*Typoskript i. S. d. § 38 Abs. 4 UrhG (Original veröffentlicht in DVBl. 2014, S. 1481-1489)*

*Mario Martini*[*]

# Big Data als Herausforderung für den Persönlichkeitsschutz und das Datenschutzrecht

**»Big Data« ist zum Inbegriff eines neuen Zeitalters der Informationsanalyse geworden. Der Trendbegriff hat den Nischendiskurs der technologischen Elite verlassen und ist zu einem Hype geworden. Manche sehen in dem Phänomen einen Garanten effizienten Ressourceneinsatzes, der das Schwungrad des Fortschritts antreibt, gar den Hoffnungsträger einer neuen industriellen Revolution und eine Schlüsseltechnologie, die ein Füllhorn von Annehmlichkeiten des täglichen Lebens über den Homo Digitalis ausschüttet. Doch Big Data hat eine Achillesferse: den Persönlichkeitsschutz. Die massenhafte Auswertung von Daten legt tiefe Einblicke in unseren digitalen Alter Ego frei. Manchem gilt Big Data daher als eine Zeitbombe, die das Datenschutzrecht bislang nicht zuverlässig entschärft. Der Beitrag gibt einen Überblick über die Gefährdungspotenziale und zeichnet die Konfliktlinien nach. Er leitet daraus ein Plädoyer für ein Algorithmenkontrollrecht ab.**

Die Science-Fiction von gestern ist häufig die Realität von morgen – dies gilt auch für den amerikanischen Film »Minority Report« aus dem Jahr 2002.[1] Tom Cruise spielt darin John Anderton, den Mitarbeiter einer außergewöhnlichen Washingtoner Polizeieinheit: der Abteilung Precrime. Deren Aufgabe ist es, auf der Grundlage von Wahrscheinlichkeitsurteilen sog. »Precogs« Verdächtige zu ermitteln und dadurch Straftaten zu verhindern. Im Washington des Jahres 2054, so die filmische Utopie, hat es deshalb sechs Jahre lang keinen Mord mehr gegeben. Doch das System frisst seine eigenen Kinder. Anderton gerät als Mitarbeiter der Abteilung bei einem Precrime-Screening selbst in den Verdacht, bald einen Mord zu begehen. Die Person, die er laut Vorhersage ermorden soll, wird tatsächlich wenig später tot aufgefunden – und Anderton verhaftet. Nur durch glückliche Fügungen gelingt es ihm, sich von dem Mordverdacht zu entlasten. Das vermeintliche Opfer hatte sich selbst umgebracht.

Die filmische Fiktion kommt der heutigen Wirklichkeit verblüffend nahe. *Predictive Policing* ist in den USA zu einem festen Bestandteil moderner Polizeiarbeit geworden. New Yorks *Real*

---

[*] Der Autor ist Lehrstuhlinhaber an der Deutschen Universität für Verwaltungswissenschaften Speyer. Der Beitrag ist aus einem Vortrag hervorgegangen, den er am 17.01.2014 im Rahmen des von Prof. Marion Albers veranstalteten Workshops »Vergessen im Internet« in Hamburg gehalten hat. Eine Langfassung des Beitrages erscheint in dem Tagungsband Hill/Martini/Wagner, Wie wir morgen leben wollen, 2015.

[1] Er geht auf eine Kurzgeschichte aus dem Jahr 1956 zurück; siehe P. K. Dick, The Collected Stories of Philipp K. Dick, Vol. 4: The Minority Report, 1956.

*Time Crime Center* entsendet auf der Grundlage einer Zusammenführung polizeilicher Datensätze Einsatzkräfte zu Gefahrenherden, die erhöhte Wahrscheinlichkeiten für Straftaten aufweisen.[2] Komplexe digitale Netzwerksysteme übernehmen die Arbeit der »Precogs«. Die Computeranalyse professionalisiert mithilfe feinmaschiger Datenrasterung den kriminalistischen Instinkt des Polizisten vor Ort.[3]

Auch in Deutschland hält die datenbasierte Unterstützung polizeilicher Tätigkeit Einzug: Das BKA nutzt seit dem Jahr 2000 die Software »Analyst's Notebook« für ihre Ermittlungstätigkeit.[4] Sie visualisiert Beziehungsgeflechte zwischen Personen und hilft bei der Verifizierung bzw. Falsifizierung von Hypothesen. Die Sicherheitsbehörden setzen damit auf einem allgemeinen Trend auf, der unsere Lebenswirklichkeit immer stärker durchdringt: »Big Data«. Welche Spannungen seine technologischen Möglichkeiten zum Schutz unseres Persönlichkeitsrechts und zu unserem Datenschutzrecht aufbauen, gehört zu den brisanten gesellschaftlichen und rechtlichen Fragen unserer Zeit. Der Beitrag stellt sich dieser Diskussion. In einer Tour d´Horizon erläutert er Wesen und Spielarten des Phänomens »Big Data« (I.) sowie sein Gefährdungspotenzial für das Persönlichkeitsrecht (II.), zeichnet seine Konfliktlinien mit zentralen Datenschutzprinzipien nach (III.) und analysiert die rechtlichen Grenzen moderner Big-Data-Praktiken (IV.), um daraus rechtspolitische Forderungen abzuleiten (V.).

## I. Wesen und Anwendungsbereiche von Big Data

Big Data[5] steht für eine disruptive technologische Entwicklung, die immer größere, heterogene Datenmengen immer schneller und immer tiefgliedriger auswertbar macht (1.) und dadurch einen Paradigmenwechsel in der Datenverarbeitung einläutet (2.). Das Phänomen versetzt Computersysteme in die Lage, Erfahrung und Intuition als Spezifika menschlicher Entscheidungen auf der Grundlage massendatenbasierter Wahrscheinlichkeitsurteile und Lernmechanismen ein Stück weit nachzubilden und dadurch künftiges Verhalten berechenbar zu machen.

### 1. Strukturmerkmale

Bisher war die professionelle Datenanalyse durch enge Begrenzungen des vorhandenen Datenvolumens, der Verarbeitungsgeschwindigkeit, der Möglichkeiten zur Datenzusammenführung sowie der Analysetechnologie limitiert. So nutzt die Menschheit Schätzungen zufolge bisher lediglich ca. 12 % der vorhandenen Datenmenge.[6] Big Data wird

---

2 Vgl. vertiefend C. Brücher, Rethink Big Data, 2013, S. 75 ff.; V. Mayer-Schönberger/K. Cukier, Big Data, 2013, S. 199 ff.

3 Vgl. dazu auch die Antwort der Bundesregierung auf die Kleine Anfrage der Fraktion DIE LINKE, BT-Drucks. 17/11582. Vorgaben für die automatisierte Datenverarbeitung des Informationsverwaltungssystems von Europol enthalten die Art. 14 ff. des Beschlusses 2009/371/JI des Rates vom 06.04.2009 zur Errichtung des Europäischen Polizeiamts, ABl. EG Nr. L 121/37 vom 15.05.2009.

4 BT-Drucks. 17/11582 (Fn. 3), S. 6.

5 Die Wortschöpfung knüpft an die seit den neunziger Jahren gebräuchlichen, mit ähnlicher Zielrichtung und Konnotation verwendeten Begriffe »Data Mining« und »Business Intelligence« an.

6 D. Bornemann, RDV 2013, 232 (233).

diese Quote substanziell erhöhen[7] und mit seinen technischen Möglichkeiten die bislang bestehenden Fesseln sprengen. Dafür sind vier Strukturmerkmale verantwortlich, die Big Data ausmachen: »volume«, »variety«, »velocity« und »analysis«.[8]

Nie zuvor standen so viele abrufbare Daten zur Verfügung (*volume*). Das Datenvolumen, welches allein das Internet zutage fördert, wächst jährlich um 50 %. Bis zum Jahr 2020 wird die Datenmenge voraussichtlich auf 40 bis 100 ZB ansteigen.[9] Neue Rechnergeschwindigkeiten machen dieses exponentiell wachsende Datenaufkommen nunmehr binnen eines Wimpernschlags analysierbar (*velocity*). Speichergröße und Verarbeitungsgeschwindigkeit verdoppeln sich entsprechend der Faustregel des Moore'schen Gesetzes gegenwärtig spätestens alle zwei Jahre.[10] Immer feinere Analysemethoden (*analysis*) ermöglichen eine immer tiefgliedrigere Durchdringung der Datenbestände. Sensorik, biometrische Erkennungsverfahren, Sentimentanalyse und Technologien der Linguistik sowie Semantik sind wichtige Katalysatoren dieser Entwicklung.[11] Während die Analyse heterogener, unstrukturierter Daten bislang entweder gar nicht oder nur um den Preis sehr ungünstiger Kosten-Nutzen-Relation möglich war, lassen sich nunmehr auch viele disparate Daten aus völlig unterschiedlichen Quellen und Kontexten – von Textbeiträgen aus sozialen Netzwerken über YouTube-Videos bis hin zu Standortdaten eines Smartphones – technisch leicht zu neuen Informationsgehalten zusammenführen (*variety*).[12]

Big-Data-Algorithmen erschließen auf dieser technischen Grundlage das Erkenntnispotenzial scheinbar wertloser Daten, das bislang brach lag, und überführen sie in eine nutzbare Auswertungsform. Sie durchkämmen Datenbestände auf Zusammenhänge, entwickeln aus den Korrelationen Muster und leiten auf dieser heuristischen Grundlage Handlungsempfehlungen und Schlussfolgerungen ab. Lernverfahren passen die Modellparameter den verfügbaren Datensätzen an.[13] Aus einer amorphen Datenmasse entsteht so durch Neustrukturierung neues Erkenntnispotenzial bislang unerkannter Korrelationen, das ihr durch intelligente Verknüpfung einen neuen Sinn einhaucht. Daraus

---

7 Für die Fähigkeit zur verantwortungsvollen und nachhaltigen Nutzung dieses Potenzials bürgert sich zusehends der Begriff »Datability« ein.

8 Vgl. zu diesen Kennzeichen von Big Data auch etwa C.-D. Ulmer, RDV 2013, 227 (227 f.); vgl. auch die Begriffsbeschreibungen bei T. Weichert, in: Geiselberger (Hrsg.), Big Data, 2013, S. 131 und 133; D. Offenhuber/C. Ratti, in: Geiselberger (Hrsg.), Big Data, 2013, S. 149 (153); Mayer-Schönberger/Cukier (Fn. 2), S. 13. Als viertes Merkmal findet sich statt »analysis« häufig »value«, teilweise auch »veracity« (Glaubwürdigkeit, Wahrhaftigkeit); vgl. z. B. R. Bachmann/G. Kemper/T. Gerzer, Big Data - Fluch oder Segen?, 2014, S. 28; BITKOM, Big Data im Praxiseinsatz, 2012, S. 7 und 19). Dabei handelt es sich lediglich um eine Beschreibung möglicher Einsatzzwecke, nicht aber um ein notwendiges Wesensmerkmal: Big-Data-Anwendungen sind zwar nur so gut, wie die Daten, die in den Analysekreislauf eingespeist werden. Eine Anwendung bleibt aber auch dann eine Big-Data-Analyse, wenn sie im Einzelfall keinen Mehrwert generiert oder die Zuverlässigkeit der Daten nicht gewährleistet ist.

9 Vgl. zu den unterschiedlichen Schätzungen Bornemann (Fn. 6), 232; BITKOM (Fn. 8), S. 12.

10 BITKOM (Fn. 8), S. 22.

11 Vgl. Ulmer (Fn. 8), 227 mit Fn. 3; BITKOM (Fn. 8), S. 27.

12 Die Umwandlung der unterschiedlichen Datenformate in einheitliche, maschinenlesbare Arbeitswerkzeuge, die den Verwendungskontext von Informationen (etwa die unterschiedliche Verwendung des Wortes »Schimmel«), im Idealfall auch mitschwingende Stimmungen (sog. Sentimentanalyse), mitzulesen vermögen, ist die Herausforderung und Kunst von Big-Data-Technologien. Vgl. T. F. Dapp, Big Data – die ungezähmte Macht, 2014, S. 8.

13 Dies ist auch der Grund dafür, weshalb die Informatik statt von »Big Data« vorzugsweise von »lernenden Systemen« bzw. »cognitive computing« spricht. Während »Big Data« stärker die Datenmenge fokussiert, stellt »cognitive computing« eher die Analysemethoden künstlicher Intelligenz in den Vordergrund, die Big-Data-Anwendungen möglich machen.

erwachsen im Idealfall evidenzbasierte Grundlagen für bessere Entscheidungen und Verhaltensprognosen.

Big Data ergänzt den Augenschein um Algorithmen, das Gedächtnis um Datenbanken und das Bauchgefühl um Statistik. Den Anspruch auf Exaktheit gibt die Analysetechnologie dabei zugunsten einer umfassenden Sammlung von Daten ein Stück weit auf.[14] Je mehr Daten zur Verfügung stehen, umso eher lassen sich nämlich Unschärfen in der Datenerhebung akzeptieren.[15] Dafür sind Big-Data-Analysen jedoch in der Lage, neue Zusammenhänge zu identifizieren, die dem herkömmlich Auslesenden verborgen bleiben – wie ein intelligenter Zuhörer, der zwischen den Zeilen lesen und Nuancen erkennen kann, ohne überdies etwas zu vergessen. Mehr Masse führt hier ausnahmsweise auch zu mehr Klasse.

## 2. Potenziale gesellschaftlichen und ökonomischen Mehrwerts

Ebenso vielfältig wie die analysierten Daten sind die Einsatzzwecke von Big-Data-Analysen. Ihre Bandbreite reicht von Bedarfsplanungen und Gestaltungsoptimierungen für öffentliche Daseinsvorsorgeaufgaben über effiziente Organisations- und Produktionsabläufe bis zur gezielten Kundenansprache und häuslichen Energieeinsparung.

Gerade mobile Anwendungen eröffnen bislang ungeahnte Möglichkeiten einer digitalen Auswertung unserer analogen Verhaltensspuren; Standortdaten sind das Datengold des 21. Jahrhunderts. In Smart Cities machen sie ein digitales Optimierungsmanagement der Verkehrs- und Versorgungsinfrastruktur als Baustein eines Mobile Government möglich, das Staus und Unfällen ebenso wie dem ineffizienten Einsatz von Umweltressourcen entgegenwirkt. Möglich werden maßgeschneiderte Marketing- und Vertriebsstrukturen mit minimalen Streuverlusten, insbesondere Location-based-Marketing, oder individualisierte Angebote, wie etwa Kfz-Versicherungstarife nach dem Modell »pay as you drive«. Medizinisches Monitoring und digitale Assistenzsysteme gehören bei der Pflege älterer Menschen ebenso wie bei der Lifestyle-Optimierung des modernen Managers zum Alltag der Zukunft. Fitness-tracking-Bänder, wie z. B. »fitbit«, machen den Anfang. Sie rechnen aus der Fülle von Sensordaten, die der Träger des Armbandes durch seine Bewegung auslöst, valide Bewegungsinformationen heraus. Dass sich Krankenversicherungstarife bald nach einem auf der Grundlage solcher »Wearables« ermittelten Health Score bemessen könnten, ist keine Utopie mehr.[16]

Wenn Gebrauchsgegenstände und Maschinen miteinander vernetzt sind, ebnet das einer selbstadaptiven Logistik des Alltags sowie einem vorbeugenden Instandhaltungsmanagement in der industriellen Produktion den Weg. Die Vernetzung der Maschinen und Anlagen von der Bestellung der Fertigungsbestandteile, über den Konstruktionsplan bis zur Auslieferung und zum Kundenservice über einen RFID-Chip perfektioniert die Automatisierung, Überwachung und Steuerung von industriellen

14 Vgl. Mayer-Schönberger/Cukier (Fn. 2), S. 179; das übersieht N. Leopold, vorgänge 2012, 74 (77).
15 Mayer-Schönberger/Cukier (Fn. 2), S. 21; BITKOM (Fn. 8), S. 27.
16 Zu weiteren Beispielen siehe etwa P. Welchering, RDV 2014, 11 ff.

Prozessabläufen. Darin liegt der Kern des Internets der Dinge – und damit der Code einer vierten industriellen Revolution, die einen Quantensprung neuer Fertigungsmethoden auslösen wird.

Big Data ist längst fester Bestandteil unseres Alltags. *Facebook* schlägt uns Freunde vor, *Amazon* empfiehlt uns auf der Grundlage verwandter Suchanfragen Produkte und *Google* unterbreitet mit seiner Autocomplete-Funktion massendatengenerierte Suchvorschläge. Die Empfehlungen sind häufig erstaunlich treffsicher. Auf dieser breiten Datengrundlage experimentieren Amazon & Co. bereits mit Geschäftsmodellen wie dem *anticipatory shopping*, welches Kunden bedarfsgerecht Waren zusendet, bevor sie diese überhaupt bestellt haben.

Gegenwärtig lässt sich nur erahnen, welche Wissensschätze und Anwendungsinnovationen sich im Einzelnen aus den Datenvorkommen bergen lassen. Klar scheint nur: Die digitale Datenalchemie bahnt sich unaufhaltsam ihren Weg. Eine digitale Goldgräberstimmung greift um sich.

## II. Gefährdungspotenzial

Doch Big Data ist nicht einfach nur Big Business. Spätestens seit den Enthüllungen von *Edward Snowden* hat Big Data seine Unschuld verloren. Das Phänomen ist nunmehr auch zum Inbegriff systematischer staatlicher Überwachung privater Lebensführung unbekannten Ausmaßes, einer neuen Form des Imperialismus, geworden, welche die Privatheit mithilfe digitaler Massenausforschungsfeldzüge zu kolonisieren droht.[17] In einer Welt digitaler Vernetzung und algorithmenbasierter Auswertung können Daten den Menschen im Wege technologischer Vermessung lesbar machen und sein Innerstes wie einen Code dechiffrieren, der den Zugang zu seinem Bau- und Verhaltensplan verschafft.[18] Das Abgreifen von Daten in Echtzeit und ihre Umwandlung in Analyse- und Planungssysteme zur Verhaltenskontrolle hat der Welt deutlich gemacht: So groß die Chancen sind, so groß ist auch das Missbrauchs- und Gefährdungspotenzial von Big Data.

Big-Data-Anwendungen zielen darauf, künftige Verhaltensmuster einer Person oder Personengruppe zu berechnen, sei es als Analyse der Wahrscheinlichkeit eines bestimmten vertraglichen Verhaltens (*Scoring*), als Akkumulierung inkonnexer Daten zu einem detailgetreuen digitalen Persönlichkeitsprofil (*Profiling*), sei es als Auswertung bestimmter Merkmale einer Person, etwa ihres Gesundheitszustands, ihrer persönlichen Vorlieben oder ihrer Zuverlässigkeit (*Personalizing*), oder als Verfolgung auf der Grundlage einer Spurenbildung (*Tracking*).

Die Deutschen stehen einer derartigen Ausforschung der engeren persönlichen Lebenssphäre durch eine staatliche Überwachungsmaschinerie oder eine digitale Wertschöpfungsmatrix, welche Wissen über Verhaltensmuster monetarisiert, besonders

---

17 F. Schirrmacher, in: Geiselberger (Hrsg.), Big Data, 2013, S. 273; Weichert (Fn. 8), S. 131; S. Zuboff, FAZ 13.02.2014, 33.

18 M. Mühl, FAZ 14.02.2014, 35.

sensibel gegenüber. Als bspw. das Telekommunikationsunternehmen O2 ankündigte, die Standortdaten seiner Mobilfunkkunden Dritten gegen Entgelt zur anonymisierten Auswertung für Marketingzwecke zugänglich zu machen, entfachte das einen Sturm der Entrüstung. Das Unternehmen musste seine Pläne sehr schnell zurücknehmen.[19] Ähnlich groß war die Empörung, als die SCHUFA bzw. der BND ankündigten, das Potenzial sozialer Netzwerke für die Bonitätsprüfung bzw. die nachrichtendienstliche Ermittlungstätigkeit auszuloten.[20]

Diese Sensibilität hat einen guten Grund: Der Schutz der Entfaltung von Privatheit ist die Funktionsbedingung selbstbestimmter Entwicklung der Persönlichkeit und des demokratischen Gemeinwesens.[21] Angriffe gegen den digitalen Zwilling richten sich gegen den Menschen selbst. Wer unter ständiger Beobachtung steht oder sich beobachtet glaubt, entfaltet sich nicht frei. Das diffuse Gefühl des Beobachtetwerdens[22] geht mit Einschüchterungseffekten (*chilling effects*) einher, die eine an die gesellschaftliche Verhaltenserwartung angepasste Selbstbeschränkung des Handelns auslösen können und damit diejenige Entfaltung konterkarieren, die das Persönlichkeitsrecht als Grundlage der Selbstbestimmung verbürgen soll. Freiheit und Unbefangenheit, Kreativität und Innovationskraft fallen dem zum Opfer. Denn diese implizieren die Abweichung von tradierten Verhaltensmustern.

## III. Konflikt mit den Prinzipien des deutschen Datenschutzrechts

Indem Big Data auf eine breite Datengrundlage angewiesen ist, um seinen Mehrwert entfalten zu können, tritt es in einen Zielkonflikt mit den Geboten der Datenvermeidung und Datensparsamkeit (1.), der Zweckbindung (2.) und der Transparenz (3.). Das fordert unser Datenschutzrecht heraus.

### 1. Prinzipien der Erforderlichkeit, Datenvermeidung und Datensparsamkeit

Datenschutz verlangt Datenaskese. Denn das Risiko für die informationelle Selbstbestimmung wächst proportional zur Menge der gespeicherten und verarbeiteten Daten. Die Datenverarbeitung ist entsprechend in ihrem Inhalt, Umfang und ihrer zeitlichen Erstreckung auf das zur Zweckerreichung Erforderliche zu begrenzen (vgl. insbesondere § 35 Abs. 2 S. 2 Nrn. 3 und 4 BDSG). Technische Datenverarbeitungssysteme sowie die Erhebung, Verarbeitung und Nutzung personenbezogener Daten sind daher so zu gestalten, dass sie ihre Funktion mit möglichst wenig personenbezogenen Daten erfüllen können (§ 3a S. 1 BDSG). Sobald Daten nicht mehr benötigt werden, sind sie zu löschen.

---

19 Vgl. Brücher (Fn. 2), S. 117 f.

20 http://www.spiegel.de/netzwelt/web/schufa-will-kreditdaten-bei-facebook-sammeln-a-837454.html (10.09.2014) bzw. www.sueddeutsche.de/digital/auslandsgeheimdienst-bnd-will-soziale-netzwerke-live-ausforschen-1.1979677 (10.9.2014).

21 Vgl. auch BVerfGE 65, 1 (43).

22 BVerfGE 125, 260 (320).

Die darin liegende Funktion des sozialen Vergessens läuft bei Big Data tendenziell leer.[23] In seiner Anwendungslogik gibt es kein unnützes Datum mehr. Sein Funktionsprinzip beruht vielmehr darauf, aus Datenmyriaden neue, bisher unbekannte Muster herauszufiltern. Aus der Perspektive einer Massendatenverwertung sind Daten immer erforderlich. Denn ihre Löschung erzeugt einen Informationsverlust, der aufzuspürende Korrelationsmuster möglicherweise unentdeckt lässt. Das setzt einen Anreiz, auch ältere Daten in möglichst großer Zahl und möglichst lange vorzuhalten. Eben dies aber kann Profile versteinern, die sich in der realen Welt bereits weiterentwickelt haben. Insoweit wirkt das Internet wie eine Asservatenkammer für Daten. Daten, die einmal den Weg hineingefunden haben, finden den Weg aus dem Informationspool nicht ohne Weiteres wieder heraus. In der realen Welt sind Geschehnisse den Gesetzen der Flüchtigkeit und der »Gnade des Vergessens« unterworfen. In der digitalen Welt lassen sich jedoch längst vergessene Momentaufnahmen zu neuen Mustern zusammensetzen und aus den zugrunde liegenden Mosaiksteinen Profile entwickeln. Sie machen den Einzelnen schnell zum Gefangenen seiner vergangenen Routinen. Erst wenn veraltete Datenbestände auch im Regime von Big Data aussortiert werden wie makulierte Bücher, liefern Analysen persönlichkeitsgerechte Abbilder der Realität.

## 2. Grundsatz der Zweckbindung

Daten dürfen nur für den Zweck genutzt werden, für den sie erhoben worden sind. Der Verarbeitungszweck begrenzt die Verarbeitungsbefugnis (§ 88 Abs. 3 Satz 2 TKG, § 12 Abs. 2 i. V. m. § 14 Abs. 1 und § 15 Abs. 1 TMG, § 28 Abs. 1 Satz 2, Abs. 2 und Abs. 5, §§ 31 und 39 BDSG, § 78 Abs. 1 Satz 1 SGB X; vgl. auch Art. 5 lit. b, 23 Nr. 2 Datenschutz-Grundverordnung-E). Das soll es dem Betroffenen ermöglichen, die Preisgabe von Daten entsprechend seiner sozialen Rolle und seinem Selbstverständnis autonom zu steuern.

Big Data löst Daten aber aus ihrem ursprünglichen Verwendungskontext heraus und stellt sie bewusst in andere Zusammenhänge. Der Analyseansatz zielt darauf, heterogene Daten zieloffen zu screenen und miteinander zu verschneiden, um bislang unerkannte Korrelationen hervorzubringen.[24] Enge Zweckbegrenzungen laufen dieser Mission zuwider;[25] Big Data bevorratet Daten zu ergebnisoffenen, also unbestimmten Zwecken. Der Analyseprozess liefert Antworten auf Fragen, die bisher gar nicht gestellt wurden. Erst am Ende des Veredelungsvorgangs steht dann fest, welche genaue Zielsetzung der Prozess hatte. Die damit einhergehende Neutralisierung des ursprünglichen Erhebungs- und Verwendungszwecks entfernt sich von dem Grundgedanken des Datenschutzrechts, den Einzelnen darüber im Klaren zu lassen, zu welchen Zwecken die Verarbeitung von Daten erfolgt.

---

23 A. Roßnagel, ZD 2013, 562 (564).
24 Bornemann (Fn. 6), 235.
25 Roßnagel (Fn. 23), 564; BITKOM (Fn. 8), S. 26.

### 3. Transparenz

Das Datenschutzrecht ist von dem Gedanken der Transparenz durchdrungen. Für den Betroffenen muss nachvollziehbar sein, welche Stellen welche Daten zu welchem Zweck und in welchem Umfang sammeln.[26] Um das sicherzustellen, gesteht das BDSG ihm insbesondere eigene Informations-, Mitteilungs- und Auskunftsansprüche zu (§ 4 Abs. 3, §§ 33 ff. BDSG). Individuelle Entfaltung kann nämlich nur dann gelingen, wenn der Einzelne die Folgen einer Preisgabe von Daten überblicken kann und seine selbstbestimmte Hoheit über die Verwendung seiner Daten immer aufrechterhalten bleibt.[27]

Big-Data-Analysen verwehren dem Einzelnen demgegenüber die Möglichkeit, zu antizipieren, welche Schlüsse aus welchen Daten in bestimmten Kontexten gezogen werden. Die Analyse-Algorithmen gehören zu den am besten gehüteten Geheimnissen von Geheimdiensten und Unternehmen gleichermaßen. Den Betroffenen bleibt nicht nur unklar, welche Daten, sondern auch mit welcher Analysemethode die Daten in den Prozess eingehen. Sie können nicht mehr nachvollziehen, wie das Auswertungsergebnis zustande kommt; es bleibt für sie eine Blackbox. Die Nachvollziehbarkeit der Verfahren ist freilich zentrale Voraussetzung für das Vertrauen darin, keinen willkürlichen oder strukturell diskriminierenden Entscheidungen ausgesetzt zu sein.

### 4. Zwischenfazit

Die Prinzipien des Datenschutzrechts erscheinen in einer Big-Data-Welt geradezu anachronistisch wie ein Relikt aus Zeiten von Lochkarten und Formularbögen. Dabei sind sie – adjustiert an die Herausforderungen des digitalen Zeitalters – zum Schutz der Persönlichkeit aktueller und notwendiger denn je. Denn erst heute werden aufgrund der fortschreitenden technischen Möglichkeiten die Bedrohungen der Persönlichkeit in besonderer Weise real, die abzuwehren die Prinzipien gedacht sind.

## IV. Datenschutzrechtliche Zulässigkeitstatbestände

Nicht allein auf der Ebene seiner Prinzipien, sondern auch in seinen Einzelbestimmungen steht das Datenschutzrecht den Möglichkeiten von Big Data nicht sprachlos gegenüber. Für einzelne Anwendungsformen hält es bereits sehr konkrete Teilregelungen vor. Das gilt insbesondere für das Profiling und Personalizing (1. a) sowie das Scoring (1. b). Auch die allgemeinen Zulässigkeitsregelungen zur Datenverarbeitung setzen Big-Data-Analysen Schranken (2.). Als Königsweg ihres Einsatzes erweist sich eine Anonymisierung der Daten (3.).

---

[26] Vgl. Weichert (Fn. 8), 141.
[27] Roßnagel (Fn. 23), 563.

## 1. Big-Data-spezifische Regelungstatbestände

### a) Profiling und Personalizing

Für das Profiling und Personalizing trifft das BDSG in § 6a ansatzweise eine Regelung[28] – noch deutlicher und prominenter Art. 20 der geplanten Datenschutz-Grundverordnung der Europäischen Union.[29] Beide verbieten Profiling zwar nicht generell.[30] Ausschließlich automatisiert ablaufende Personenbewertungen erklären sie aber grundsätzlich für unzulässig, soweit sie – wie z. B. ein Vertragsschluss – für den Betroffenen eine rechtliche Folge nach sich ziehen oder ihn – wie z. B. eine negative Bewerberauswahl – erheblich beeinträchtigen. Eine solche gesetzliche Grenzziehung ist auch geboten. Sachgerechter Persönlichkeitsschutz bewahrt den Einzelnen vor einem Bewertungsautomatismus. Denn Big-Data-Analysen ergründen Korrelationen, nicht aber Kausalitäten.[31] Es braucht immer eine Ergänzung durch die Deutungskunst und -hoheit menschlicher Analyse. Algorithmen haben natürliche Erkenntnisgrenzen. Sie erkennen zwar mit unbestechlicher Genauigkeit Wahrscheinlichkeiten, Intuition aber ist ihnen fremd. Sie fördern nicht nur mitunter Zufallskorrelationen zutage, wie etwa den – unter Börsianern bekannten – Zusammenhang zwischen dem Ausgang des Super-Bowl-Endspiels und dem Anstieg von Börsenkursen.[32] Algorithmen verleiten auch schnell zu »*cum hoc ergo propter hoc*«-Fehlschlüssen: So lässt sich etwa zwischen der Storchenpopulation und der Geburtenrate in Niedersachsen zwischen 1972 und 1985 eine statistisch valide Korrelation ausmachen. Ein Big-Data-Algorithmus würde auf dieser Datengrundlage womöglich die Ansiedlung von Störchen in Niedersachsen vorschlagen, um die menschliche Geburtenrate zu erhöhen. Darin läge eine paradoxe Verwechslung von Ursache und Wirkung: Die höhere Geburtenrate in storchenreichen Gegenden ist die Folge des Umstandes, dass Störche typischerweise dort leben, wo die Geburtenrate wegen des örtlichen Umfeldes höher ist, nämlich auf dem Land.
Der Ursachenzusammenhang besteht also in der Urbanisierung, nicht in der Storchendichte. Auch der komplexeste Algorithmus vermag den Menschen in seiner komplexen Struktur und Eigenwilligkeit autonomer Steuerung in Freiheit und eigener Verantwortung nicht zu erfassen. Algorithmen sind nicht zuletzt fehleranfällig. Anlasslose Kursstürze, die Computeralgorithmen des Hochfrequenzhandels im Gefolge automatisierter Kauf- bzw. Verkaufsentscheidungen als Kettenreaktionen an der Börse auslösen, machen das deutlich.[33] Das erzeugt ein Bedürfnis nach staatlicher Algorithmenkontrolle – sowohl zum Schutz des Einzelnen als auch der Funktionsfähigkeit der gesellschaftlichen Ordnung.

---

28 Dazu etwa S. Golla, PinG 2014, 61 ff.

29 Zur Unzulässigkeit, die gesamte Persönlichkeit eines Menschen zwangsweise staatlich zu registrieren und zu katalogisieren, bereits BVerfGE 27, 1 (6).

30 Für den Entwurf zur EU-Datenschutz-Grundverordnung ergibt sich das auch aus Erwägungsgrund Nr. 58.

31 Mayer-Schönberger/Cukier (Fn. 2), S. 248.

32 Vgl. etwa G. W. Kester, The Journal of Investing 2010, 82 ff.

33 Vgl. auch Mayer-Schönberger/Cukier (Fn. 2), S. 203.

### b) Scoring

Einen besonders sensiblen Unterfall der Profilbildung bildet das Scoring.[34] Es verdichtet einzelne Merkmalsvektoren zu statistisch begründeten Prognosen künftigen Verhaltens. Die Analyse der Wahrscheinlichkeit, etwa eines Kreditausfalls, schließt vom Gruppenverhalten auf das Verhalten einzelner Merkmalsträger. Welche Faktoren das sind, bleibt häufig unklar. Deshalb zieht § 28b Nr. 1 BDSG[35] der Analysemethode ausdrücklich eine Grenze: Es muss sich um ein wissenschaftlich anerkanntes mathematisch-statistisches Verfahren handeln, das nachweisbar für die Berechnung der Wahrscheinlichkeit eines Verhaltens relevant ist. Diese normative Forderung und der Grundgedanke des Persönlichkeitsschutzes insinuieren, dass Betroffene die Dechiffrierung des Verfahrens zur Ergebnisermittlung verlangen können.

Mit eben diesem Begehren trat denn auch eine Bürgerin an die SCHUFA heran und verlangte von der Wirtschaftsauskunftei auf der Grundlage des § 34 Abs. 4 BDSG gerichtlich die Offenlegung des Algorithmus. Nach dem Willen des Gesetzgebers unterliegt dieser aber dem Geheimnisschutz als vorrangigem Schutzgut.[36] Dem Auskunftsinteresse Betroffener sieht der BGH entsprechend genüge getan, wenn die SCHUFA die in die Wahrscheinlichkeitsberechnung konkret eingeflossenen personenbezogenen Daten mitteilt.[37] Zwar erstreckt § 6a Abs. 3 BDSG den Auskunftsanspruch auch auf den »logischen Aufbau der automatisierten Verarbeitung« ihn betreffender Daten (dessen Anwendung ist auch nicht durch § 28b BDSG gesperrt).[38] Der gesetzliche Auskunftsanspruch bezieht sich aber nur auf die tragenden Funktionsprinzipien und grundlegenden Entscheidungsmaßstäbe, insbesondere ihre Wertigkeit. Gleiches gilt nach § 34 Abs. 4 Satz 1 Nr. 4 BDSG: Der Betroffene soll nachvollziehen können, in welcher Weise die Entscheidung zustande kommt und welche Faktoren in sie eingeflossen sind.[39] Der Quellcode ist davon demgegenüber nicht umfasst.[40] Der Betroffene kann seine Offenlegung nicht erzwingen.[41]

Rechtspolitisch gibt das Betroffenen Steine statt Brot. Sie können eine Überprüfung der Sachgerechtigkeit und Diskriminierungsfreiheit von Algorithmen nicht erreichen. Insbesondere wird das dem Gebot sachgerechten Persönlichkeitsschutzes nicht gerecht. Zwar braucht und verdient die in Algorithmen steckende geistige Leistung einen Investitionsschutz; die Pflicht zur Offenlegung der Berechnungsformel kann die technologische Innovationsfreude ersticken. Sachgerechter Persönlichkeitsschutz setzt aber umgekehrt eine Kontrolle solcher Algorithmen voraus, soweit sie sensible

---

34 Zu der Pflicht, vor Abschluss von Verbraucherkreditverträgen die Kreditwürdigkeit »anhand ausreichender Informationen« zu bewerten, vgl. Art. 8 Abs. 1 der Richtlinie 2008/48.

35 Für den Bereich des Kreditwesens, namentlich den Umgang mit personenbezogenen Daten im Zusammenhang mit Risikobemessungsverfahren, ist § 10 Abs. 1 Satz 3-8 KWG lex specialis. Vgl. dazu auch BT-Drucks. 16/1335, S. 48.

36 BT-Drucks. 16/10529, S. 17: »(…) nicht die Scoreformel, an deren Geheimhaltung die Unternehmen ein überwiegendes schutzwürdiges Interesse haben«.

37 BGH, NJW 2014, 1235 (1235).

38 Vgl. BT-Drucks. 14/4329, S. 37.

39 Zur umstrittenen Reichweite des Anspruchs im Einzelnen siehe BGH, NJW 2014, 1235 (1237) m. w. N.

40 BT-Drucks. 14/4329, S. 38; Erwägungsgrund Nr. 41 der EG-Datenschutzrichtlinie; in diesem Sinne auch P. Scholz, in: Simitis (Hrsg.), BDSG, 8. Aufl. 2014, § 6a Rn. 40.

41 BGH, NJW 2014, 1235 ff.

Persönlichkeitsmerkmale oder grundlegende Lebensentfaltungschancen berühren. Denn Algorithmen steuern in Zukunft unser Leben. Sie entscheiden, welche Angebote wir erhalten, was wir lesen, mit wem wir in Kontakt treten. Das macht sie diskriminierungsanfällig. Deshalb ist nicht nur Betroffenen ein Einblick in die grundsätzliche Gewichtung der in das Scoring einfließenden Faktoren zu gewähren, sondern ist auch eine behördliche Algorithmenkontrolle jedenfalls in den Fällen geboten, die sensible Auswirkungen in persönlichkeitssensitiven Bereichen zeitigen. Ebenso wie das Vertrauen, dass etwa Google bei seinen Suchalgorithmen seine eigenen Angebote gegenüber denjenigen von Konkurrenten absolut gleichbehandelt, nicht ohne Weiteres gerechtfertigt ist, ist nicht ohne Weiteres sichergestellt, dass die Kriterien der SCHUFA entscheidungsrelevant sind und nicht – direkt oder indirekt – unzulässig diskriminieren. Um den erforderlichen Geheimnisschutz zu verbürgen, ist eine Ausgestaltung als In-camera-Kontrolle sachgerecht. Die (technisch entsprechend zu rüstenden) Datenschutzbehörden erlangen dann als unabhängige Kontrollinstanzen exklusiven Einblick in die Algorithmenstruktur, um deren Vereinbarkeit mit dem Datenschutzrecht überprüfen zu können. Grundsätzlich reicht insoweit eine Befugnis zur repressiven Ex-post-Kontrolle. Nur in besonders sensitiven Bereichen, etwa Analysen religiöser Überzeugungen oder der Gesundheit (vgl. § 3 Abs. 9 BDSG), ist eine Ex-ante-Kontrolle als Kontrollmechanismus angezeigt. Dass ein solches Überwachungsregime nicht leicht realisierbar ist, entbindet den Staat nicht von seiner Verantwortung, wirksamen Persönlichkeitsschutz sicherzustellen.

### 2. Allgemeine Zulässigkeitstatbestände

Jenseits der spezifischen Tatbestände der §§ 6a und 28b BDSG richtet sich die Zulässigkeit von Big-Data-Anwendungen nach den allgemeinen Regeln des BDSG bzw. der jeweiligen leges speciales. Es gilt das allgemeine Verbotsprinzip des Datenschutzrechts. Eine Verarbeitung personenbezogener Daten bedarf also entweder einer informierten Einwilligung oder einer gesetzlichen Verarbeitungserlaubnis (§ 4 Abs. 1 BDSG, § 12 Abs. 1 TMG).

#### a) Einwilligung

Wirksam einwilligen kann man nur in eine Verarbeitung, deren Reichweite man überblicken kann. Eine Einwilligung in Big-Data-Analysen muss den Analysezweck daher hinreichend klar benennen. Allgemeine Zweckangaben, wie z. B. »zum Zwecke der Werbung«, sind zu unbestimmt.[42] Big Data lebt als Phänomen aber gerade davon, dass Daten für vorher nicht ausdrücklich bestimmte Zwecke genutzt werden sollen.

[42] So auch Roßnagel (Fn. 23), 564; vgl. auch LG Berlin, NJW 2013, 2605 (2606 f.).

De lege ferenda sollte sich die Einwilligung explizit auch auf die Zusammenführung unterschiedlicher Daten und ihre Auswertung mithilfe von Algorithmen beziehen. Einwilligungen sollten für Big-Data-Analysen auch grundsätzlich nur zeitlich begrenzt möglich sein. So lässt sich dem Nutzer der erhöhte Gefährdungsgrad bewusst machen und dem Risiko der Entkopplung von aktuellen Persönlichkeitspräferenzen und Einstellungen begegnen.

### b) Gesetzliche Verarbeitungsbefugnis

Auch der Spielraum für gesetzliche Verarbeitungserlaubnisse ist de lege lata eng gesteckt. Das gilt sowohl für Standort- (aa), Nutzungs- (bb) als auch für Inhaltsdaten (cc).

#### AA) STANDORTDATEN

Für Standortdaten, etwa das Bewegungsprofil von Smartphones, limitieren die §§ 96 und 98 TKG die Verarbeitungszwecke. Sie begrenzen diese (jenseits von Einwilligungserklärungen) auf die Zielsetzung der Nachrichtenübermittlung sowie deren Abrechnung (§ 96 Abs. 1 Satz 1 und 2 TKG) und die Erbringung eines angeforderten Dienstes mit Zusatznutzen, etwa eines Lokalisationsdienstes (§ 98 Abs. 1 Satz 1 TKG). Eine allgemeine und breite personenbezogene Datenauswertung, etwa für Zwecke der Werbung oder des Direktmarketings, ist unzulässig.[43]

#### BB) TELEMEDIENRECHTLICHE NUTZUNGS- UND BESTANSDATEN

Vordergründig mehr Möglichkeiten gesteht das TMG den Telemedienanbietern, etwa Facebook, Google & Co. zu. Sie dürfen für Zwecke der Werbung und bedarfsgerechten Gestaltung der Telemedien pseudonymisierte Nutzungsprofile erstellen (§ 15 Abs. 3 Satz 1 TMG).[44] Analysieren dürfen sie etwa, wann der Nutzer die Webseite normalerweise aufruft, welche Angebote er auf der Grundlage von Empfehlungen angeklickt hat und in welchem Preissegment sich seine Kaufentscheidungen bewegen. Zulässig sind aber nur (pseudonymisierte) *Nutzungs*profile, keine übergreifenden *Persönlichkeits*profile (§ 15 Abs. 3 TMG).[45] Soweit der Anbieter personenbezogene Daten erhoben hat, sind sie nach Ablauf des Zugriffs zu löschen (§ 13 Abs. 4 Satz 1 Nr. 2 TMG). Eine Zusammenführung mit anderen Daten über den Nutzer – deren Sammlung und Verknüpfung Big Data konzeptionell immanent ist – ist ausdrücklich untersagt (§ 15 Abs. 3 Satz 3 TMG).

#### CC) INHALTSDATEN

##### (1) Für eigene Geschäftszwecke – § 28 BDSG

---

43 Ulmer (Fn. 8), 229; M. Martini/W. Weiß/J. Ziekow, Rechtliche Zulässigkeit flächendeckender Alarmierungen der Bevölkerung in Katastrophenfällen per SMS (KatWarn), 2013, S. 76 ff. und 95 ff.

44 Zu Cookies siehe etwa M. Martini, in: Hill/Schliesky (Hrsg.), Neubestimmung der Privatheit, 2014, S. 193 (229 ff.).

45 A. Dix/P. Schaar, in: Roßnagel (Hrsg.), Beck'scher Kommentar zum Recht der Telemediendienste, 2013, § 15, Rn. 62.

Für Inhaltsdaten, also den Inhalt der Kommunikation,[46] bestimmt sich die Zulässigkeit der Verarbeitung für eigene Geschäftszwecke grundsätzlich nach § 28 BDSG. Es kommt dann entscheidend darauf an, ob das Auswertungsinteresse gegenüber dem schutzwürdigen Interesse des Betroffenen überwiegt. Je heterogener die Daten, je systematischer und umfänglicher die Auswertung, umso eher überwiegen die Interessen des Betroffenen.[47] Das gilt insbesondere bei Auswertungen, die auf ein Persönlichkeitsprofil mit rechtlichen Folgen oder erheblichem Beeinträchtigungspotenzial gerichtet sind (§ 6a BDSG), sowie bei besonderen Arten personenbezogener Daten (§ 28 Abs. 6 i. V. m. § 3 Abs. 9 BDSG), wie etwa der ethnischen Herkunft oder der politischen Meinung.[48] Ein überwiegendes Auswertungsinteresse vermutet das Gesetz demgegenüber grundsätzlich bei allgemein zugänglichen Daten (§ 28 Abs. 1 Nr. 3, Abs. 6 Nr. 2 i. V. m. § 10 Abs. 5 Satz 2 BDSG).[49] Darunter fallen viele klassische Big-Data-Internetquellen, etwa Tweets ebenso wie öffentliche Daten der Facebook-Profile.[50]
Der automatisierten Verarbeitung kann der Betroffene widersprechen – allgemein nach § 35 Abs. 4 BDSG bzw. speziell der Datennutzung für Zwecke der Werbung oder Marktforschung nach § 28 Abs. 4 Satz 1 BDSG. Das Widerspruchsrecht nach § 35 Abs. 4 BDSG ist jedoch zwei Hürden ausgesetzt: Das schutzwürdige Löschungsinteresse des Betroffenen muss zum einen gerade wegen seiner besonderen persönlichen Situation überwiegen (§ 35 Abs. 5 Satz 1 BDSG). Zum anderen muss er von der Verarbeitung überhaupt erfahren. An beidem wird es bei Big-Data-Analysen regelmäßig mangeln. Der Löschungsanspruch des § 35 Abs. 5 BDSG ist insoweit für den Schutz gegen Big-Data-Analysen ein relativ stumpfes Schwert.

Strenger ist hingegen – in der Deutung des EuGH – das Unionsrecht.[51] Er verlangt von Suchmaschinenbetreibern als Big-Data-Kollektoren auf Antrag die Entfernung solcher Einträge aus ihren Ergebnislisten, die den Verarbeitungszwecken aufgrund Zeitablaufs nicht mehr entsprechen oder dafür nicht oder nicht mehr erheblich sind bzw. darüber hinausgehen.[52] Der EuGH legt damit die sensible Entscheidung über das Rangverhältnis zwischen dem Informationsinteresse der Allgemeinheit und dem kollidierenden Persönlichkeitsschutz des Einzelnen in die Verantwortung des privaten Diensteanbieters. Das ist zwar im Äußerungsrecht nichts gänzlich Ungewöhnliches. Die Konfliktentscheidung ist bei dem Suchmaschinenbetreiber allerdings nicht in den richtigen Händen. Denn zum einen ist

---

46 Das TKG regelt die Anforderungen an die technische Ebene der Signalübertragung, das TMG und das BDSG demgegenüber die inhaltliche Ebene der Kommunikation. Das TMG nimmt dabei die Beziehung zwischen dem Nutzer und dem Telemedienanbieter in den Blick, das BDSG demgegenüber den Inhalt der Kommunikation, den der Telemediendienste übermittelt.

47 S. Venzke-Caprese, DuD 2013, 775 (779).

48 T. Weichert, in: Däubler/Weichert/Wedde (Hrsg.), BDSG, 2014, § 28 Rn. 52, 55.

49 Die Privilegierung beschränkt sich dann aber auf die aus der Quelle stammenden Daten, wie sie in ihrem dort veröffentlichten Zustand vorgelegen haben. Die Verknüpfung dieser Daten mit neuen Informationen erzeugt neue Daten, die nicht mehr allgemein zugänglich sind und nicht mehr den Schutz der Privilegierung genießen. Siehe P. Richter, DÖV 2013, 961 (964); P. Gola/R. Schomerus, in: Gola/Schomerus (Hrsg.), BDSG, 11. Aufl. 2012, § 28, Rn. 31.

50 Venzke-Caprese (Fn. 47), S. 776.

51 EuGH, NJW 2014, 2257 ff.; dazu etwa N. Nolte, NJW 2014, 2238 ff.

52 EuGH, NJW 2014, 2257 (2264, Rn. 93 ff.).

er (entsprechend dem Gebot effizienter Konfliktlösung) im Verhältnis zum Äußernden nur subsidiär verantwortlich. Zum anderen ist er als Intermediär weder dazu bestimmt noch dazu legitimiert, solche komplexen, für die demokratischen Grundlagen einer Gesellschaft wesentlichen Abwägungsentscheidungen als Clearingstelle der digitalen Identität unbefangen und sachgerecht zu treffen. Die Abwägung ist in der Entscheidungsmacht einer plural zusammengesetzten, die konfligierenden Interessen abbildenden Schlichtungsstelle sachgerechter verortet.

Zum Schutz des Einzelnen davor, an überholten Persönlichkeitsmustern festgehalten zu werden, sollte der Gesetzgeber darüber hinaus einer Auswertung zu Big-Data-Zwecken entweder eine zeitliche Verwertungsschranke ziehen oder der Verarbeitung älterer Daten einen abdiskontierten Gewichtungsfaktor auferlegen: Daten, deren Generierung so lange zurückliegt, dass ihre Aussagekraft durch den Zeitablauf gefährdet ist – denken lässt sich (in Analogie zu § 12 Abs. 3 BVerfSchG oder § 35 Abs. 2 Satz 1 Nr. 4 BDSG) an einen Zeitraum von z. B. 3-5 Jahren –, sollten grundsätzlich keine oder nur noch eine schwächer gewichtete Verwendung für Big-Data-Anwendungen finden dürfen.[53] Technisch absichern lässt sich dies durch entsprechende Schlüssel als Metadaten, welche den Taten als zweckbestimmende »tags« mit auf den Weg gegeben werden. Das wirkt einer Abkopplung der Big-Data-Auswertung von der Realität eines Persönlichkeitsbildes entgegen.

### (2) Datenverarbeitung öffentlicher Stellen

Noch größeren Einschränkungen der Verarbeitungsbefugnis als Private sind öffentliche Stellen unterworfen. Für sie hat der Grundsatz der Zweckbindung besondere verfassungsrechtliche Bedeutung: Er begrenzt die Verarbeitung auf den jeweils eigenen gesetzlichen Aufgabenbereich. Einem allgemeinen, zuständigkeitsüberschreitenden Informationsaustausch, der jegliche Informationsgrenzen abbaut, schiebt das einen Riegel vor.[54] Die föderale und fachliche Aufgliederung von Behörden entfaltet insoweit eine grundrechtliche Nebenwirkung.[55] Daraus erwächst ein informationelles Trennungsprinzip etwa für den Austausch zwischen offen arbeitenden Polizeibehörden und verdeckt arbeitenden Nachrichtendiensten: Der Verfassungsschutz darf sich gewünschte Informationen nicht auf dem Umweg über die Polizei beschaffen.[56] In Übereinstimmung damit sehen die Landesdatenschutzgesetze bereits ausdrücklich die Pflicht vor, zu unterschiedlichen Zwecken erhobene Daten getrennt zu verarbeiten.[57] Für öffentliche Stellen ist der Einsatz von Big-Data-Modellen daher nachhaltig eingeschränkt.[58]

---

53 Etwas anderes gilt für öffentlich zugängliche Daten. Deren Zugänglichkeit ggf. zu beseitigen, wenn sich relevante Änderungen ergeben haben, fällt in den Verantwortungsbereich eines jeden Einzelnen.

54 BVerfGE 133, 277 (321, Rn. 106).

55 BVerfGE 133, 277 (323, Rn. 113).

56 BVerfGE 133, 277 (329, Rn. 123). Zum Trennungsgebot zwischen Polizei und Geheimdiensten bereits C. Gusy, Die Verwaltung 1991, 467 (476); H. P. Bull, PinG 2013, 1 (6 f.); C. Streiß, Das Trennungsgebot zwischen Polizei und Nachrichtendiensten, 2011, S. 153 ff., 178 ff.

57 Vgl. etwa § 5 Abs. 5 BlnDSG, § 7 Abs. 4 Satz 2 Nr. 8 BremDSG, § 9 Abs. 2 Satz 2 Nr. 8 LDSG RhPf.

58 Strenger noch T. Weichert, ZD 2013, 251 (254).

### 3. Anonymisierung als Königsweg

Für zahlreiche Big-Data-Anwendungen bietet das Datenschutzrecht gegenwärtig keine gesetzliche Verarbeitungsgrundlage. Es ermöglicht ihren Einsatz dann nur in zwei Gestaltungsformen: bei der Verarbeitung allein gerätebezogener Sachdaten, wie etwa der Werkleistung einer Maschine mit RFID-Chip (soweit sie keine Rückschlüsse auf Personen zulassen), oder im Wege einer Auflösung des Personenbezugs durch Anonymisierung (§ 3 Abs. 6 BDSG). Diese Datenverarbeitungen fallen aus dem Regime des Datenschutzrechts heraus.

Anonymisierungsmethoden erweisen sich insofern datenschutzrechtlich als Königsweg des Einsatzes von Big-Data-Techniken.[59] Sie müssen allerdings zuverlässig sicherstellen, dass eine Zuordnung zu einer konkreten Person mit nach menschlichem Ermessen hinreichender Wahrscheinlichkeit ausgeschlossen ist. Das wird unter Big-Data-Bedingungen immer neuralgischer.[60] Auch anonymisierte Daten können nämlich unter Umständen durch massenhafte Big-Data-Verknüpfung Rückschlüsse auf konkrete Personen zulassen und Lebensentwürfe vermessbar machen. Je umfangreicher und detaillierter die Merkmalsdaten sind, je länger Bewegungsmuster gespeichert werden, umso höher ist die Wahrscheinlichkeit, dass ein Abgleich der Merkmalsdaten eine Reidentifizierung ermöglicht.[61] Häufig genügen insoweit wenige Merkmale,[62] gerade wenn die Aggregierung etwa auf politischen, religiösen oder ethnischen Merkmalen aufsetzt.[63]

Ein Weg hinreichenden Reidentifizierungsschutzes kann in einem Austausch des Anonymisierungsschlüssels nach kurzen Zeitabständen bestehen,[64] beispielsweise für Standortdaten durch kurzfristige Vergabe einer Einweg-Hash-ID, die – entsprechend der Wortbedeutung »hash« – Daten »verstreut« bzw. »zerhackt«.[65] Denkbar sind auch Methoden einer *Differential Privacy*: Datenbankabfragen und -antworten passieren dann einen Filter, der Unschärfen hinzufügt, um einen Rückschluss auf bestimmte Personen auszuschließen.[66] Ob solche Methoden den Personenbezug sicher auflösen und es dadurch gelingt, die Nutzungschancen der Big-Data-Technologie mit dem Persönlichkeitsschutz Betroffener zu versöhnen, stellt die Weichen für die Daten»spende«bereitschaft Betroffener sowie das darauf aufbauende gesellschaftliche und ökonomische Entfaltungspotenzial.

## V. Regulierungsstrategien de lege ferenda: Big-Data-Kontrolle als Algorithmen-Kontrolle

---

59 In diesem Sinne auch Ulmer (Fn. 8), 229; die anonymisierte Verwendung bereits erhobener Daten ist auch kein neuer Verwendungszweck im Sinne des Gesetzes, der einer eigenen Verarbeitungsgrundlage bedürfte.

60 Vgl. auch G. Baum, DuD 2013, 583.

61 Weichert (Fn. 58), 258; M. Wójtowicz, PinG 2013, 65 (67).

62 Vgl. mit Beispielen Bornemann (Fn. 6), 233; P. Katko/A. Babaei-Beigi, MMR 2014, 360 (361 f.).

63 Bedenklich ist daher aus deutscher Perspektive der Plan Großbritanniens, eine Online-Datenbank aufzubauen, welche Informationen zu den Krankheiten der Briten in pseudonymisierter Form vorhält, die Versicherungs- sowie Pharma-Unternehmen zum Ankauf zur Verfügung stehen. Vgl. FAZ 21.01.2014, 9.

64 Ulmer (Fn. 8), 330.

65 In einen Zeitabstand von 90 Minuten tauscht das System den jeweiligen Schlüssel gegen einen neuen aus. Vgl. dazu Ulmer (Fn. 8), 230.

66 Weichert (Fn. 58), 258.

Big Data gilt als Zeitenwende unserer Datenverarbeitung. Der Megatrend verheißt, bisher brach liegende Datenrebstöcke abzuernten, um daraus ein wertvolles Cuvée zu komponieren. Die Analyse liefert die Erntemaschine, welche die mühselige und kostspielige Handarbeit der Datenlese durch intelligente Algorithmen ersetzt.[67] Zugleich arbeitet sie etwas ungenauer als der »Arbeiter im Weinberg«, nimmt (angesichts der großen Verarbeitungsmenge) namentlich Unschärfen in Kauf und begnügt sich mit Korrelations- statt Ursachenforschung.

Vollständig neu sind ihre Möglichkeiten nicht. Big Data ist weniger eine Revolution als eine Evolution, die eine neue Stufe erklimmt. Umgekehrt handelt es sich aber auch nicht lediglich um alten Wein in neuen Schläuchen. Die Entwicklung birgt enormes Wertschöpfungs- und gesellschaftliches Problemlösungspotenzial. Wenn Daten der Rohstoff des 21. Jahrhunderts sind, dann wird die Qualität und Schnelligkeit der Datenverarbeitung zu einem zentralen Produktionsfaktor der digitalen Infrastruktur. Denn der Wert von Daten hängt von der Schnelligkeit und Qualität ihrer Verarbeitung ab.

Big Data kann aber auch einer neuen Überwachungsarchitektur den Boden bereiten. Big Data und Big Brother liegen nahe beieinander. Die Technologie weist in mancher Hinsicht strukturelle Ähnlichkeit mit der Atomkraft auf: Ihr Potenzial ist enorm, solange sie unter Kontrolle ist. Gerät sie in die falschen Hände oder wird sie falsch eingesetzt, richtet sie kaum beherrschbaren Schaden an. Bei unzureichender Risikovorsorge birgt die Delegation immer weitreichenderer Entscheidungen an immer komplexer konfigurierte und autonom agierende Systeme die Gefahr eines nachhaltigen Kontrollverlustes. Geboten ist daher eine regulatorische Umgehung des Algorithmeneinsatzes in besonders kritischen, da diskriminierungsanfälligen oder persönlichkeitsrechtlich sensiblen Entscheidungszusammenhängen, mithin ein Algorithmen-Risikotechnologierecht. Seine Aufgabe ist es, eine ausgewogene Balance zwischen den Risiken für den Schutz personenbezogener Daten und dem Interesse der Gesellschaft an innovativen Nutzungsformen herzustellen. Hierbei sollte stets das Recht die Technik, nicht die Technik das Recht bestimmen.

Das deutsche Datenschutzrecht setzt Big Data bereits heute enge Grenzen. Diese sind allerdings noch nicht hinreichend auf das digitale Zeitalter und die Chancen sowie Herausforderungen, die von Big-Data-Analysen und ihren Algorithmen ausgehen, abgestimmt. Algorithmen sind ein neuer Steuerungsmechanismus unserer digitalen Umwelt. Sie legen Filter über unsere Rekonstruktion der sozialen Realität. Als Schaltstellen in der digitalen Infrastruktur der Zukunft steuern sie Such- und Wissensprozesse. Sie machen soziale Netzwerke zu Beobachtungslabors des Sozialverhaltens, deren Analysen intransparent und beeinflussbar sind. Damit bilden sie einen zentralen Machtfaktor im »digital capitalism« (*Dan Schiller*). Ihre Manipulation nach dem Leitmuster »Ich mache mir die Welt, wie sie mir gefällt« fällt ihrem Schöpfer bzw. Anwender leicht. Die damit

[67] Vgl. auch Weichert (Fn. 8), 132.

verbundene Steuerungsmöglichkeit setzt dazu auch entsprechende Anreize. Bei ihren Adressaten sprechen Analysealgorithmen demgegenüber das Grundvertrauen in die Richtigkeit von Zahlen an. Ihrem Zauber haftet der Nimbus wissenschaftlich unangreifbarer Objektivität an. Er verleitet zu einem blinden Glauben in die Analyserichtigkeit und -redlichkeit. Ein wichtiger Schlüssel zum verantwortungsvollen Umgang mit Big Data ist insofern die Herstellung von Transparenz der Analysepraxis.[68]

Unter den Bedingungen moderner Datenanalysen geht bereits von der Zugehörigkeit zu einer Gruppe ein Risiko für die unbefangene gesellschaftliche Selbstentfaltung aus. Der Einzelne tritt dort immer häufiger als Teil einer Gruppe in Erscheinung, der Algorithmen aufgrund berechneter Korrelationen bestimmte Eigenschaften zusprechen. Ihre Bewertungsschablonen ordnen das Individuum kollektiven Handlungsmustern zu und laden damit zur Selektivität der Realitätserfassung und zu vorverurteilender Diskriminierung ein.[69] Wer in einem Gebiet wohnt, das für eine schlechte Zahlungsmoral bekannt ist, den wird womöglich manch ein Anbieter nicht oder nur gegen Vorkasse beliefern. Wer zum Shoppen nach New York möchte, muss damit rechnen, dass er diesen Plan durch Eingabe sensibler Suchbegriffe im Internet, wie z. B. »Al Qaida«, akut gefährden kann. Big Data kollektiviert insoweit das Risiko für die informationelle Selbstbestimmung, indem es auf intransparenten Erhebungen beruhenden Differenzierungen Raum gibt.[70] Das macht den Einzelnen zum Gefangenen von Wahrscheinlichkeiten, die spekulative Rückschlüsse vom Gruppenverhalten auf seine Person zulassen, jedoch nicht zutreffen müssen.[71] Bei seinen Bemühungen, eine Versicherung abzuschließen, eine Arbeitsstelle zu suchen oder Produkte im Internet zu erwerben, läuft der Einzelne dann schnell gegen unsichtbare Wände, ohne den Grund zu kennen. Gerade dadurch können Big-Data-Analysen die persönliche Freiheit und die soziale Funktionsfähigkeit von Kommunikationsinfrastrukturen[72] bedrohen.

Das muss nicht heißen, personenbezogene Daten deshalb generell Big-Data-Anwendungen zu verschließen. Das schösse über das Ziel hinaus.[73] Algorithmengesteuerte Massenentscheidungen dürfen aber – trotz ihrer Komplexität und Dynamik in lernenden Systemen – nicht frei von staatlicher Kontrolle bleiben. Sachgerechte Big-Data-Regulierung ist Algorithmenkontrolle. Sensible Big-Data-Anwendungen müssen daraufhin überprüfbar sein, welche Daten erhoben und verwertet, wie sie verschmolzen und wie die Analyseergebnisse verwendet werden. Staatliche Stellen müssen eine repressive – in sensiblen Bereichen, insbesondere im Gesundheitswesen, nach dem Vorbild des § 4d Abs. 5 Satz 1 BDSG auch die präventive – Kontrollmöglichkeit über die Struktur der Algorithmensteuerung ausüben, um einen sachgerechten Persönlichkeitsschutz außerhalb der Blackbox »Algorithmus« sicherstellen zu können. Die Risiken umfassender

---

68 Ebenso etwa J. Schneider/N. Härting, CR 2014, 306 (310); N. Härting, CR 2014, 528 (531 f.); S. Noller, PinG 2013, 20 f.

69 Roßnagel (Fn. 21), 566. Optimistischer insoweit Mayer-Schönberger/Cukier (Fn. 2), S. 203; vgl. dazu (wenn auch zu eng in der Schlussfolgerung) Leopold (Fn. 14), 80.

70 Weichert (Fn. 58), 254; Roßnagel (Fn. 23), 566.

71 Mayer-Schönberger/Cukier (Fn. 2), S. 206.

72 W. Hoffmann-Riem, AÖR 134 (2009), 513 (535).

73 So aber P. Ohm, UCLA Law Review 2010, 1701 (1742 f.).

Persönlichkeitsausforschung sowie einer Gruppendiskriminierung gilt es, durch vorsorgenden Datenschutz, insbesondere Ansprüche Betroffener auf Zugang zu gespeicherten personenbezogenen Informationen, feste Löschungsfristen für Daten, Big-Data-adäquate Anonymisierungstechniken und die Kontrolle auf potenziell diskriminierende Suchkriterien, im Zaum zu halten.[74] Die Zertifizierung und Auditierung entsprechender Analyseinstrumente, insbesondere durch die staatlichen Datenschutzbeauftragten, sowie Datenschutz-Folgeabschätzungen für besonders persönlichkeitssensitive Verarbeitungsvorgänge können insoweit ein wirksames ergänzendes Kontrollinstrument sein.[75] Sie stellen sicher, dass Herausforderungen für das Selbstbestimmungsrecht bereits bei der Konzeption neuer Technologien Berücksichtigung finden. Als Grundlage des Persönlichkeitsschutzes in einer freiheitlichen Gesellschaft ist der Staat aufgerufen, sichere Rahmenbedingungen für den Umgang mit dem Instrument »Big Data« zu entwickeln, die den Bürger Vertrauen in den Nutzen und die Ungefährlichkeit seines Einsatzes schöpfen lassen. Erst dann sieht er sich durch Big-Data-Analysen nicht lediglich zum Objekt eines informatorischen *Foucault*schen Panoptikums herabgewürdigt. Anderenfalls wird er aufseufzen wie eine der »Precogs« in dem Film »Minority Report«: »Ich brauche Ruhe, ich brauche Ruhe vor der Zukunft.«

[74] Roßnagel (Fn. 23), 566.
[75] Roßnagel (Fn. 23), 566.