zeiten von Big-Data-Analysen, (2014)

A. Dix, Rede im Abgeordnetenhaus von Berlin am 24. September 2015, Plenarprotokoll 17/69, 7068 f.

N. Härting/J. Schneider, Das Ende des Datenschutzes – es lebe die Privatsphäre, Computer und Recht 2015, 819 ff.

D. Helbing/B. Frey/G. Gigerenzer/E. Hafen/M. Hagner/Y. Hofstetter/J. van den Hoven/R. Zicari/A. Zwitter, Digital-Manifest I u. II, Spektrum der Wissenschaft 1/2016, 51 ff.

Konferenz der Datenschutzbeauftragten des Bundes und der Länder, Big Data zur Gefahrenabwehr und Strafverfolgung: Risiken und Nebenwirkungen beachten, Entschließung v. 19. März 2015

H. Maas, Unsere digitalen Grundrechte, DIE ZEIT v. 10.12.2015, 9

V. Mayer-Schönberger/K.Cukier, „Big Data" – Die Revolution, die unser Leben verändern wird (2013)

Privacy Bridges: EU and US Privacy Experts in Search of Transatlantic Privacy Solutions (2015), https://privacybridges.mit.edu/

P. Richter, Datenschutz zwecklos? – Das Prinzip der Zweckbindung im Ratsentwurf der DSGVO, Datenschutz und Datensicherheit 2015, 735 ff.

S. Sädtler, Demokratische Willensbildung in der Cloud im Kontext von Big Data und Datenschutz, in: P. Richter (Hrsg.), Privatheit, Öffentlichkeit und demokratische Willensbildung in Zeiten von Big Data, Baden-Baden 2015, 69 ff.

S. Simitis (Hrsg.), Bundesdatenschutzgesetz, 8. Aufl., Baden-Baden 2014

M. Weber, Gesammelte Politische Schriften (1920)