

Big Data als Herausforderung für das Datenschutzrecht und den Persönlichkeitsschutz¹

Mario Martini

Übersicht

I.	Big Data: Von der Science-Fiction zur Alltagsrealität	3
1.	Wesensmerkmale	4
a)	Volume	5
b)	Velocity	6
c)	Variety	6
d)	Analysis	7
2.	Folgen für die Datenverarbeitung	7
a)	Datengestützte Handlungsempfehlungen als Zielsetzung: Big Data als digitales Orakel von Delphi	8
b)	Referenzfelder mit Wertschöpfungs- und Gemeinwohlpotenzial	10
aa)	Mobile Government	12
bb)	Neue Wertschöpfungsressourcen der Wirtschaft	13
cc)	Social-Media-Analyse und -Monitoring	15
dd)	Gesundheit und Lifestyle	15
ee)	Erkenntnispotenziale der Wissenschaft	17
II.	Missbrauchspotenzial	17
1.	Verarbeitung von Sachdaten	18
2.	Persönlichkeitssensitive Auswertungen	18
III.	Konfliktlinien mit datenschutzrechtlichen Prinzipien	21
1.	Prinzip der Erforderlichkeit, Datenvermeidung und Datensparsamkeit	21
2.	Grundsatz der Zweckbindung	22
3.	Transparenz	23
4.	Zwischenfazit	24
IV.	Rechtliche Zulässigkeit von Big-Data-Analysen	24
1.	Big-Data-spezifische Regelungstatbestände	24
a)	Profiling	25
aa)	Antworten der Rechtsordnung	25
bb)	Algorithmen als Teil einer <i>Hume'schen</i> Metaphysik zwischen »cum hoc ergo propter hoc«-Fehlschlüssen und Apophänie	25
b)	Scoring	27
2.	Allgemeine Zulässigkeitstatbestände	29
a)	Einwilligung	30

1 Eine gekürzte Fassung des Beitrages ist abgedruckt unter DVBl 2014, S. 1481 ff.

b) Gesetzliche Verarbeitungsbefugnis	31
aa) Standortdaten	31
bb) Telemedienrechtliche Bestands- und Nutzungsdaten	32
cc) Inhaltsdaten	32
(1) Verarbeitung für eigene Geschäftszwecke	
– § 28 BDSG	32
(2) Datenverarbeitung öffentlicher Stellen	35
3. Anonymisierung als Königsweg	36
a) Anonymisierung unter Big-Data-Bedingungen	37
b) Lösungswege	38
V. Regulierungsstrategien und Fazit	38
1. Kollektivierung des Risikos für die informationelle Selbstbestimmung durch Big-Data-Algorithmen	39
2. Transparenzsichernde Handlungsinstrumente	41

I. Big Data: von der Science-Fiction zur Alltagsrealität

Mit ihren spekulativen Extrapolationen technischer Veränderungen und deren Ausstrahlungen auf die Gesellschaft kommen Science-Fiction-Filme oftmals der Gegenwartsrealität zuvor – häufig auch zwiespältigen ethischen Entwicklungen. Was heute als Science-Fiction erdacht wird, muss man – wie *Norman Mailer* treffend formulierte – »morgen vielleicht als Reportage zu Ende schreiben«. Der US-amerikanische Thriller »Minority Report« aus dem Jahr 2002 ist dafür ein gutes Beispiel.² *Tom Cruise* spielt darin *John Anderton*, den engagierten Mitarbeiter einer außergewöhnlichen Washingtoner Polizeieinheit: der Abteilung Precrime. Deren Aufgabe besteht darin, auf der Grundlage von Wahrscheinlichkeitsprognosen spezieller Auguren, sog. Precogs, Verdächtige zu ermitteln und dadurch schwere Straftaten zu verhindern. In Washington hat es deshalb, so die filmische Utopie, sechs Jahre lang keinen Mord mehr gegeben. Doch das System frisst seine eigenen Kinder: *John Anderton* gerät als Mitarbeiter der Abteilung bei einem Precrime-Screening selbst in den Verdacht, bald einen Mord zu begehen. Er besucht die Wohnung der (ihm bis dahin unbekannten) Person, die er laut Vorhersage ermorden soll. Kurze Zeit später wird diese tatsächlich 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.

Was als filmische Projektion auf das Jahr 2054 konzipiert war, ist heute der Realität verblüffend nahe. *Predictive Policing* ist in den USA ein immer wichtiger werdender Bestandteil der Polizeiarbeit. Das *Real Time Crime Center* in New York führt ermittlungsrelevante Datensätze zusammen und entsendet auf deren Grundlage Einsatzkräfte an Gefahrenherde, die erhöhte Wahrscheinlichkeiten für Straftaten aufweisen.³ Die Computeranalyse⁴ perfektioniert mithilfe feinmaschiger Raster die Intuition sowie die kriminalistische Erfahrung des Polizisten vor Ort, indem sie ihm einen digitalen Assistenten zur Seite stellt.⁵ Ähnlich pilotieren US-amerikanische Sicherheitsbehörden Roboter, die mit Hilfe von Gesichtserkennungs- und Kennzeichenerfassungstechnologien ihre Umgebung durchscannen und die gesammelten Erkenntnisse mit polizeilichen Datenbanken abgleichen, um Gefahren frühzeitig zu erkennen und abzuwehren.

Auch in Deutschland bedienen sich die Sicherheitsbehörden zusehends der Funktionalitäten, die *Predictive Policing* ermöglicht. Das Land Bayern testet in den Pilotstädten München und Nürnberg die Soft-

2 Der Film geht auf die gleichnamige Kurzgeschichte des amerikanischen Autors *Philipp K. Dick* aus dem Jahre 1956 zurück.

3 Zu weiteren Beispielen des Einsatzes, insbesondere in Los Angeles und Memphis, sowie im Rahmen der Entlassung von Straftätern auf Bewährung, vgl. *Brücher*, Rethink Big Data, 2013, S. 75 ff.; *Mayer-Schönberger/Cukier*, Big Data, 2013, S. 199 ff.; *Meinecke*, Big Data und Data Mining: Automatisierte Strafverfolgung als neue Wunderwaffe der Verbrechensbekämpfung?, in: *Taeger* (Hrsg.), Big Data & Co, 2014, S. 183 (184 f.).

4 Die Möglichkeit effektiver und effizienter Überwachung bildete im Rahmen der militärischen Forschung, welcher die Computerentwicklung einen guten Teil ihrer Fortschritte verdankt, stets eine wichtige Triebfeder. Der Computereinsatz war gerade in Zeiten des Kalten Krieges sowohl in Ost als auch West von der Motivation beseelt, feindliche Aggressionen zu erkennen, bevor sie begangen werden, jedenfalls bevor sie sich auswirken konnten.

5 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 6.4.2009 zur Errichtung des Europäischen Polizeiamts, ABl. EG Nr. L 121/37 vom 15.5.2009.

ware »Precobs« zur Steuerung seiner Polizeistreifen. Deren sprachliche Anleihe an den Film »Minority Report« ist kein Zufall (und zugleich nicht unbedingt ein gutes Omen⁶): Auftrag der Software ist es, die Polizeikräfte an Orte erhöhter Wohnungseinbruchsgefahr zu entsenden.⁷ Sie macht sich dabei die Erkenntnis zunutze, dass Serientaten typischen Routinen folgen.

Die Sicherheitsbehörden in den USA wie in Deutschland springen damit auf einen allgemeinen Trend auf, der unsere Lebenswirklichkeit unaufhaltsam durchdringt: Big Data.⁸ Der Trendbegriff⁹ steht für eine technologische Entwicklung, die immer größere Datenmengen immer schneller und immer tiefgliedriger auswertbar macht (unten 1.), und dadurch einen Paradigmenwechsel in der Datenverarbeitung einläutet (unten 2.).

1. Wesensmerkmale

Schätzungen zufolge verarbeitet die Menschheit bisher lediglich ca. 12 % der vorhandenen Datenmenge.¹⁰ Big Data wird diese Ausbeutungsquote substanziell erhöhen. Bislang scheinbar wertlosen, unstrukturierten Daten¹¹ haucht die Analyse durch intelligente Verknüpfung und das Aufspüren bislang unerkannter Muster neues Leben ein: Das digitale Goldschürfen durchsiebt bereits vorhandene Datenbestände algorithmisch nach Korrelationen, strukturiert die Datenmassen um und versucht, die in ihnen schlummernden Schätze mit dem Ziel der Generierung neuen Wissens zu heben. Möglich wird das durch vier Faktoren, die schlaglichtartig das Phänomen »Big Data« beschreiben: *volume* [a)], *velocity* [also erhöhte Verarbeitungsgeschwindigkeit, b)], *variety* [sc. die technische Möglichkeit, Daten aus unterschiedlichen Kontexten zusammenzuführen, c)] sowie verfeinerte Analysemethoden [*analysis*, d)].¹²

6 Siehe dazu insbesondere II., S. 16 ff.

7 *Anonymous*, »Big Data gegen böse Buben«, FAZ vom 1.12.2014, S. 26.

8 Statt von »Big Data« spricht die Informatik vorzugsweise von »Cognitive Computing« (dazu z.B. *Haun*, Handbuch Robotik, 2. Aufl., 2013, S. 476) bzw. »lernenden Systemen«. Die Begriffe sind dabei nicht ganz deckungsgleich. Während »Big Data« stärker die Datenmenge fokussiert, stellt »Cognitive Computing« eher die Analysemethoden künstlicher Intelligenz und maschinellen Lernens in den Vordergrund, die Big-Data-Anwendungen möglich machen, insbesondere das manuelle Programmieren von Algorithmen durch lernende Systeme (Lernalgorithmen) ersetzen.

9 Er knüpft an die in der Vergangenheit die Diskussion beherrschenden, mit ähnlicher Zielrichtung und Konnotation verwendeten Begriffe »Data Mining« und »Business Intelligence« an. »Data Mining« bezeichnet Instrumente datenbasierter Mustererkennung und betont dabei das Ziel der Erkenntnisgewinnung. Der Begriff ist gleichwohl ein Stück weit irreführend, da der Prozess vorrangig nicht auf die Gewinnung von Daten, sondern von Informationen ausgerichtet ist. Die Wendung »Business Intelligence« hat sich in den neunziger Jahren in der privaten Wirtschaft als Beschreibung für die Optimierung von Entscheidungsvorgängen auf der Grundlage von Daten etabliert. Dazu etwa *Bachmann/Kemper/Gerzer*, Big Data – Fluch oder Segen?, 2014, S. 117 f., 165 ff.; *Hill*, DÖV 2010, 789 (789 f.) m. w. N.

10 *Bornemann*, RDV 2013, 232 (233).

11 Gerade das (etwa in sozialen Netzwerken) schnell anwachsende Datenvolumen des Internets produziert eine Vielzahl amorpher Daten, die mit den herkömmlichen Methoden der Datenauswertung keiner sachgerechten Analyse zuzuführen waren. Schätzungen zufolge sind heute nur 15 % der Daten (z.B. als Kundenstammdaten) strukturiert und 85 % der Daten unstrukturiert, *TNS Infratest*, Quo vadis Big Data, 2012, S. 5.

12 Vgl. zu diesen typischen Merkmalen von Big Data auch etwa *Ulmer*, RDV 2013, 227 (227 f.). Als viertes Merkmal werden statt »analysis« häufig »value«, teilweise auch »veracity« (Glaubwürdigkeit, Wahrhaftigkeit) genannt; vgl. z.B. *Bachmann/Kemper/Gerzer* (Fn. 9), S. 28 ff. »Value« bzw. »veracity« bilden freilich lediglich typische Zweckbeschreibungen des Einsatzes, nicht jedoch konstitutive Wesensmerkmale des Phänomens: Big-Data-Anwendungen sind zwar nur so gut wie

a) Volume

Je mehr sich der Cyberspace zum universellen Umschlagplatz sozialer Interaktion entwickelt, umso tiefere und präzisere Einblicke in alltägliche Kommunikations- und Präferenzmuster sowie das Seelenleben der Bevölkerung gewährt er. Noch nie zuvor in der Geschichte der Menschheit ließ sich das ganze private und kollektive Denken und Handeln so umfassend nachzeichnen.

Allein von 2000 bis 2002 hat die Menschheit mehr Daten gesammelt als in den 40.000 Jahren zuvor.¹³

Heute fassen die Daten des Internets rund 1,9 Zettabyte (ZB).¹⁴ Und das war erst der Anfang: Jede Minute kommen auf YouTube 72 Stunden neues Videomaterial hinzu. Pro Stunde laden die Nutzer auf Facebook mehr als 10 Millionen neue Bilder hoch. Das gesamte verfügbare Datenvolumen wächst jährlich um 50 %.¹⁵ Insbesondere der rasante Anstieg des weltweiten mobilen Datenverkehrs verleiht dieser Entwicklung besondere Dynamik. Er stieg im Jahre 2013 gegenüber dem Vorjahr um 81 %. Das zwischenzeitlich erreichte mobile Datenvolumen übertrifft den gesamten Internetdatenverkehr des Jahres 2000 bereits um das 18-fache.¹⁶

Triebfedern dieses Prozesses sind drastisch reduzierte Speicherkosten und die stete Verbindung des Einzelnen zum digitalen Datenstrom: Rund die Hälfte der Deutschen besitzt ein Smartphone; 4,5 Milliarden Menschen sind weltweit in sozialen Netzwerken miteinander verbunden. Nahezu jeder Transaktionsvorgang und immer mehr alltägliche Handlungen, die sich bislang ausschließlich »im realen Leben« vollzogen haben, hinterlassen digitale Spuren – vom Einkauf in Onlineshops über Bankgeschäfte, den Fahrkarteneinkauf bis hin zum Hotel-Check-In.¹⁷ Mit dem Siegeszug des E-Payments wird die Anonymität des Bargeschäfts des täglichen Lebens womöglich schon bald der Vergangenheit angehören. Der Trend verstärkt sich durch die flächendeckende Integration digital vernetzter Mess-, Steuer- und Regelsysteme in Alltagsgegenstände. Schon im Jahr 2020 sind Schätzungen zufolge womöglich 50 Milliarden Geräte miteinander

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. Die *BITKOM* versteht unter dem Begriff »Big Data« die »Analyse großer Datenmengen aus vielfältigen Quellen in hoher Geschwindigkeit mit dem Ziel, wirtschaftlichen Nutzen zu erzeugen«, *BITKOM*, Big Data im Praxiseinsatz, 2012, S. 7. Die Eingrenzung auf die Erzeugung wirtschaftlichen Nutzens ist zu eng. *Bornemann* (Fn. 10), 234 versteht unter dem Begriff »besonders große Datenmengen, die bisher noch nicht mithilfe von Standarddatenbanken und Datenmanagement-Tools verarbeitet werden können«. Vgl. auch die Begriffsbeschreibung bei *Härtig*, CR 2014, 528 (528 f.); *Mayer-Schönberger/Cukier* (Fn. 3), S. 13; *Offenhuber/Ratti*, Drei Mythen über Smart Cities und Big Data, in: Geiselberger (Hrsg.), Big Data, 2013, S. 149 (153); *Weichert*, Big Data – eine Herausforderung für den Datenschutz, in: Geiselberger (Hrsg.), Big Data, 2013, S. 131 (133).

13 *BITKOM* (Fn. 12), S. 12. Noch drastischer die Schätzungen bei *Heuer*, Kleine Daten, große Wirkung, 2013, S. 6.

14 Vgl. etwa *Fessler*, datareport 1/2013, 11 (11); *Weichert*, ZD 2013, 251 (252). Bisher hat die Menschheit Schätzungen zufolge digitale Daten im Umfang von 2,8 Zettabyte produziert, vgl. etwa <http://www.spiegel.de/netzwelt/web/das-internet-der-dinge-erzeugt-2-8-zettabyte-daten-a-872280.html> (27.3.2015).

15 *Fraunhofer-IAIS*, Big Data – Vorsprung durch Wissen, 2012, S. 6.

16 www.heise.de/netze/meldung/Weltweiter-mobiler-Datenverkehr-wuchs-2013-um-80-Prozent-2106758.html?wt_mc=nl.heise-netze (9.2.2015).

17 Auch dafür sind Smartphones wichtige Taktgeber, etwa durch die Einbindung von biometrischen Erkennungstechnologien, wie z.B. Fingerabdrucksystemen, Höhenmessern oder Bewegungssensoren. Sie schaffen neue Möglichkeiten der Informationsgewinnung, Authentifizierung und Auswertung, die das Lebensumfeld der Menschen in bisher unbekanntem Ausmaß vermessbar machen.

vernetzt.¹⁸ Die Geräte tauschen dann eigenständig Informationen aus und steuern sich wechselseitig. Die Möglichkeit zur Auslagerung von Daten in die virtuelle *cloud* trägt zusätzlich zur Vergrößerung des bislang auf lokale Daten begrenzten »Data Warehouse« bei. So gewinnen Daten, neben den klassischen Produktionsfaktoren Kapital, Arbeit und Rohstoffe, im 21. Jahrhundert immer mehr an Bedeutung.¹⁹ Den ungeschliffenen Rohdiamanten, den die gewachsene Datenmenge bildet, veredeln Big-Data-Analysten zu einem wertvollen Substrat.

b) Velocity

Schon in der Vergangenheit horteten staatliche Einrichtungen und Unternehmen häufig riesige Mengen an Informationen und Entscheidungsparametern, waren aber außerstande, diese zu ordnen und zeitgerecht bzw. in den Grenzen eines adäquaten Kosten-Nutzen-Verhältnisses auszuwerten. Neue Rechengeschwindigkeiten machen die verfügbaren Daten nunmehr binnen eines Wimpernschlags und damit in Echtzeit analysierbar. Speichergröße und Verarbeitungsgeschwindigkeit verdoppeln sich gegenwärtig entsprechend der Faustregel des *Moore'schen* Gesetzes spätestens alle zwei Jahre.²⁰

Dieser Trend prädestiniert die technischen Möglichkeiten des Phänomens »Big Data« als Assistenzinstrument für Entscheidungsprozesse. Denn nur, wenn die Verarbeitungsgeschwindigkeit mit dem Datenwachstum Schritt hält, können digitale Technologien zeitgerecht auf Lebenssachverhalte einwirken.

c) Variety

Neue Auswertungstechnologien gestatten es, heterogene Daten aus völlig disparaten operativen Quellen und Kontexten – von Textbeiträgen in sozialen Netzwerken über YouTube-Videos bis hin zu Standortdaten eines Smartphones – zu neuen Informationsgehalten und kontextübergreifenden Mustern zusammenzusetzen.²¹ Die medienübergreifende Analyse unstrukturierter Daten stößt die Türen zu neuen Dimensionen der Datennutzung auf.

18 Evans, *Das Internet der Dinge*, 2011, S. 3.

19 BITKOM (Fn. 12), S. 7 und 34; Bornemann (Fn. 10), 233.

20 BITKOM (Fn. 12), S. 22. Deutlich wird die Veränderung auch am Beispiel der Genomentschlüsselung: Während die Entschlüsselung der ersten menschlichen Genome noch mehrere Jahre brauchte, benötigt ein Genomsequenzierungsautomat der neuesten Generation dafür heute zwei Tage.

21 Die Umwandlung der unterschiedlichen Datenformate in einheitliche, maschinenlesbare Arbeitswerkzeuge, die den Verwendungskontext von Informationen (etwa im Falle von Homonymen, wie bspw. »Schimmel« als Bezeichnung für ein weißes Pferd versus ein Lebensmittel, das von einem Pilz befallen ist), im Idealfall auch mitschwingende Stimmungen, zu lesen vermögen, ist die Herausforderung und Kunst von Big-Data-Analysen. Sie müssen bei ihren Verarbeitungsvorgängen nicht nur datenschutzrechtliche Erfordernisse beachten, sondern auch urheberrechtliche Schranken sondieren sowie Sprachhürden überwinden. Vgl. Dapp, *Big Data – die ungezähmte Macht*, 2014, S. 8.

d) Analysis

Immer feinere Analysemethoden ermöglichen die Erkennung von Mustern in ungeordneten Datenmengen. Sensorik, biometrische Erkennungsverfahren, Sentimentanalyse und Technologien der Linguistik sowie Semantik²² sind wichtige Katalysatoren dieser Entwicklung.²³ Sie lassen Maschinen das Verhalten ihrer Umgebung erlernen, Sprache verstehen, Gesichter erkennen und Sinneseindrücke nachempfinden. Computer- und Neurowissenschaften simulieren in einem Prozess des »Deep Learning« die Neuronenverbindungen des menschlichen Gehirns und stellen digitale Assistenten zur Verfügung, die durch das Lernen aus und Denken in Zusammenhängen komplexe Fragen beantworten können. Einen Vorschmack auf diese neue Physik eines digitalen Bewusstseins²⁴ liefert der Kinofilm »Her«. Er erzählt die Geschichte eines Mannes, der sich in ein intelligentes Computerbetriebssystem namens »Samantha« verliebt. Im Gespräch mit ihr verarbeitet der Protagonist seine (kurz bevorstehende) Scheidung. Die beiden kommen sich dabei geistig immer näher. Die fehlende körperliche Nähe sowie die »Zweitbeziehung« von »Samantha« zu einem hyperintelligenten, dem Philosophen *Allan Watts* nachgebildeten Betriebssystem entwickeln sich jedoch zu einem unüberbrückbaren Problem.

So weit wie in dem Film »Her« ist die technische Entwicklung zwar (glücklicherweise) noch nicht fortgeschritten. Immerhin ist der künstlichen Intelligenz aber ein entscheidender Durchbruch gelungen:²⁵ Das erste Computersystem hat den sogenannten Turing-Test bestanden. Dessen Herausforderung besteht darin, einen Menschen Glauben zu machen, er habe es statt mit einem Computer mit einem anderen Menschen zu tun: Die Testperson unterhält sich in einem Text-Chat mit zwei Gesprächspartnern, die er weder sehen noch hören kann. Einer davon ist ein Mensch, der andere ein Computer. Gelingt es dem Computer, die Testperson davon zu überzeugen, dass er von den beiden Gesprächspartnern derjenige mit menschlichen Eigenschaften ist, gilt das Experiment als bestanden. Die Software »*Eugene Goostmann*« machte 10 der insgesamt 30 Testpersonen glauben, dass es sich bei ihr um eine reale Person handele.²⁶ Viele sehen eine nunmehr neue Ära der Maschinenintelligenz angebrochen, in der Computer früher oder später die kognitiven Fähigkeiten des Menschen übertreffen.

2. Folgen für die Datenverarbeitung

In dem Wettlauf um die besten Startplätze bei der Verteilung der digitalen Dividende bringen sich die großen Softwareanbieter IBM, Microsoft, Oracle, SAP und Salesforce.com mit immer ausgefeilteren

22 Diese beiden Instrumente strukturieren Freitexte in einer Weise, die quantitative Auswertungen ermöglicht. Sie greifen dazu insbesondere auf die Auswertung von Schlüsselworten zurück. Vgl. dazu etwa *Brücher* (Fn. 3), S. 83 ff.

23 Vgl. *BITKOM* (Fn. 12), S. 27; *Ulmer* (Fn. 12), 227 mit Fn. 3.

24 Dazu *Kaku*, *Die Physik des Bewusstseins*, 2014.

25 Auf spielerischer Ebene fließen menschliche und maschinelle Kommunikation schon länger einander: Das NintendoDS-Spiel »Love Plus«, das sich v. a. im asiatischen Raum immer größerer Beliebtheit erfreut, simuliert bereits heute eine Beziehung mit den fiktiven Charakteren *Rinko*, *Nene* oder *Karaka*, vgl. http://www.huffingtonpost.com/2014/01/21/loveplus-video-game_n_4588612.html (26.2.2015).

26 Siehe *Kremp*, Durchbruch bei künstlicher Intelligenz: Der unheimlich menschliche Eugene Goostmann, Spiegel online vom 9.6.2014. Zu den Ergebnissen der Software »Cleverbot«, die 59,3 % der Versuchsteilnehmer für einen Menschen hielten, vgl. www.newscientist.com/article/dn20865-software-tricks-people-into-thinking-it-is-human.html#.VWAiJUZkRyt (26.2.2015).

Analyseprodukten in Stellung, die menschlicher Intelligenz nahekommen – IBM etwa mit »Watson«, SAP mit »Hana«.²⁷ Sie durchleuchten das »digitale Panoptikum« immer tiefgliedriger und schneller – und eröffnen ihren Herstellern glänzende Geschäftsperspektiven. Big-Data-Anbietern steht ein goldenes Zeitalter bevor. Schon im Jahr 2012 lag der globale Umsatz für Big-Data-Anwendungen bei 4,6 Milliarden EUR.²⁸ Der Umsatz mit Hard- und Software rund um die Erfassung, Speicherung und Auswertung sehr großer digitaler Datenmengen wird Schätzungen zufolge²⁹ bereits in diesem Jahr auf rund 73,5 Milliarden EUR steigen. Das sind 66 % mehr als im Vorjahr. Im Jahr 2016 sollen es bereits 160 Milliarden EUR sein. Damit ist Big Data das wachstumsstärkste Segment des ohnehin dynamisch expandierenden IT-Marktes und eine Schlüsseltechnologie, die Garant für die Zukunftsfähigkeit eines Wirtschaftsstandorts sein kann.³⁰

Auf Deutschland entfällt gegenwärtig nur ein Fünftel der europäischen Big-Data-Umsätze.³¹ Viele seiner (vor allem kleinen und mittleren) Unternehmen beobachten die Umwälzungen noch ebenso aufmerksam wie zurückhaltend vom Spielfeldrand aus; rund ein Viertel der mittelständischen Unternehmen sieht ihr Geschäftsmodell vielmehr von der digitalen Entwicklung, die insbesondere Jahrzehnte alte Branchengrenzen zwischen Mechanik, Elektrotechnik und Software einreißt, bedroht.³² Zahlreiche Internetpioniere dringen mit innovativen personalisierten Angeboten in das Feld klassischer deutscher Industriebastionen vor. Deutschlands Industrie droht diejenige technologische Souveränität einzubüßen, die es in der industriellen Produktion unangreifbar zu machen schien; ihr droht die Reduzierung auf die Rolle eines bloßen Zulieferers der digitalen Ökonomie. In deren Königsklasse spielen bislang nur wenige deutsche Unternehmen. Doch Deutschland will sich für das Spiel der digitalen Zukunft warm laufen und den inzwischen eingetretenen Rückstand aufholen. Seine Unternehmen haben fraglos das Potenzial, an die Rolle als den Globus umspannende Industriewerkbank und Weltmarktführer in der Automatisierung der Fertigung anzuknüpfen – und auch in diesem neuen digitalen Wettstreit die Rolle eines Spielmachers zu übernehmen.³³ Das setzt allerdings das Anbrechen einer digitalen Gründerzeit voraus.

a) Datengestützte Handlungsempfehlungen als Zielsetzung: Big Data als digitales Orakel von Delphi

Bislang konzentrierte sich Datenverarbeitung sowohl in der allgemeinen Auswertungspraxis als auch in der Wissenschaft auf hochwertige Daten, gleichsam die Rosinen im Datenkuchen. Die Verarbeitungskapazitäten reichten nicht aus, um Vollerhebungen durchzuführen, also ohne Stichproben auszukommen. In Zeiten schier unbegrenzter Speicher- und Verarbeitungskapazität braucht es jedoch keine Vorauswahl repräsentativer Daten mehr. Eine Vollauswertung wird möglich.³⁴

27 Hana ist in der Lage, große und hochgradig verteilte Datenbestände binnen kürzester Zeit zu verarbeiten. Die Datenmengen werden dabei auf viele verschiedene Rechner aufgeteilt und dezentral bearbeitet. Großer Beliebtheit erfreut sich auch das Open-Source-Produkt »Hadoop«.

28 BITKOM (Fn. 12), S. 47.

29 BITKOM, Weltmarkt für Big Data wächst rasant, 2014.

30 BITKOM (Fn. 12), S. 48.

31 BITKOM (Fn. 12), S. 50. Zu Förderbemühungen der Bundesregierung vgl. z.B. Stöcker, APuZ 2015, 8 (9 f.).

32 Commerzbank, Management im Wandel, 2015, S. 23

33 BITKOM (Fn. 12), S. 47.

34 Zu den darin schlummernden Gemeinwohlpotenzialen siehe im Einzelnen S. 15.

Den Anspruch auf Exaktheit lässt die Big-Data-Analyse dabei zugunsten einer umfassenden Sammlung von Daten ein Stück weit hinter sich; sie nimmt auch eine geringere Validität der Daten in Kauf.³⁵ Zwar sind auch Big-Data-Analysen nur so gut wie die Qualität der Daten, die in ihren Datenkreislauf einfließen. Je mehr Daten zur Verfügung stehen und je intelligenter sie zu neuen Informationsgehalten verknüpft werden, umso eher lassen sich entsprechend dem Gesetz der großen Zahl jedoch Unschärfen in der Datenerhebung akzeptieren.³⁶ Mehr Masse führt hier auch zu mehr Klasse.³⁷ Dafür ist das Verarbeitungsparadigma von Big Data in der Lage, Zusammenhänge zu erkennen, die der herkömmlichen Datenlese verborgen bleiben – wie ein intelligenter Zuhörer, der zwischen den Zeilen lesen und Nuancen erkennen kann – ohne überdies etwas zu vergessen. Big-Data-Rechner, wie Watson, verarbeiten nicht nur große Datenmengen und sind nicht nur statisch programmiert. Sie lernen auf der Grundlage von Interaktion und der Analyse eigener Rechercheergebnisse stets hinzu. Mit jedem Klick verbessern sie ihre Leistung und passen ihre Modellparameter veränderten Bedingungen an.³⁸

Ziel der digitalen Sieb- und Rastertechnik sind evidenzbasierte Grundlagen für bessere Entscheidungen und Verhaltensprognosen. Big-Data-Technologien ergänzen den Augenschein um Algorithmen, das Gedächtnis um Datenbanken und das Bauchgefühl um Statistik. Ihre Daten-Alchemie läutet einen Paradigmenwechsel in der Auswertung von Daten zur Unterstützung menschlicher Entscheidungen ein: In einer Big-Data-Welt durchforsten Designmuster für Algorithmen das virtuelle Chaos nach Korrelationen und spüren dort bislang unerkannte Muster auf, aus denen sie Handlungsempfehlungen und Schlussfolgerungen ableiten. Die Analyse identifiziert Potenziale und beantwortet Fragestellungen, die bislang in der Fülle von Daten verborgen geblieben sind. Die Ergebnisse erlauben entscheidungsrelevante Rückschlüsse und generieren Herrschaftswissen, das sich Big-Data-Generatoren nutzbar machen können. Der im Orakel von Delphi aufscheinende, alte Menschheits Traum, die Zukunft präziser vorherzusehen, scheint sich zu verwirklichen. Eine neue Ära der Entscheidungsunterstützung bricht an.

Gegenwärtig lässt sich nur erahnen, welche Wissensschätze sich noch aus den Bergen von Daten gewinnen lassen. Doch bereits heute greift eine Goldgräberstimmung um sich. Sie ist der Leitidee verschrieben: Wer die Zukunft kennt, dem gehört sie.³⁹ Das Feuer der Begeisterung, welches das Phänomen »Big Data« in der digitalen Wirtschaft entfacht, erinnert an die Planungseuphorie der 60er und 70er Jahre, als man

35 Vgl. *Bachmann/Kemper/Gerzer* (Fn. 9), S. 183 f.; *Mayer-Schönberger/Cukier* (Fn. 3), S. 179; das übersieht *Leopold*, vorgänge 2012, 74 (77).

36 *BITKOM* (Fn. 12), S. 27; *Mayer-Schönberger/Cukier* (Fn. 3), S. 21. Kritisch dazu *Boyd/Crawford*, Big Data als kulturelles, technologisches und wissenschaftliches Phänomen – Sechs Provokationen, in: Geiselberger (Hrsg.), Big Data, 2013, S. 187 (197).

37 Umgekehrt können sich aber auch Unschärfen in ihrer verzerrenden Wirkung potenzieren und die Richtigkeit von Ergebnissen suggerieren, die sich von der Realität signifikant entfernen. Deutlich macht dies das Phänomen des sogenannten »Schmetterlingseffekts«, den als erster *Konrad Lorenz* beschrieb. *Lorenz* versuchte, auf der Grundlage von Computermodellen Wetterprognosen zu erstellen. Bei dem Versuch, unter nahezu gleichen Ausgangsbedingungen die Prognose zu replizieren, stellten sich jeweils unterschiedliche Ergebnisse ein. Geringe Modifikationen der eingegebenen Parameter veränderten die Ergebnisse erheblich. Darin offenbarte sich die Erkenntnis: Komplexe, nichtlineare dynamische Systeme reagieren empfindlich auf kleine Abweichungen der Ausgangsbedingungen.

38 Daher auch die Bezeichnung »lernende Systeme«, siehe dazu oben Fn. 8; vgl. auch *Türpe*, DuD 38 (2014), 31 (32 f.).

39 *Klausnitzer*, Das Ende des Zufalls, 2013, S. 31.

davon überzeugt war, durch gute Planung dem Gemeinwohl zu dienen und die Gesellschaft mithilfe kybernetischer Modelle zu durchschauen. In dieser Welt wird der Statistiker zum Helden des Informationszeitalters und die Big-Data-Analyse zur Basis für Prozess- und Entscheidungsoptimierungen.

b) Referenzfelder mit Wertschöpfungs- und Gemeinwohlpotenzial

Big Data ist keine Zukunftsvision des Silicon Valley. Die Technologie ist längst Teil unserer Alltagsrealität. Viele ihrer Prototypen möchten wir nicht mehr missen, etwa die Routenplanung. Sie arbeitet nach den Funktionsprinzipien von Big-Data: Navigationsgeräte aggregieren unzählige Ortsdaten, die wir uns früher mühsam mit Karten Schritt für Schritt erschlossen haben, und fügen diese mittels einer komplexen Berechnung individueller Präferenzen und aktueller Verkehrsinformationen zu einer Route zusammen.

Die ersten Ursprünge von Big-Data-Anwendungen liegen viel länger zurück. Es war die Schutzgemeinschaft für Absatzfinanzierung, welche im Jahr 1927 als erste ihre Entscheidungen auf ein »System zur Beurteilung des Zahlungsverhalten« stützte. Daraus ist im Jahre 1952 die Bundes-SCHUFA hervorgegangen; sie mündete im Jahr 2000 bzw. 2002 in die SCHUFA Holding AG. Wo konventionelle Methoden der Datenverarbeitung, die auf strukturierten Daten aufbauen, etwa die Kreditbewertung, ihre Grenzen erreichen, fängt Big Data an.⁴⁰ Es handelt sich um eine logische Fortentwicklung bestehender Technologien, die einen Quantensprung der Datenauswertung ermöglichen, um belastbare Aussagen zur Bewertung sowie Vorhersage von Ereignissen zu treffen.

Amazon, Facebook und Google führen uns das Prinzip in perfektionierter Reinform vor Augen. *Amazon* sammelt Myriaden von Daten über seine Kunden, durchsucht sie nach Mustern und destilliert daraus Empfehlungen, die das Kaufverhaltensgenom entschlüsseln. *Facebook* ist als Marktplatz menschlicher Eitelkeiten der größte Maskenball unserer Zeit. Er vereinnahmt die Nutzer in informationellen Kokons einer Aufmerksamkeitsmaschinerie, aus denen auszubrechen vielen schwerfällt. Als kollektive Autobiografie und personalisierte Zeitung wertet Facebook seine globale, von Milliarden von Menschen gefütterte Datenbank mit Hilfe von »Facebook Insights« & Co. aus, dechiffriert den Social-Graph und schlägt uns auf dieser Grundlage neue Freunde sowie Produkte und Dienstleistungen vor, die unseren Präferenzen entsprechen könnten. Je mehr Inhalte die Internetnutzer in das Netz einspeisen, desto genauer lassen sich deren Konsumverhalten und persönliche Neigungen dokumentieren und gewinnbringend verwerten. »Facebook sagt mir, wer ich bin. Amazon sagt mir, was ich will. Google sagt mir, was ich denke«,⁴¹ hat *George Dyson* die komplementären Geschäftsmodelle der prominentesten Big-Data-Kollektoren auf den Punkt gebracht. In ihrer radikalen Effizienzorientierung und technologischen Überlegenheit überrollen die großen Internetkonzerne die Strukturen der überkommenen analogen Ökonomie wie eine Feuerwalze. Insbesondere *Google* hat sich wie kein zweites Unternehmen als Brandbeschleuniger dieser Entwicklung sowie

40 *BITKOM* (Fn. 12), S. 8 und 11.

41 *Dyson, Turing's cathedral*, 2012, S. 308.

als Informationsgoldmine des 21. Jahrhunderts in Stellung gebracht.⁴² In den USA verdient der Internetkonzern inzwischen mehr mit Werbung als sämtliche gedruckten Zeitungen und Zeitschriften zusammen. Das Backup des globalen Gedächtnisses, das Google erstellt hat und den Menschen als Schaufenster in das digitale Fantasieland öffnet, die Perfektionierung der Internetrecherche mit Hilfe seiner Autocomplete-Funktion,⁴³ die geodatenbasierte Erfassung der Welt via Google Street View, Google Maps und Google Earth sowie seine Big-Data-Übersetzungssoftware (die zwar sprachliches Feingespür und Intuition – noch – vermissen lässt, aber erstaunliche mechanisierte Übersetzungsleistungen erbringt) setzen Maßstäbe in der Welt der Datenauswertung. Bereits im Jahr 2009 sagte Google Flu – auf der Grundlage einer Analyse der Suchanfragen seiner Nutzer – die weitere Ausbreitung der Schweinegrippe vergleichsweise zuverlässig voraus.⁴⁴ So erstaunlich treffsicher die Suchergebnisse sind, so sehr dokumentieren sie die ausgereifte Magie der ihnen zugrunde liegenden Algorithmen. Mithilfe der neuen Möglichkeiten experimentieren Amazon & Co. bereits mit Geschäftsmodellen, die den Kunden bedarfsgerecht Waren zusenden, bevor sie diese bestellt haben (*anticipatory shopping*).⁴⁵

Genauso vielfältig wie die analysierten Datenmengen selbst sind auch ihre Einsatzzwecke: Marktforschung, individualisierte Werbung (das sogenannte »Targeted Advertising«), Kreditvergabe, Qualitätsforschung oder Personalplanung gehören ebenso zu dem Leistungsportfolio moderner Big-Data-Anwendungen wie die Optimierung von Verkehrsleitsystemen, die Bedarfsplanung für öffentliche Infrastrukturvorhaben oder die Auswertung von Internetdaten zur Bekämpfung der Kriminalität oder missbräuchlicher Inanspruchnahme sozialer Leistungen.⁴⁶

42 Glaser, Erinnerung der Zukunft, in: Geiselberger (Hrsg.), Big Data, 2013, S. 281 (283). Warnend Hofstetter, APuZ 2015, 33 (34): »Die Macht verschiebt sich weg vom demokratisch legitimierten Staat, dessen Repräsentanten durch den Souverän wählbar und kontrollierbar sind, hin zu Wirtschaftsbetrieben, die über unsere Daten verfügen [...]«.

43 Aus der Zusammenführung massenhaft gesammelter Nutzerdaten ergibt sich womöglich auch eine marktbeherrschende Stellung der Unternehmen. Während ihre technischen Features sich vergleichsweise leicht imitieren lassen, können Konkurrenten das Datenpotenzial, das den Humus für die Funktionalität der Dienste bildet, nur schwer abbilden. Das verschafft den Platzhirschen einen nahezu nicht einholbaren Wachstumsvorsprung und begründet für Newcomer eine Marktzutrittschürde im Wettbewerb der Suchmaschinen- bzw. Sozialen Netzwerke-Dienste. Die Europäische Kommission hat deshalb gegen Google ein Kartellverfahren eingeleitet, welches die Bevorzugung seiner eigenen Angebote gegenüber konkurrierenden Angeboten im Rahmen des Werbesystems Google AdWords bzw. Google Shopping zum Gegenstand hat. Vgl. Pressemitteilung der Kommission »Kommission leitet Untersuchung gegen Google wegen unfairen Wettbewerb ein« vom 15.4.2015, abrufbar unter ec.europa.eu/deutschland/press/pr_releases/13234_de.htm (17.4.2015); siehe auch Körber, NZKart 2014, 293 (293); Paal, GRUR-Beilage 2014, 69 (71 f.); ders., ZRP 2015, 34 ff.

44 Vgl. <https://www.google.org/flutrends/intl/de/de/#DE> (15.3.2015); Mayer-Schönberger/Cukier (Fn. 3), S. 7 ff.; Dugas/Hsieh/Levin et al., CID 2015, 463 (465 ff.); vgl. auch Dapp (Fn. 21), S. 26 f. Zur Fehleranfälligkeit von Google Flu bei der Vorhersage der Wintergrippe 2012 in den USA Mayer-Schönberger, APuZ 2015, 14 (17).

45 Der aggregierte Datenpool kann dabei nicht nur für die Unternehmen selbst, sondern auch für eine Vielzahl von Geschäftspartnern, etwa Banken, Versicherungen und Gesundheitsunternehmen, hochgradig interessant sein. Das macht die Auswertungsmöglichkeiten der Datensammlung sensibel. Dazu im Einzelnen unten II., S. 16 ff.

46 Zieger/Smirra, MMR 2013, 418 (418); Kreibich, APuZ 2015, 20 (23 f.); Fraunhofer-IAIS (Fn. 15), S. 8.

aa) Mobile Government

Gerade mobile Anwendungen eröffnen bislang ungeahnte Möglichkeiten gemeinwohlorientierter Datenauswertung.⁴⁷ Auf der Grundlage von Positions- und Bewegungsdaten der Verkehrsteilnehmer lassen sich Schwerpunkte der Verkehrsauslastung ermitteln, Verkehrsströme lenken und Staus vermeiden, typische Unfallherde erfassen und vorhersagen sowie die Transportpotenziale von Buslinien und Bahnen optimieren. Intelligente Mobilitätssysteme vernetzen den öffentlichen und privaten Verkehr nahtlos zu einem übergreifenden Verkehrsökosystem.

Die Städte Dublin und Stockholm etwa, aber auch New York, machen es vor. Dublin hat im Jahr 2013 mithilfe der Software »Infosphere Streams« ein Projekt lanciert, das die Verkaufszahlen der Fahrkartenautomaten, Abgaswerte, Aufzeichnungen von Straßensensoren und Videokameras sowie die GPS-Daten der in Dublin eingesetzten Busse analysiert, um den Verkehrsfluss zu optimieren. Auf ähnliche Weise wertet Stockholm die Daten von 250.000 GPS-Geräten sowie Sensor- und Videoinformationen über Wetter-, Unfall- und Staumeldungen aus, um die Verkehrsteilnehmer auf dem besten Weg zu ihrem Ziel zu leiten. Die Fahrzeiten im öffentlichen Personen- und Nahverkehr ließen sich dadurch deutlich reduzieren. Die Stadt New York nutzt den elektronischen Mautpass (EZ-Pass) auch dazu, den Verkehrsfluss in Echtzeit zu erfassen und zu optimieren, insbesondere Störungen rechtzeitig entgegenzuwirken.⁴⁸ Die Zahl der Städte, die diesem Beispiel folgen, nimmt auch in Deutschland rasant zu. Das Nürnberger Verkehrsunternehmen VAG wertet die Bewegungsprofile seiner Nutzer auf der Grundlage angekaufter Daten von Kunden der Deutschen Telekom zur Verbesserung seiner Services aus.⁴⁹

Nicht nur im Bereich der Optimierung der Verkehrsinfrastruktur, sondern auch für die sachgerechte Analyse von Investitionsentscheidungen insgesamt, etwa für die Ermittlung der optimalen Position einer Windenergieanlage, für die Versorgungsplanung und den Bereich des Energiesparens, liefern mobile (im Verbund mit stationären) Daten als Bausteine wertvolle Grundlagen. Die öffentliche Bedarfs- und Energieversorgungsplanung lässt sich so auf eine neue Effizienzstufe heben;⁵⁰ ein digitales Optimierungsmanagement der gesamten Verkehrs- und Versorgungsinfrastruktur wird möglich. Die Telekommunikationsunternehmen, deren Gewinnmargen im harten Wettkampf um Marktanteile darben, sehen mit der Analyse anonymisierter und aggregierter Kundendaten für sich bereits neue Geschäftsfelder aufblühen.⁵¹

47 Dazu auch *Buschauer*, (Very) nervous systems. Big Mobile data, in: Reichert (Hrsg.), Big Data, 2014, S. 405 ff.

48 Vgl. dazu *Spies*, ZD-Aktuell 2013, 03734.

49 *Blaß*, Bewegungsstatistik: Verkehrsbetrieb wertet Mobilfunkdaten aus, www.spiegel.de/netzwelt/netzpolitik/daten-schutz-vag-in-nuernberg-greift-telekom-daten-ab-a-1024001.html; *Anonymous*, 17 Telekommitarbeiter und ihr Blick in die Zukunft, FAZ vom 21.3.2015, S. 26.

50 Vgl. dazu etwa auch *Bachmann/Kemper/Gerzer* (Fn. 9), S. 203 ff.; *Ulmer* (Fn. 12), 228.

51 Vgl. etwa am Beispiel der Deutschen Telekom: www.optout-service.telekom-dienste.de/public/index.jsp (18.12.2014).

bb) Neue Wertschöpfungsressourcen der Wirtschaft

Der Wirtschaft eröffnen sich durch Big Data neue Wertschöpfungsketten effizienten Ressourceneinsatzes.⁵² Unternehmen wüssten für ihre Dispositionen nur zu gerne, welche Menge welchen Artikels in Zukunft gebraucht wird, wie die Schadensprognose für ihre Versicherungsnehmer aussieht und wie es um die Wechselbereitschaft eines Telekommunikationskunden zu einem Wettbewerber bestellt ist. Big-Data-Analysen verheißen, den Zugang zu diesem Wissensreservoir zu erschließen. Dazu gehören maßgeschneiderte Marketing- und Vertriebsstrukturen mit minimalen Streuverlusten ebenso wie zeitsensitiv in ständiger Rückkopplung mit dem Kunden eingesetzte Impulse, insbesondere Location-based-Marketing (z.B. »Google Places«)⁵³ und In-Store-Verhaltensanalysen mit ausgetüftelten, individuell zugeschnittenen Kundenangeboten⁵⁴ – nach dem überkommenen Vorbild von Tante Emma als analoger Datenbank, die immer genau wusste, zu wem welches Produkt passt.⁵⁵ Big-Data-Technologien optimieren die Markt- und Wettbewerbsbeobachtung, die Betriebsabläufe und Fertigungsprozesse. Sie ermöglichen Produktgestaltungen und -weiterentwicklungen, welche das Nutzungsverhalten sowie die Präferenzen der Kunden zielgenauer befriedigen, und helfen bei der Ermittlung des Lieferbedarfs.⁵⁶

In der Logistik lassen sich durch die Erfassung von Verbrauchs- und Positionsdaten sowie des Zustands von Verschleißteilen Wartungs- und Stillstandskosten minimieren und Transporte einfacher disponieren, insbesondere Leerfahrten verringern und Beiladungen steuern.⁵⁷ Produktsensoren entlang der gesamten Produktions- und Lieferkette machen das möglich. Sie beziehen Maschinen und Geräte durch IP-Vernetzung in den Datenkreislauf ein.

52 Vgl. zu Beispielen möglicher Innovationen bspw. *BITKOM*, Big Data und Geschäftsmodell-Innovationen in der Praxis: 40+ Beispiele, 2015. Nach einer Online-Umfrage des Instituts Fraunhofer-IAIS sehen die befragten Unternehmen das größte Potenzial von Big Data im Aufbau strategischer Wettbewerbsvorteile (69 %), gefolgt von der Umsatzsteigerung (61 %) und der Kosteneinsparung (55 %), *Fraunhofer-IAIS* (Fn. 15), S. 8. Im Bereich des Handels liegt der Anwendungsschwerpunkt in der Absatzprognose, der umsichtigen Steuerung sowie dem Marktmonitoring. Den Banken und Versicherungen ist es vor allen Dingen um Betrugserkennung sowie die Risikoabschätzung bestellt, *Fraunhofer-IAIS* (Fn. 15), S. 23 f.

53 Zu den rechtlichen Herausforderungen von Location-based-Services *Brandenburg/Leuthner*, Location based Services und Local Commerce, in: Taeger (Hrsg.), Big Data & Co, 2014, S. 651 (656 ff.).

54 Die amerikanische Supermarktkette Walmart hat auf der Grundlage einer Analyse seiner Kundendaten etwa festgestellt, dass in den Abendstunden Bier und Windeln auffällig oft gemeinsam den Weg in den gleichen Einkaufskorb finden. Entsprechend hat das Unternehmen seine Regaloptimierung an dieses Einkaufsverhalten junger Väter ausgerichtet, vgl. *Dapp* (Fn. 21), S. 28 f. Besondere Aufmerksamkeit rief ein Fall im Bundesstaat Minnesota hervor: Die Kaufhauskette »Target« hatte aufgrund einer Analyse des Einkaufsverhaltens seiner Kunden u.a. erkannt, dass Schwangere ab dem dritten Schwangerschaftsmonat verstärkt parfümfreie Lotionen kaufen. Entsprechend hatte das Unternehmen einer Kundin im Highschool-Alter Werbung für Babyprodukte zukommen lassen. Ein Vater brachte diese Werbung auf die Zinnen: Er warf dem Unternehmen vor, seine Tochter durch Werbung zur Schwangerschaft anzustiften. Das Unternehmen wusste zu diesem Zeitpunkt freilich bereits mehr als er: Die Tochter war schon schwanger. Vgl. *Dapp* (Fn. 21), S. 27. Zu den Rechtsfragen verhaltensbezogener Online-Werbung siehe *Arning/Moos*, ZD 2014, 242 ff.; *Zeidler/Brüggemann*, CR 2014, 248 ff.

55 Das Fenster zum Kunden des »unbekannten Verbrauchers« öffnen insbesondere soziale Netzwerke oder Micro-Blogging-Dienste, wie der Kurznachrichtendienst »Twitter«, in zunehmendem Umfang aber auch mobile Bezahlendienste. Datenanalysen durchleuchten sie auf Präferenzen der Kunden. Zu den technischen Möglichkeiten von *Beacons* als Instrument von Location-based-Services siehe *Schürmann/von der Heide*, (i)Beacons- technische Hintergrund und Datenschutz – rechtliche Anforderungen, in: Taeger (Hrsg.), Big Data & Co, 2014, S. 637 ff.

56 *BITKOM* (Fn. 12), S. 35 und 37; *Ulmer* (Fn. 12), 228.

57 *BITKOM* (Fn. 12), S. 10 und 34.

In Zukunft werden Haushaltsgeräte,⁵⁸ Fahrzeuge und andere Maschinen, vom Kühlschrank bis zur Waschmaschine, von der Ampel bis zum Heizkörper, im Datenkreislauf via eigener IP-Adressen in Interaktion treten. Jeder Gegenstand ist in dem steuernden *Cyber-Physical System* über seinen gesamten Lebenszyklus identifizier- und lokalisierbar – sowohl sub specie seiner Herkunft, seines Funktionsstatus als auch seiner Aufgabe in der Systemkette. Von der Bestellung der Fertigungsbestandteile über den Konstruktionsplan sowie den Produktionsprozess bis hin zur Auslieferung und zum anschließenden Kundenservice – alles ist über einen RFID-Chip, also einen mit Funkwellen ansteuerbaren Transponder, vernetzt. Virtuelle und physische Welt greifen nahtlos ineinander; kollaborative Produktionsstrukturen lösen überkommene Befehlsketten und Silostrukturen ab. In dieser Integration internetfähiger Chips in Alltagsgegenstände liegt der genetische Code eines neu entstehenden Kommunikationsnetzes: des Internets der Dinge.⁵⁹ Es lässt die Maschinen miteinander reden. In sein Ökosystem ist die gesamte dezentrale Produktionstechnik eingebunden. Die Glieder der Lieferkette vernetzen sich branchen- und technologieübergreifend. Das ebnet einer selbstadaptiven Logistik des Alltags ebenso den Weg wie einem Echtzeitmanagement in der industriellen Produktion, das Just-in-time-Konzepte optimiert und Warenströme über Fabrikttore hinweg effizienter zusammenführt (sog. *Lean Logistics*).

Arbeitsabläufe und Fertigungsstrukturen werden effizienter, schneller und flexibler. An die Stelle großer Fertigungseinheiten tritt die individuelle Fertigung, die passgenaue Einzelartikel zum Preis von Massenprodukten herstellt.⁶⁰ Print-on-Demand-Konzepte des Buchhandels führen das Prinzip bereits heute vor Augen. Eine neue Ära der Automatisierung, Überwachung und Steuerung von industriellen Prozessen bricht an. Sie ermöglicht eine punktgenaue und bedarfsgerechte Steuerung von Produktionsabläufen auf der Grundlage einer vollständigen Integration aller Herstellungsschritte – ausgehend vom Kunden bis zum Netzwerk der Lieferanten – und verheißt dadurch einen Quantensprung der Vielfalt und Schnelligkeit industrieller Produktion. Der Kunde mit seinen individuellen Wünschen und Anforderungen tritt in den Mittelpunkt des Produktionsprozesses. Indem die Prozesse sich den immer schneller und flexibler wandelnden Kundenanforderungen anpassen, setzen sie die Effizienzbemühungen der Automatisierungsindustrie in einer vierten industriellen Revolution⁶¹ logisch fort: Während die vorherigen industriellen Revolutionen die physische Arbeit von Menschen und Tieren durch Maschinen ersetzen und Produktionsabläufe verbesserten, ergänzt das Internet der Dinge die menschliche Prozesssteuerung durch maschinelle Kommunikation und eine Optimierung des Informationsstroms.

58 Vgl. *Wagner*, Datenschutz in Connected Homes, in: Peters/Kersten/Wolfenstetter (Hrsg.), *Innovativer Datenschutz*, 2012, S. 205 ff.

59 Dazu etwa *Andelfinger/Hänisch*, Internet der Dinge, 2015, S. 9 ff.; *Bräutigam/Klindt*, NJW 2015, 1137 ff.; *Evans* (Fn. 18), S. 1 ff.

60 Seit *Henry Ford* gilt das Paradigma: Eine Fabrik ist umso wirtschaftlicher, je mehr einheitliche Produkte sie fertigt. Dieses Paradigma löst sich in der Welt der Industrie 4.0 ein Stück weit auf.

61 Dazu bspw. *Bauernhansl*, Industrie 4.0 in Produktion, Automatisierung und Logistik, 2014, S. 5 ff.; *Hirsch-Kreinsen*, Digitalisierung industrieller Arbeit, 2015; *Pinnow/Schäfer*, Industrie 4.0 (R)Evolution für Wirtschaft, Politik und Gesellschaft, 2015, S. 1 ff.

cc) Social-Media-Analyse und -Monitoring

Via Big Data lassen sich Präferenzartikulationen der Bevölkerung erfassen und typische Verhaltensmuster erkennen, um daraus Potenziale für die Gemeinwohlorentwicklung ebenso wie die Produktoptimierung zu erschließen.⁶² Strömungen in der Bevölkerung lassen sich identifizieren und begleiten, bevor sich Fronten verhärten oder Fehlplanungen in die Verschwendung von Ressourcen münden. Technologien der Sentimentanalyse und der Auswertung digitaler Stimmungsbilder sozialer Interaktion in Netzwerken sind bereits weit entwickelt und vielfach im praktischen Einsatz. Sie machen sich das Kommunikationsbedürfnis, aber auch den Narzissmus und den Hang zur Selbstdarstellung des Homo digitalis zunutze. Unternehmen können auf der Grundlage der Signale, die er in den Cyberspace aussendet, ihre Werbestrategien und ihre Produktpolitik auf seine Bedürfnisse ausrichten. Mehr noch: Im Zeitalter sozialer Netzwerke sind die Kunden nicht nur das Produkt, sondern selbst die beste Werbung, ja die authentischsten Multiplikatoren für erfolgreiche Marken. Das Spektrum der Anwendungssphären reicht von der Unterstützung des Issue-, Reputations- und Krisenmanagements über die Erfolgsmessung der eigenen Öffentlichkeitsarbeit, die Optimierung des Kundenservice, Meinungsführeridentifikation und Markt- sowie Wettbewerbsbeobachtungen bis hin zum Benchmarking und zur Event-Detection.⁶³

Öffentliche Stellen folgen diesem Trend mit Verzögerung, aber immer nachhaltiger. Sie nutzen entsprechende Echtzeit-Analyseformate zur Trend- und Meinungsforschung, zur politischen Strategieoptimierung und als Stethoskop politischer Krisenintervention, bildet sich doch die gesamte Bandbreite gesellschaftspolitischer Dialoge und Meinungen mittlerweile auch im Netz ab. Social-Media-Monitoring kommt als Baustein der Gefahrenprävention bei Großveranstaltungen, als Sensorium zur Aufdeckung allgemeiner Gefahren für die öffentliche Sicherheit und Ordnung sowie als Überwachungsinstrument der Geheimdienste ebenso zum Einsatz wie bei der Evaluation behördeneigener Social-Media-Präsenzen. Zu den Schattenseiten des Monitorings gehört aber auch das Risiko einer digitalen Blockwart-Mentalität, die demokratische Selbstentfaltung und Entwicklung der Gesellschaft durch vollständige Erfassung zu ersticken droht.

dd) Gesundheit und Lifestyle

In der Auswertungswelt der Big-Data-Analyse bilden Gesundheitsdaten die Kronjuwelen. Medizinisches Monitoring⁶⁴ und digitale Assistenzsysteme gehören bei der Pflege älterer Menschen ebenso zum Alltag der Zukunft wie bei der Lifestyle-Optimierung des modernen Managers. Die Datenspur, die der Einzelne hinterlässt, wird immer länger. Fitness-Tracking-Bänder, wie »Fitbit« oder das »Nike Fuelband«, welche die gesunde Lebensweise, insbesondere die Bewegungsintensität, »auf Schritt und Tritt« überwachen, machen den Anfang. Das Smartphone entwickelt sich im Big-Data-Zeitalter zur Gesundheitszentrale der Selbstoptimierung; Gesundheitsschutz wird zum digitalen Informationsmanagement des Körpers.

62 Dazu auch *President's Council of Advisors on Science and Technologie*, Report to the President: Big Data and Privacy: A technological Perspective, 2014, S. 28 f.; *Pentland*, Social physics, 2014.

63 Vgl. *Hofmann*, Methoden des Social-Media-Monitoring, in: König/Stahl/Wiegand (Hrsg.), Soziale Medien, 2014, S. 161 ff.

64 So können Implantate den Herzstatus von Patienten überwachen und das Herz im Falle eines Herzversagens durch elektrisches Schocksignal wieder zum Schlagen bringen, *Stöcker* (Fn. 31), 8. Zu den Anwendungsmöglichkeiten von Big Data in der Medizin siehe bspw. *Langkafel*, APuZ 2015, 27 ff.

Dass der im Wege des Fitness-Trackings ermittelte Health-Score auch Einfluss auf den Krankenversicherungstarif haben könnte, ist längst keine Utopie mehr,⁶⁵ vermutlich eher die nächste Stufe einer Entwicklung. Die AOK Nordost und andere Krankenversicherer experimentieren bereits mit solchen Modellen.⁶⁶ Sie planen, den Health-Score in bestehende Bonusprogramme zu integrieren.⁶⁷ Ausländische Anbieter tun dies bereits, so etwa der italienische Marktführer Generali⁶⁸ und die amerikanische Versicherungsgesellschaft United Health. Die bietet ihren Versicherten einen Tarif an, der diese verpflichtet, mithilfe eines Messgeräts festzuhalten, wie viele Schritte sie am Tag gehen.

Die Kontrolle des Gesundheitsverhaltens schützt die Versicherer ein Stück weit vor einer versicherungstypischen strukturellen Vertragsasymmetrie: Kunden können ihr Risikoverhalten im Anschluss an den Vertragsschluss selbst steuern und verändern, ohne dass der Versicherer darauf Einfluss hat. Das künftige Verhalten, insbesondere bewusst in Kauf genommene Selbstschädigungen, kann der Versicherer bei seiner Versicherungsprämienanpassung fortan berücksichtigen und damit für eine verursachergerechte Zurechnung von Schadensfolgen sorgen.

Im Extremfall könnten »Wearables« jedoch einer neuen Gesundheitsökonomie den Boden bereiten, die via »Fitcoin« eine datenbasierte Metrik für die solidarische Finanzierung von Gesundheitsleistungen entwickelt und eine Kommerzialisierung sensibler Daten einläutet, welche die Logik des Marktes in die letzten Nischen unserer Persönlichkeitsentfaltung vordringen lässt. Für die Gewährung wirtschaftlicher Vorteile sind viele Menschen bereit, gesundheitsbezogene Daten mit ihrem Krankenversicherer zu teilen. Die – unter Rückgriff auf den Anreizmechanismus des Preisvorteils wirkende – »Karotte vor der Nase« gestattet nicht nur eine gezielte Ansprache von insbesondere jungen Kunden,⁶⁹ sondern als Kehrseite auch eine vollständig Erfassung und Überwachung der privaten Lebenswelt, insbesondere des Gesundheitsverhaltens. Sie stellt zugleich den Solidargedanken des Prinzips der gesetzlichen Krankenversicherung auf die Probe. Schließlich beruht dieses seinem Wesen nach auf dem Gedanken eines solidarischen Ausgleichs von Risiken. Sind diese auf der Grundlage einer Preisgabe sensibler Gesundheitsdaten individuell messbar, kommt es zu einer Risikoselektion.

Das gleiche ambivalente Potenzial wohnt der Nutzung von Big-Data-Analysen als Instrument der medizinischen Therapie inne. So erleichtern die Massendatenanalyse und Echtzeitauswertung komplexer Daten Durchbrüche in der medizinischen Forschung, etwa bei der Analyse der Nebenwirkungen medizinischer Therapien oder der Korrelationen zwischen Krankheitsbildern und genetischen sowie sozioökonomischen

65 Vgl. zu entsprechenden Pay-as-you-drive-Modellen für KfZ-Versicherungen *Armbrüster*, NJW 2015, Editorial Heft 6; *LDI NRW*, 22. Datenschutzbericht, 2015, S. 37 ff.; *Lüdemann/Sengstacken/Vogelpohl*, RDV 2014, 302 ff.; *Schwichtenberg*, DuD 2015, 378 ff.

66 Siehe das entsprechende App-Angebot unter www.aok.de/portale/nordost/mobil-vital/ (25.2.2015).

67 *Heller*, Zeigt mir, wie fit du bist, FAS vom 6.4.2014, S. 55. Ob eine solche Durchleuchtung persönlicher Lebensweisen in diesem Umfang datenschutzrechtlich unter dem Gesichtspunkt der Erforderlichkeit (§ 284 Abs. 1 SGB V) gerechtfertigt ist, ist zu bezweifeln.

68 Er lässt seinen Kunden im Gegenzug zur Übermittlung mobiler Gesundheitsdaten Gutscheine zukommen. Vgl. »Versicherer Generali will Fitnessdaten von Kunden sammeln«, <http://www.sueddeutsche.de/news/wirtschaft/versicherungen-versicherer-generali-will-fitnessdaten-von-kunden-sammeln-dpa.urn-newsml-dpa-com-20090101-141121-99-02990> (4.3.2015).

69 *BITKOM* (Fn. 12), S. 36.

Einflussfaktoren.⁷⁰ Big-Data-Analysen versprechen dort grundlegende Erkenntnisse für die Entwicklung moderner Krebsbehandlungen und Therapien personalisierter Medizin.⁷¹ Sie erhöhen zugleich aber den Bestand diskriminierungsgeneigten Wissens um individuelle genetische oder sonstige physiologische Dispositionen.

ee) Erkenntnispotenziale der Wissenschaft

Big-Data-Analysen erweitern den Horizont unserer Erkenntnisse über die Rekonstruktion der Wirklichkeit. Insbesondere die Einbeziehung des Bürgers in den Prozess der Datengenerierung (sog. »Citizen Science«) eröffnet der Forschung ganz neue Datengrundlagen und Auswertungsmöglichkeiten. Ungeahnt große statistische Grundgesamtheiten ersetzen die konventionelle Stichprobenanalyse, vereinfachen und perfektionieren Prognosen sowie Simulationen. Die Hypothesenaufstellung kann auf Korrelationsmuster zurückgreifen, welche die wissenschaftliche Wahrheitssuche auf die relevanten Fragen ausrichtet. Prognosen über den Verlauf von Epidemien, Konjunkturentwicklungen, Klimaveränderungen und demografische Verschiebungen erzielen eine bislang unerreichte Treffsicherheit. Auch der experimentellen Psychologie, Soziologie und Ökonomik erwächst eine neue Spielwiese, auf deren Nährboden eine Probandenanalyse im Verhaltenslabor »Wirklichkeit« stattfinden kann, um Hypothesen algorithmisch überprüfen zu lassen.⁷²

Mit dieser Verbreiterung der Datengrundlage verändert sich auch die herkömmliche Form wissenschaftlichen Arbeitens. Statt der Verifizierung einer Hypothese auf der Grundlage der Frage nach dem »Warum?«, welche die wissenschaftliche Theorienbildung umtreibt, stellen Big-Data-Interferenzmodelle die induktive Frage nach dem »Was?« in den Vordergrund.⁷³ Der Journalist *Chris Andersen* glaubt gar, dass sich das überkommene Wissenschaftsmodell der Hypothesenaufstellung, Modellerstellung und des Modelltests vollständig überlebt hat und durch eine Analyse von Korrelationen als neues Forschungsparadigma abgelöst werden wird.⁷⁴

II. Missbrauchspotenzial

Bis zum Jahr 2013 stand der Trendbegriff »Big Data« aus dem Silicon Valley noch primär für die ökonomischen Ideen verschriebene Aggregierung und Analyse großer unstrukturierter Datenmengen mithilfe von Algorithmen – also eine Evolution der privaten Welt. Seit den Enthüllungen von *Edward Snowden* steht der Begriff aber auch für eine systematische staatliche Überwachung privater Lebensführung unbekannten Ausmaßes und eine neue Form des digitalen Imperialismus, welche die Privatheit mithilfe

70 BITKOM (Fn. 12), S. 37 und 40.

71 Dazu *Dehmer/Holzinger/Emmert-Streib*, Personalized medicine by means of complex networks: a Big Data challenge, in: Weber/Thouvenin (Hrsg.), Big Data und Datenschutz - Gegenseitige Herausforderungen, 2014, S. 37 ff.

72 Dazu auch allgemein *Mayer-Schönberger*, FuL 2014, 706 (706); *Rudder*, Dataclysm, 2014.

73 Zu dieser Logik siehe auch unten S. 23 ff.

74 *Anderson*, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, Wired Magazine vom 23.6.2008; in diese Richtung bereits ähnlich *Wolfram*, A new kind of science, 2002, der eine Welt voraussah, in der Computerexperimente, welche Klassifikationen und Korrelationen ermitteln, die Stelle mathematischer Beweise und Theorien einnehmen.

digitaler Massenausforschungswaffen zu kolonisieren droht.⁷⁵ Staatliche Überwachungsprogramme wie PRISM haben der Welt deutlich gemacht: So groß die Chancen sind, so groß ist auch das Missbrauchspotenzial neuer technologischer Auswertungsinstrumente. Die algorithmische Einhegung von Entscheidungsabläufen eröffnet tiefe Einblicke in unser digitales Alter Ego, perfektioniert die Berechenbarkeit menschlichen Verhaltens und damit die Möglichkeit, die engere persönliche Lebenssphäre des Einzelnen auszuforschen. Wie ein »Fahrtenschreiber« registriert und verarbeitet sie Informationen über das Verhalten der Menschen und offenbart damit die dunkle Seite einer Technologie, die aus Beobachtungsdaten auf zugrunde liegende Gesetzmäßigkeiten schließt. In einer Welt der Algorithmen können personenbezogene Daten den Menschen im Wege technologischer Vermessung lesbar machen und sein Innerstes wie eine codierte Betriebsanleitung entziffern.⁷⁶

Dass Big-Data-Anwendungen einer neuen Überwachungslogik sowie Techniken der Verhaltenskontrolle, des Abgreifens von Daten in Echtzeit und ihrer Umwandlung in Kontroll- und Planungssysteme den Boden bereiten können, fordert unsere Datenschutzverfassung heraus; sie lässt das Bedürfnis nach zeitgemäßen Antworten der Rechtsordnung offen zutage treten.

1. Verarbeitung von Sachdaten

Nicht alle Big-Data-Analysen zielen allerdings auf persönlichkeitsensitive Auswertungen. Manche verarbeiten allein Sachdaten, etwa die Werkleistungen einer Maschine mittels RFID-Chip. Diese Auswertungsprozesse unterfallen dann nicht dem Regime des Datenschutzrechts – jedenfalls solange sie nicht (sei es unmittelbar, sei es mittelbar) im Wege der Identifizierung von Lebensmustern und -gewohnheiten, Rückschlüsse auf Personen zulassen. Betroffen ist der Schutzgedanke des Datenschutzrechts nämlich nicht generell, sondern nur, soweit personenbezogene Daten erhoben, verarbeitet oder genutzt werden (§ 1 Abs. 1 und 2 BDSG).

2. Persönlichkeitssensitive Auswertungen

Die meisten Big-Data-Anwendungen sind freilich wesensmäßig darauf gerichtet, das Verhalten einer Person oder Personengruppe zu analysieren und/oder zu antizipieren, sei es als Berechnung von Wahrscheinlichkeitswerten für künftiges vertragliches Verhalten (*Scoring*), als Akkumulierung inkonnexer Daten zu einem detailgetreuen digitalen Persönlichkeitsprofil (*Profiling*),⁷⁷ sei es als Beantwortung einer

75 In diesem Sinne Zuboff, Die neuen Massenausforschungswaffen, FAZ vom 13.2.2014, S. 33; Schirrmacher, Der verwetete Mensch, in: Geiselberger (Hrsg.), Big Data, 2013, S. 273 (273).

76 Mühl, Korrekturen ausgeschlossen, FAZ vom 14.2.2014, S. 35.

77 Das Europäische Parlament schlägt in seiner legislativen Entschließung vom 12.3.2014 zu dem Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr COM(2012)0011 folgende Legaldefinition des Profilings vor: »jede Form automatisierter Verarbeitung personenbezogener Daten, die zu dem Zweck vorgenommen wird, bestimmte personenbezogene Aspekte, die einen Bezug zu einer natürlichen Person haben, zu bewerten oder insbesondere die Leistungen der betreffenden Person bei der Arbeit, ihre wirtschaftliche Situation, ihren Aufenthaltsort, ihre Gesundheit, ihre persönlichen Vorlieben, ihre Zuverlässigkeit oder ihr Verhalten zu analysieren oder vorauszusagen.« (Art. 4 Abs. 3a des Verordnungsentwurfs).

Fragestellung in Bezug auf eine bestimmte Person, z.B. der Wahrscheinlichkeit krankheitsbedingter Arbeitsausfälle, (*Personalizing*) oder als Rückverfolgung auf der Grundlage einer Spurenbildung (*Tracking*).⁷⁸

Die Deutschen stehen einer derartigen Durchleuchtung ihrer engeren persönlichen Lebenssphäre durch eine digitale Wertschöpfungsmatrix, welche Wissen über Verhaltensmuster monetarisiert bzw. einer staatliche Überwachungsmaschinerie zuführt, besonders sensibel gegenüber. Das dokumentieren nicht alleine die helle Entrüstung über die NSA-Überwachung sowie die hierzulande besonders hohe Aufmerksamkeit für den dystopischen Roman »The Circle« des US-Amerikaners *Dave Eggers* über das Innenleben eines Internetgiganten, der mit Parolen wie »Alles Private ist Diebstahl« die volle soziale Kontrolle anstrebt. In Deutschland regt sich auch Protest gegen Vorhaben internationaler Unternehmen, die in anderen Ländern ohne größere Anteilnahme der Öffentlichkeit gestartet sind.⁷⁹ Als bspw. das Telekommunikationsunternehmen O2 ankündigte, - ähnlich wie bereits in Großbritannien - die Standortdaten seiner deutschen Mobilfunkkunden anonymisiert für Marketingzwecke gegen Entgelt an Dritte weiterzugeben, entfachte das einen Sturm der Entrüstung. Das Unternehmen nahm seine Pläne umgehend zurück.⁸⁰ Groß war auch die Empörung, als die SCHUFA ankündigte, das Potenzial sozialer Netzwerke für die Bonitätsprüfung auszuloten.⁸¹ Nach überschäumender Kritik hat die SCHUFA diesen Plan aufgegeben. Ein anderes deutsches Unternehmen setzt ihn demgegenüber derweil bereits um: Nach eigenen Angaben nutzt das Hamburger Unternehmen Kreditech Lokalisationsdaten, Verweildauern auf Webseiten und Profilangaben in sozialen Netzwerken als Kern seines Geschäftsmodells, um im Wege eines Big-Data-Scoring Entscheidungen über die Vergabe von Krediten zu treffen.⁸²

Die Sensibilität der Bevölkerung gegenüber einem Zugriff auf ihr digitales Kommunikationsverhalten hat einen guten Grund: Die Verfügungsgewalt über Daten ist ein zentraler Machtfaktor der digitalen Welt. Die systematische Datenauswertung erzeugt eine Informationsasymmetrie, aus der ein Gefühl des Überwachtwerdens und eine Schieflage der Chancen zur Verwirklichung von Lebensplänen erwachsen können. Die historischen Erfahrungen der NS-Zeit und der Bespitzelung durch die Stasi wirken insoweit in den Köpfen der Deutschen in besonders intensiver Weise nach.

Das Verfassungsrecht hegt diesen besonderen Respekt vor der Individualität des Einzelnen normativ ein: Der Schutz der Privatheit ist das unverzichtbare Fundament für die selbstbestimmte Entfaltung der Persönlichkeit – und damit einer demokratischen Gesellschaft.⁸³ Sie überlässt dem Einzelnen die Ausfüllung

78 Weichert (Fn. 14), 255.

79 Vgl. aber auch die Erfahrungen, welche der Karten- und Navigationssystemanbieter TomTom in den Niederlanden sammelte: Als bekannt wurde, dass das Unternehmen anonymisierte Verkehrsbewegungsdaten an die niederländische Polizei verkaufte, damit diese ihre Verkehrskontrollen optimieren kann, strafften viele Kunden das Unternehmen mit Verachtung und Boykott. Vgl. <http://www.spiegel.de/netzwelt/gadgets/standortsuche-fuer-radarfallen-tomtom-entschuldigt-sich-fuer-deal-mit-der-polizei-a-759464.html> (9.2.2015).

80 Vgl. zur Chronik *Brücher* (Fn. 3), S. 117 f.; zur rechtlichen Zulässigkeit *Mantz*, K&R 2013, 7 ff. Ähnliches gilt auch für das Pilotprojekt der VAG Nürnberg zur Verbesserung der Verkehrsplanung auf der Grundlage anonymisierter Telekommunikationsdaten; siehe dazu *Anonymous* (Fn. 49), S. 26 und Fn. 49.

81 Vgl. <http://www.spiegel.de/netzwelt/web/schufa-will-kreditdaten-bei-facebook-sammeln-a-837454.html> (16.3.2014).

82 Vgl. <http://www.kreditech.com/what-we-do/> (25.2.2015).

83 Vgl. BVerfGE 65, 1 (40); *Bornemann* (Fn. 10), 233.

seiner gesellschaftlichen Rolle, seines Selbstverständnisses und seiner Identität ohne vorgeformte Muster. Widersprüchliches, paradoxes und emotionsgeleitetes Verhalten gehört dazu ebenso wie Persönlichkeitsverschiebungen und die Sprunghaftigkeit menschlichen Verhaltens. Big-Data-Technologien vermögen dieses Wechselspiel menschlichen Seins ihrer Natur nach nur unzureichend zu erfassen. Sie legen vielmehr technische Schablonen über ihre Erfassungsobjekte. Menschen sind hinsichtlich ihres Nutzungs- und Kommunikationsverhaltens in vielen Lebenssituationen zugleich aber auch leicht berechen- und vorhersehbar.⁸⁴ Aus den Verbindungsdaten allein zweier Wochen lassen sich aussagekräftige Vorhersagen für den künftigen Aufenthaltsort einer Person zu einem bestimmten Zeitpunkt ableiten.⁸⁵ Wenn Menschen wissen oder vermuten, dass sie beobachtet werden, verhalten sie sich nicht unbefangen. Sie neigen dann – bewusst oder unbewusst – dazu, den (unterstellten) Erwartungen des Beobachters zu entsprechen.⁸⁶ Schon der Verdacht, dass das Gegenüber über Zusatzwissen verfügt oder Informationen (offen oder verdeckt) aufzeichnet, verändert eine Kommunikationsbeziehung. Wer unter ständiger Beobachtung steht, begibt sich so der Chance, sich frei zu entfalten. Kreativität, Freiheit und Innovationskraft fallen dem zum Opfer, gründen sie doch auf Abweichungen von tradierten Verhaltensmustern.

Das Allgemeine Persönlichkeitsrecht sichert die Freiheit individueller Lebensgestaltung daher gegen eine umfassende Beobachtung und eine davon ausgehende Beeinträchtigung individueller Freiheitsentfaltung umfänglich ab; als Vorfeld-Grundrecht verleiht es insbesondere der Verwirklichung anderer Freiheitsrechte, die auf der Freiheit individueller, unbefangener Entfaltung des Einzelnen aufbauen (wie z.B. die Vereinigungsfreiheit oder die Versammlungsfreiheit), normativen Flankenschutz.⁸⁷

In diese Unbefangenheit der Kommunikationsbeziehungen dringen Big-Data-Technologien ein. Sie lösen in einer auf freiheitlichen Werten beruhenden Gesellschaft die Furcht vor einer Verdinglichung des in ein Internet der Dinge inkorporierten Menschen aus. Die Vorstellung, dass der digitale Zwilling als Schatten des Individuums (etwa im Wege eines Pre-Employment-Screenings) komplexe Persönlichkeitsstrukturen ähnlich einer Rasterfahndung⁸⁸ durchsichtig macht, erinnert an mahnende Schreckensbilder der Science-Fiction-Welt. Big-Data-Technologien lassen diese aber Realität werden. Deutlich macht das etwa der Plan der amerikanischen Investmentbank JP Morgan Chase & Co, ihre Mitarbeiter mithilfe von Algorithmen und eines dadurch gewonnenen Mitarbeiterprofils auf der Grundlage einer Auswertung ihrer E-Mails und

84 Vgl. dazu auch *Weichert* (Fn. 14), 255.

85 *Martini*, Vom heimischen Sofa in die digitale Agora: E-Partizipation als Instrument einer lebendigen Demokratie?, in: Hill/Schliesky (Hrsg.), Neubestimmung der Privatheit, 2014, S. 193 (213); *Ulmer* (Fn. 12), 230 m. w. N.

86 Vgl. auch BVerfGE 65, 1 (43); 100, 313 (358 f.); 109, 279 (354 f.); 120, 378 (405 f.); 122, 342 (369); 125, 260 (320); vgl. auch zur Theorie der objektiven Selbst-Aufmerksamkeit in der Psychologie *Wiekens/Stapel*, *Social Psychology* 41 (2010), 10 ff.

87 Vgl. auch BVerfGE 120, 378 (397).

88 Die Rasterfahndung darf als Urtyp einer Big-Data-Analyse gelten. Zu ihren gesetzlichen Regelungen siehe insbesondere § 5 G 10. Von einer Rasterfahndung unterscheidet sich die Big-Data-Analyse dadurch, dass der Nutzer bei der Rasterfahndung wissen muss, was er sucht. Die Big-Data-Analyse stellt demgegenüber Zusammenhänge her und identifiziert Suchfaktoren. Die Förderung der Rasterfahndung verdankt ihren Ursprung nicht zuletzt dem besonderen Einsatz des BKA-Chefs *Horst Herold*, der Ende der sechziger Jahre als Nürnberger Polizeipräsident mit den Möglichkeiten der modernen Computertechnik experimentierte.

des sonstigen digitalen Kommunikationsverkehrs automatisiert auf verdächtige Verhaltensmuster zu überprüfen, um Verstöße gegen Risikostandards und Verhaltensrichtlinien aufzudecken.⁸⁹

Angriffe gegen den digitalen Zwilling richten sich gegen den Menschen und seine Freiheit selbst. Treffen lernende Algorithmen Entscheidungen, ohne dass diese – aufgrund der Komplexität der Korrelationen und der Echtzeitverwertung im Einzelfall – noch vom Menschen nachvollziehbar und damit kontrollierbar ist, beschwört das überdies die Furcht vor einer autonomen Steuerung herauf, deren Zauberlehrling die Geister, die er rief, nicht wieder los wird.

III. Konfliktlinien mit datenschutzrechtlichen Prinzipien

Indem Big-Data-Analysen auf eine breite Datengrundlage angewiesen sind, um ihren Mehrwert generieren zu können, treten sie in einen Zielkonflikt mit zentralen Grundgedanken des Datenschutzrechts, insbesondere den Prinzipien der Datensparsamkeit sowie Erforderlichkeit (unten 1.), der Zweckbindung (unten 2.) und der Transparenz (unten 3.).⁹⁰

1. Prinzip der Erforderlichkeit, Datenvermeidung und Datensparsamkeit

Datenschutz verlangt Datenaskese. Die Datenverarbeitung ist – in ihrem Inhalt, Umfang und in ihrer zeitlichen Erstreckung – auf das für den zulässigen Zweck Erforderliche zu begrenzen (vgl. etwa § 6b Abs. 1 S. 1, § 13 Abs. 1, § 14 Abs. 1 S. 1, § 28 Abs. 1 S. 1 Nr. 1 BDSG [»erforderlich ist«]). Das Prinzip der Datensparsamkeit und Datenvermeidung gründet auf eine einfache Erkenntnis: Das Risiko für die informationelle Selbstbestimmung wächst proportional zur Menge der gespeicherten und verarbeiteten Daten. Entsprechend sind als Ausfluss des Verhältnismäßigkeitsprinzips technische Systeme so zu gestalten, dass sie ihre Funktion mit möglichst wenigen personenbezogenen Daten erfüllen können (§ 3a S. 1 BDSG); Verarbeitung und Vorhaltung sind nur so lange zulässig, wie sie erforderlich sind, um den konkreten Zweck zu erreichen. Vorrangig sind anonymisierte oder pseudonymisierte Daten zu verwenden. Sobald Daten nicht mehr benötigt werden, sind sie zu löschen (§ 35 Abs. 2 S. 2 Nr. 3 und 4 BDSG; Art. 23 Abs. 2 S. 1 u. 2 Datenschutz-Grundverordnung-E).

Die darin liegende Funktion des sozialen Vergessens läuft bei Big Data tendenziell leer.⁹¹ In seiner Welt gibt es nämlich kein irrelevantes oder unnützes Datum mehr, legt doch gerade ein neuer Datensatz womöglich die entscheidende Spur zu einer neuen Korrelation frei, die wertvolle Erkenntnisse hervorbringt.

Das setzt einen Anreiz, auch ältere Daten in möglichst großer Zahl und für möglichst lange Zeit vorzuhalten, um auf dieser breiten Datengrundlage neue Informationen zu generieren. Informationen, die einmal den Weg in die Komplexität des Verarbeitungsvorgangs von Big-Data-Prozessen gefunden haben, finden den Weg dann kaum wieder heraus. Während in der realen Welt Geschehnisse den Gesetzen der Flüchtigkeit und der »Gnade des Vergessens«, die Erinnerungen verblassen lässt, unterworfen sind, ist die digitale

89 www.bloomberg.com/news/articles/2015-04-08/jpmorgan-algorithm-knows-you-re-a-rogue-employee-before-you-do (6.5.2015).

90 Vgl. auch die Bündelung der Datenschutzprinzipien in Art. 5 des Entwurfs zur EU-Datenschutz-Grundverordnung.

91 *Roßnagel*, ZD 2013, 562 (564).

Welt in der Lage, längst vergessene Momentaufnahmen wieder wachzurufen, zu neuen Mustern zusammenzusetzen und aus den zugrunde liegenden Mosaiksteinen neue Profile zu kreieren. Dies aber kann Persönlichkeitsbilder, die sich in der realen Welt bereits weiterentwickelt haben, digital mumifizieren. Das Internet wirkt dann wie ein Sarkophag für Daten. Erst, wenn Datenbestände auch im Regime von Big Data wie makulierte Bücher aussortiert werden, liefern Analysen realitätsgerechte Persönlichkeitsprofile. Überholte Daten tragen nicht der Dynamik von Persönlichkeitsentwicklungen Rechnung – und dem Respekt, den sie verdienen; sie sind auch keine hilfreiche Grundlage für Zukunftsprognosen, welche massenhafte Datenauswertung ihrem Wesen nach hervorbringen soll.⁹² Gerade in einer Big-Data-Welt ist daher entscheidend, wer in welcher Weise Kontrolle über den Datenbestand ausübt, unter welchen Voraussetzungen Daten in die Analysemaschinerie eingespeist werden können und wie lange sie dort verbleiben sowie in welchen Rhythmen und nach welcher Logik überholte, unnütze von aktuellen, relevanten Daten geschieden werden.

Die Persönlichkeit des Menschen braucht zu ihrer freien Entfaltung die Möglichkeit zum Neustart. So wie Resozialisierung ohne Vergessen und ohne Versöhnung nicht möglich ist, ist auch Persönlichkeitsschutz ohne ein Recht auf Vergessenwerden unter den Bedingungen von Big Data freiheitssichernd kaum realisierbar.

Das Datenschutzrecht formt diesen Leitgedanken bereits in Ansätzen aus: Es verlangt eine Löschung, wenn sich der ursprüngliche Erhebungszweck erledigt hat (§ 35 Abs. 2 S. 1 Nr. 3 BDSG; dazu auch im Einzelnen unten IV. 2. b cc (1) (β), S. 33). Ob noch weitere neue Zwecke einer Big-Data-Verwendung hinzutreten, ist für die Löschungspflicht unerheblich.

2. Grundsatz der Zweckbindung

Daten dürfen nicht für jeden beliebigen, sondern nur für den Zweck genutzt werden, für den sie erhoben worden sind; der Erhebungszweck begrenzt die Verarbeitungsbefugnis (vgl. insbesondere § 88 Abs. 3 S. 2 TKG, § 12 Abs. 2 i. V. m. § 14 Abs. 1 und § 15 Abs. 1 TMG, § 28 Abs. 1 S. 2, Abs. 2 und Abs. 5, §§ 31 und 39 BDSG, § 78 Abs. 1 S. 1 SGB X; Art. 5 Nr. 1 lit. b, 23 Abs. 2 Datenschutz-Grundverordnung-E).⁹³ Nur so ist es dem Betroffenen möglich, die Preisgabe von Daten entsprechend seiner sozialen Rolle und seinem Selbstverständnis selbst zu steuern.

Big-Data-Technologien bevorraten Daten zu ergebnisoffenen, also unbestimmten Zwecken, um sie zur Gewinnung neuer Erkenntnisse zu verschneiden und auf neue Korrelationen zwischen scheinbar zusammenhangslosen Informationsbausteinen zu screenen. Erst am Ende der Auswertung lässt sich dann sagen,

⁹² Glaser (Fn. 42), 294.

⁹³ Weniger streng demgegenüber Art. 6 Abs. 1 lit. b der Richtlinie 95/46/EG (Datenschutzrichtlinie). Sie beschränkt sich auf das Verbot, Daten in einer mit den Zweckbestimmungen nicht zu vereinbarenden Weise zu verarbeiten. Zwar zielt die Richtlinie grundsätzlich auf eine Vollharmonisierung (EuGH, Urt. v. 24.11.2011 – C-468/10 –, ECLI:EU:C:2011:777, Rn. 26 ff.). Im Hinblick auf die Grundsätze des Art. 6 gesteht sie den Mitgliedstaaten jedoch einen Gestaltungsspielraum zu (vgl. auch Art. 5 der Richtlinie 95/46/EG). Eine solche Konkretisierung nimmt der deutsche Zweckbindungsgrundsatz in unionskonformer Weise vor.

welche genaue Zielsetzung der Prozess hatte. Die Analyse liefert Antworten auf Fragen, die bisher gar nicht gestellt wurden. Enge Zweckbegrenzungen laufen dieser Mission zuwider.⁹⁴

Um ihre Ziele zu erreichen, ist die Vorhaltung und Verarbeitung von Massendaten zwar erforderlich. Das Phänomen »Big Data« setzt die Daten allerdings nicht zu dem Zweck ein, zu dem sie ursprünglich erhoben worden sind. Die Analyse löst sie aus dem Erforderlichkeitszusammenhang des Erhebungszwecks heraus. Die zu einem spezifischen Zweck gespeicherten Daten mit vielen anderen Daten zusammenzuführen, verletzt die datenschutzrechtlich vorgegebene, auf das Erforderliche beschränkte Zweckbindung; der wertschöpfungsorientierte Grundgedanke von Big Data verläuft quer zur persönlichkeitsrechtlichen Philosophie der Zweckbindung.⁹⁵

3. Transparenz

Sowohl das deutsche Datenschutzrecht als auch der Entwurf für eine Datenschutz-Grundverordnung⁹⁶ sind von dem Gedanken der Transparenz beseelt. Diesem Zweck sind insbesondere die Informationspflicht bei Datenerhebung sowie neuer Speicherung und Zweckänderung (§ 33 BDSG), Auskunftsrechte (§ 34 BDSG)⁹⁷ und Lösungsrechte (§ 35 BDSG) verschrieben. Für die Bürgerinnen und Bürger muss nachvollziehbar sein, ob und ggf. welche Stellen welche Daten zu welchem Zweck und in welchem Umfang sammeln und auswerten.⁹⁸ Sie sollen wissen, ob, von wem und zu welchem Zweck Daten erfasst werden.⁹⁹ Nur so bleibt die digitale Souveränität des Einzelnen über die Verwendung seiner Daten gewahrt. Bleibt dem Bürger verborgen, »wer was wann und bei welcher Gelegenheit über ihn weiß«, kann ihn das »in seiner Freiheit wesentlich hemmen, aus eigener Selbstbestimmung zu planen oder zu entscheiden«.¹⁰⁰ Individuelle Entfaltung kann nur dann gelingen, wenn der Einzelne die Folgen einer Preisgabe von Daten überblicken und über deren Verwendung autonom entscheiden kann.¹⁰¹ In der Welt des Cyberspace fühlt sich der Einzelne noch mehr als in der analogen Welt abhängig und überwacht von Technologien, deren Wirkungsweise er immer weniger zu durchschauen vermag. Er sieht sich dem dunklen Gefühl einer steten Überwachung ausgesetzt.¹⁰² Das kann die Grundlagen freier, ungestörter Kommunikation untergraben: Diese ist nicht nur die Basis für die Entfaltung anderer grundrechtlicher

94 BITKOM (Fn. 12), S. 26; Roßnagel (Fn. 91), 564.

95 Roßnagel (Fn. 91), 565; Schaar, RDV 2013, 223 (225); vgl. auch aus der Perspektive des Schweizer Rechts Thouvenin, Erkennbarkeit und Zweckbindung, in: Weber/Thouvenin (Hrsg.), Big Data und Datenschutz - Gegenseitige Herausforderungen, 2014, S. 61 ff.

96 Insbesondere Erwägungsgrund Nr. 46 und Art. 11 ff. Datenschutz-Grundverordnung-E.

97 Dazu auch Art. 15 sowie insbesondere Erwägungsgrund Nr. 51 S. 2 Datenschutz-Grundverordnung-E: »Jede betroffene Person sollte einen Anspruch darauf haben zu wissen und zu erfahren, zu welchen Zwecken die Daten verarbeitet werden, wie lange sie voraussichtlich gespeichert werden, wer die Empfänger der Daten sind, nach welcher allgemeinen Logik die Daten verarbeitet werden und welche Folgen eine solche Verarbeitung haben kann.« (Entwurfssfassung des Europäischen Parlaments vom 12. März 2014). Zu möglichem Reformbedarf des Auskunftsanspruchs unter den Bedingungen von Big Data Liedke, K&R 2014, 709 ff.

98 Vgl. Weichert (Fn. 12), 141.

99 Erwägungsgrund Nr. 46 S. 2 Datenschutz-Grundverordnung-E.

100 Jeweils BVerfGE 65, 1 (43).

101 Roßnagel (Fn. 91), 563.

102 Vgl. dazu auch oben S. 16 ff.

Freiheiten, sondern auch »eine elementare Funktionsbedingung eines auf Handlungs- und Mitwirkungsfähigkeit seiner Bürger begründeten freiheitlichen demokratischen Gemeinwesens«.¹⁰³

Big-Data-Analysen verwehren dem Einzelnen die Möglichkeit, zu antizipieren, welche Schlüsse aus welchen Daten in bestimmten Kontexten gezogen werden. Die Betroffenen können nicht mehr nachvollziehen, wie die Ergebnisse zustande kommen.¹⁰⁴ Ihnen bleibt regelmäßig nicht nur unklar, *welche Daten*, sondern auch *welche Analysemethoden* in den Prozess eingehen. Die Unternehmen betrachten Letztere insbesondere regelmäßig als Betriebs- und Geschäftsgeheimnis.¹⁰⁵ Dies macht Big-Data-Analysen zu einer Black Box. Das läuft dem Bedürfnis nach Datentransparenz zuwider. Insbesondere ist die Nachvollziehbarkeit der Verfahren wichtige Voraussetzung für das Vertrauen in die Willkürfreiheit einer Entscheidung. Big-Data-Analysen müssen deshalb transparent machen, wie sie zu den Ergebnissen kommen.¹⁰⁶

4. Zwischenfazit

Die Prinzipien des Datenschutzrechts erscheinen in einer Big-Data-Welt geradezu anachronistisch – wie ein Relikt aus Zeiten von Lochkarten und Formularbögen. Ihr Schutzinstrumentarium zur Sicherung einer freien Entfaltung der Persönlichkeit steht der vollen Entfaltung der Möglichkeiten von Big-Data-Technologien im Wege. Dies ändert freilich nichts daran, dass diese Datenschutzprinzipien zum Schutz der Persönlichkeit heute im Grundsatz aktueller und erforderlicher denn je sind. Denn erst im digitalen Zeitalter werden mit den fortschreitenden technischen Möglichkeiten die Bedrohungen der Persönlichkeit real, zu deren Abwehr die Datenschutzprinzipien konzipiert sind. Das schließt deren Anpassung an die Herausforderungen des digitalen Zeitalters jedoch nicht aus. Denn die Datenschutzprinzipien sind kein Selbstzweck, welche die Datenauswertung zu einem technischen Museum erheben sollen, sondern Teil einer Gesamtordnung mit konkurrierenden Bedürfnissen, die es zu befriedigen gilt. Die Zielsetzung sachgerechten Persönlichkeitsschutzes und die Möglichkeit, das Wertschöpfungspotenzial zu erschließen, das moderne Technologien in sich bergen, gilt es, zu einem schonenden Ausgleich zu bringen.

IV. Rechtliche Zulässigkeit von Big-Data-Analysen

1. Big-Data-spezifische Regelungstatbestände

Das Datenschutzrecht steht den Herausforderungen algorithmengestützter Massendatenauswertung (auch jenseits seiner abstrakten Grundprinzipien) nicht vollends sprachlos gegenüber. Für einzelne, mit besonderem persönlichkeitsbezogenem Gefährdungspotenzial verknüpfte Anwendungsformen hält es bereits konkrete Teilregelungen vor. Das gilt insbesondere für das Profiling (a) und das Scoring (b).

103 BVerfGE 65, 1 (43); vgl. auch BVerfGE 120, 378 (430); 122, 342 (369).

104 Vgl. auch Koch, ITRB 2015, 13 (18); Mayer-Schönberger/Cukier (Fn. 3), S. 225; Weichert (Fn. 14), 257.

105 Weichert (Fn. 14), 257; dazu auch im Einzelnen S. 25.

106 So auch schon Mayer-Schönberger/Cukier (Fn. 3), S. 225.

a) Profiling

Profiling besteht darin, die Mosaikbausteine der Eigenschaften einer Person zu einem Persönlichkeitsbild zusammenzuführen. Interessant sind solche Puzzles bspw. als allgemeine Kundenprofile für Unternehmen, aber auch als Internet-Meinungsführer-Analysen, die Informationen über sog. *Influencer* offenlegen, welche eine zentrale Funktion für die gesellschaftliche Wahrnehmung eines Unternehmens ausüben.¹⁰⁷

aa) Antworten der Rechtsordnung

Die gesamte Persönlichkeit eines Menschen (zwangsweise) zu registrieren und zu katalogisieren, hat das BVerfG früh (im sogenannten Mikrozensus-Urteil) für unzulässig erklärt.¹⁰⁸ Für die Auswertung personenbezogener Daten zur Bewertung einzelner Persönlichkeitsmerkmale trifft das BDSG vor diesem Hintergrund in § 6a eine Regelung. Noch deutlicher und prominenter zieht aber die geplante Datenschutz-Grundverordnung in ihrem Art. 20 dem Profiling Grenzen¹⁰⁹. Beide Regelungen verbieten Profiling nicht – sehr wohl aber automatisierte, d. h. ohne jede persönliche Prüfung erfolgende Profiling-Maßnahmen, die für Betroffene eine rechtliche Folge nach sich ziehen oder sie erheblich beeinträchtigen.¹¹⁰

Der Entwurf der Datenschutz-Grundverordnung räumt dem Betroffenen darüber hinaus grundsätzlich ein Widerspruchsrecht gegen Profiling ein (Art. 20 Abs. 1 i. V. m. Art. 19 Abs. 1). Bei Maßnahmen, die dem Betroffenen gegenüber rechtliche Wirkungen entfalten oder seine rechtlichen Interessen oder Freiheiten in ähnlicher Weise berühren, stellt er Profiling zudem unter den Vorbehalt der Erforderlichkeit für einen Vertragsschluss, einer Einwilligung oder einer sonstigen ausdrücklichen Erlaubnis (Art. 20 Abs. 2).¹¹¹

bb) Algorithmen als Teil einer *Hume'schen* Metaphysik zwischen »cum hoc ergo propter hoc«-Fehlschlüssen und Apophänie

Dass sowohl das gegenwärtige als auch das künftige Datenschutzregime eine menschliche Prüfung und eine Erklärung des Ergebnisses von Profiling-Maßnahmen normativ verankern, zollt ethischen Grundwertungen unserer Rechtsordnung in inhaltlich gebotener Weise Tribut: Sachgerechter Persönlichkeitsschutz bewahrt den Einzelnen vor einem Bewertungsautomatismus. Denn Computeralgorithmen können die Komplexität menschlichen Verhaltens (jedenfalls bislang) nicht hinreichend zuverlässig abbilden. Sie bewegen sich in einer *Hume'schen* Metaphysik,¹¹² die Kausalitäten nicht durch logische

107 Zur grundrechtlichen Sensibilität des Profilings siehe bspw. *Roßnagel* (Fn. 91), 565; *Weichert* (Fn. 14), 255.

108 BVerfGE 27, 1 (6).

109 Vgl. auch Erwägungsgründe Nr. 58 und 58a des Entwurfs der Datenschutz-Grundverordnung. Siehe dazu auch mit rechtspolitischen Vorschlägen *Härtig* (Fn. 12), 532 ff.

110 § 6a Abs. 1 S. 1 BDSG; Art. 20 Abs. 5 S. 1 Datenschutz-Grundverordnung-E.

111 Der Entwurf lässt allerdings qualitative Vorgaben für die Profilerstellung schmerzlich vermissen.

112 »Wenn aber viele gleichförmige Beispiele auftreten und demselben Gegenstand immer dasselbe Ereignis folgt, dann beginnen wir den Begriff von Ursache und Verknüpfung zu bilden. Wir empfinden nun ein neues Gefühl [...]; und dieses

Schlüsse, sondern durch statistische Zusammenhänge, also regelmäßig auftretende Kovariationen von Ereignissen, ermittelt (sog. Regularitätstheorie der Kausalität). Korrelationen deuten eine Wahrscheinlichkeit, die Berechtigung einer Richtigkeitsvermutung, dafür an, dass zwei Variablen kausal miteinander verknüpft sind. Sie weisen diesen Zusammenhang jedoch nicht objektiv nach. Auch Big-Data-Technologien fördern stochastische Korrelationen zutage, ergründen aber keine Kausalitäten.¹¹³

Ohne das Verständnis der Zusammenhänge bleibt die Korrelation regelmäßig wertlos. Denn nur mit deren Hilfe lassen sich die hinter der Korrelation liegenden Gesetzmäßigkeiten verstehen. Genau dies zeichnet Regeln und Gesetze nämlich aus: Sie sind geronnene Datenkompressionen von beobachteten Ursache- und normativ erwünschten Wirkungsbeziehungen, die ein erkanntes Anleitungsmuster zum Ausdruck bringen. Die Vermutungen, die algorithmische Zusammenhänge auslösen, und die Intuition, die sie implizieren, sind als solche freilich kein uneingeschränkt tauglicher Ratgeber. Die Ergebnisse algorithmischer Korrelationssuche sind interpretationsbedürftig und fehleranfällig. Es braucht immer eine Verbindung des Datenstroms mit Reflexion und damit eine Ergänzung durch die Kunst menschlicher Analyse.¹¹⁴ Denn Korrelationen verleiten auf der Grundlage trügerischer Sicherheit schnell zu »*cum hoc ergo propter hoc*«-Fehlschlüssen: So lässt sich zwar auf der Grundlage einer Analyse von Feuerwehrdatensätzen ein stochastisch valider Zusammenhang zwischen der Höhe der Brandschäden und der Zahl der im Einsatz herbeigerufenen Feuerwehrleute ausmachen:¹¹⁵ Je mehr Feuerwehrleute am Brandeinsatz mitwirken, desto größer der Schaden. Deshalb die Zahl der Feuerwehrleute zu reduzieren, wäre jedoch eine (einem Computeralgorithmus intuitiv naheliegende) fatale Verwechslung von Ursache und Wirkung. Denn je größer der Brand, umso mehr Feuerwehrleute kommen zum Einsatz – nicht umgekehrt.¹¹⁶ Ein Big-Data-Algorithmus würde womöglich auch vorschlagen, die Ansiedlung von Störchen in Niedersachsen zu fördern, um die Geburtenrate zu erhöhen. Immerhin korrelierte dort zwischen 1972 und 1985 die Storchpopulation mit der Geburtenrate. Auch dieser Vorschlag würde aber Ursache und Wirkung verwechseln: Denn die Geburtenrate steigt nicht wegen der Präsenz von Störchen, sondern weil dort, wo mehr Störche leben – auf dem Land – die Geburtenrate aufgrund bestimmter soziodemografischer Faktoren tendenziell höher ist als in der Stadt. Der Ursachenzusammenhang besteht also in der Urbanisierung, nicht in der Storchendichte.

Big-Data-Analysen fördern mitunter auch Zufallskorrelationen zutage, wie etwa den – unter Börsianern bekannten – Zusammenhang zwischen dem Anstieg von Börsenkursen und dem Ausgang des Super-Bowl-Endspiels.¹¹⁷ Ebenso lässt sich eine statistisch valide Korrelation zwischen der Anfälligkeit für Angina Pectoris und der Trägerschaft des Sternzeichens »Wassermann« ausmachen.¹¹⁸ Daraus einen Ursachenzusammenhang abzuleiten, wäre ein analytischer Fehlschluss. Über lange Zeitreihen bleibt der Ursachenzusammenhang nicht stabil. Algorithmen haben natürliche Erkenntnisgrenzen und ihre Ergebnisse Fehlerquoten. Mitunter erkennen sie Muster, die gar nicht existieren (sog. Apophänie).¹¹⁹ Sie identifizieren zwar

Gefühl ist das Urbild jener Vorstellung [von notwendiger Verknüpfung], das wir suchen« (*Hume*, Eine Untersuchung über den menschlichen Verstand, 12. Aufl., 1993, S. 95).

113 Vgl. insbesondere *Mayer-Schönberger/Cukier* (Fn. 3), S. 248.

114 Vgl. auch *Bachmann/Kemper/Gerzer* (Fn. 9), S. 340 ff.; *Hill*, DÖV 2014, 213 (216 f.) m. w. N.

115 Ähnlich korreliert auch die Verweildauer von Patienten im Krankenhaus mit einem schlechteren Gesundheitszustand nach dem Krankenhausaufenthalt. Auch hier fehlt jedoch ein Kausalzusammenhang.

116 Vgl. *Bachmann/Kemper/Gerzer* (Fn. 9), S. 272 f. mit weiteren Beispielen.

117 Vgl. *Kester*, *The Journal of Investing* 19 (2010), 82 ff.

118 *Austin/Mamdani/Juurlink et al.*, *Journal of Clinical Epidemiology* 59 (2006), 964 ff.

119 *Boyd/Crawford* (Fn. 36), 198; *Hill* (Fn. 114), 221.

mit unbestechlicher Genauigkeit Wahrscheinlichkeiten. Intuition und eine Methodik der Selbstkorrektur sowie die Erfassung individueller Besonderheiten sind ihnen (jedenfalls auf absehbare Zeit) nicht eigen. Unsicherheiten aufgrund fehlender Berechenbarkeit mancher Ereignisse (sog. schwarze Schwäne)¹²⁰ können sie nicht eliminieren.

Auch der komplexeste Algorithmus ist nicht in der Lage, den Menschen in seiner vielfältigen Struktur und Eigenwilligkeit autonomer Steuerung vollständig zu erfassen und ein detailgetreues Abbild der Realität und ihrer Ursachenzusammenhänge zu zeichnen. Menschliches Verhalten ist nicht nur vernunft-, sondern eben auch intuitions- und emotionsgeleitet – und damit im eigentlichen Wortsinne »unberechenbar«. Wären perfekte Vorhersagen menschlichen Verhaltens möglich, hätten wir unsere Fähigkeit verloren, unser Leben autonom in Freiheit und Verantwortung zu gestalten.

Die menschliche Hoheit über den Datenverarbeitungsprozess und die Nachvollziehbarkeit seiner Ergebnisse muss entsprechend zu jeder Zeit gewährleistet sein; der Mensch muss »die letzte Instanz« der Entscheidungskette bleiben. Die Unschärfen und Fehlerquellen sowie die immer bedeutsamer werdende Rolle von Algorithmen machen verfahrensrechtliche Schutzmechanismen erforderlich: Im Falle besonderer Risiken für die Rechte und Freiheiten der Betroffenen ist eine Möglichkeit der Vorabkontrolle¹²¹ von Algorithmen geboten. Handgreiflich wird das etwa am Beispiel des Steuerungsalgorithmus eines autonomen Fahrzeugs, der darüber entscheidet, ob das Fahrzeug das auf die Straße laufende Kind oder eine große Menschengruppe gefährdet. Algorithmen als Regieanweisungen der digitalen Welt dürfen – ähnlich wie die Verarbeitung besonderer Arten personenbezogener Daten und Personenbewertungen nach § 4d Abs. 5 BDSG – einer staatlichen Kontrolle nicht vollständig entzogen sein. Das gilt insbesondere im Hinblick auf das von ihnen ausgehende Diskriminierungspotenzial. Begrüßenswert ist insoweit auch der regulatorische Schulterschluss des Art. 20 Abs. 3 Datenschutzgrund-Verordnung-E mit ausgewählten besonderen Diskriminierungsverboten aus Art. 21 Abs. 1 GrCh (bzw. Art. 3 Abs. 3 GG). Der Entwurf der Vorschrift untersagt Profiling-Maßnahmen, die nach Rasse, ethnischer Herkunft, politischer Überzeugung, Religion oder Weltanschauung, Mitgliedschaft in einer Gewerkschaft, sexueller Orientierung oder Geschlechtsidentität diskriminieren. Die sich insoweit abzeichnende Schnittmenge zwischen freiheitsrechtlicher und gleichheitsrechtlicher Schutzdimension geht über das seit jeher im Datenschutz verortete Verbot der Verarbeitung besonderer Kategorien von personenbezogenen Daten in sachlich gebotener Weise hinaus. Sie vollzieht eine funktionelle Weiterentwicklung der informationellen Selbstbestimmung vom Daten- zum Profilschutz, der stärker als bisher auf die Abwehr datenbasierter Diskriminierungen abzielt.

b) Scoring

Zu einem sensiblen Paradigma der Zusammenführung großer Datenmengen für die Wahrscheinlichkeitsanalyse ist das Scoring avanciert. Es verdichtet einzelne Merkmalsvektoren auf der Grundlage von Zurechnungsfaktoren zu statistisch begründeten Prognosen, etwa der Ausfallwahrscheinlichkeit eines Kredits. Das Verfahren baut auf der Prämisse auf, dass es einen Zusammenhang zwischen bestimmten

120 Vgl. dazu *Taleb*, Der schwarze Schwan, 2012.

121 Vgl. zum bestehenden Rechtsrahmen von Vorabkontrollen im Rahmen des BDSG *Hallermann*, RDV 2015, 23 ff.

Merkmalswerten und einer bewerteten Eigenschaft bzw. einem Verhalten gibt: Es schließt vom Gruppenverhalten auf das Verhalten einzelner Merkmalsträger. Welche Verhaltensfaktoren das jeweils sind, bleibt für die Betroffenen häufig unklar.

Deshalb setzt § 28b BDSG¹²² der Analysemethoden ausdrückliche Schranken: Das Gesetz verlangt zur Herstellung von Transparenz ein wissenschaftlich anerkanntes mathematisch-statistisches Verfahren, das für die Berechnung der Wahrscheinlichkeit eines bestimmten Verhaltens nachweisbar erheblich ist (§ 28b Nr. 1 BDSG).¹²³

Viele Menschen treibt das Bedürfnis um, überprüfen zu können, ob der Erteilung kreditrelevanter Auskünfte eine rechtmäßige Analysemethoden zugrunde liegt. So sieht sich die SCHUFA immer häufiger der Forderung ausgesetzt, den Algorithmus offenzulegen, der den Scoring-Wert ermittelt. Nach dem Willen des Gesetzgebers unterliegt dieser aber dem Geheimnisschutz als vorrangigem Schutzgut.¹²⁴ Der Betroffene kann seine Offenlegung nicht verlangen. Er ist lediglich über das Zustandekommen und die Bedeutung der Wahrscheinlichkeitswerte einzelfallbezogen und nachvollziehbar in allgemein verständlicher Form zu unterrichten (§ 34 Abs. 2 S. 1 Nr. 3 i. V. m. Abs. 8 S. 1 und 2 BDSG). Dem Betroffenen will der Gesetzgeber die Möglichkeit eröffnen, Fehler in der Berechnungsgrundlage aufzudecken, nicht aber den Berechnungsschlüssel zu enträtseln. So hat es der BGH auch im Jahr 2014 entschieden.¹²⁵

Angesichts der unbändigen Macht, die Algorithmen über wichtige wirtschaftliche und private Entfaltungschancen der Bürger in einer digitalen Welt ausüben können, bleibt das hinter dem rechtspolitisch Wünschenswerten zurück. Zwar braucht und verdient die in Algorithmen steckende geistige Leistung einen Investitionsschutz.¹²⁶ Eine Pflicht zur Offenlegung schürt die Gefahr, die technologische Innovationsfreude nachhaltig zu dämpfen. Sachgerechter Persönlichkeitsschutz setzt – ähnlich wie nach § 4d Abs. 5 BDSG – aber eine Möglichkeit der Kontrolle solcher Algorithmen, jedenfalls aber der in die Bewertung eingeflossenen Datengrundlage (nicht nur der zur Berechnung der Wahrscheinlichkeitswerte genutzten Datenarten – § 34 Abs. 2 S. 1 Nr. 2 BDSG) auf ihre Validität voraus. Denn informationelle Selbstbestimmung bedingt ihrem Wesen nach die Möglichkeit der Kontrolle über das, was Dritte über die eigene Person wissen.¹²⁷ Eine wirksame Überprüfung der Prognosekraft eingesetzter Verfahren ist aber nur möglich, wenn diese selbst einer wirksamen Kontrolle unterliegen.

Algorithmen avancieren zum zentralen Steuerungsmechanismus unseres digitalen Ökosystems.¹²⁸ Sie entscheiden, welche Angebote wir erhalten, was wir lesen, mit wem wir in Kontakt treten. Das eröffnet ein

122 Für den Bereich des Kreditwesens, namentlich den Umgang mit personenbezogenen Daten im Zusammenhang mit Risikobemessungsverfahren, ist § 10 Abs. 2 KWG lex specialis. Vgl. dazu auch BT-Drucks. 16/1335, S. 48.

123 Vgl. zur Auslegung der Tatbestandsmerkmale und zur Evaluation des Gesetzes im Einzelnen bspw. *ULD Schleswig-Holstein*, Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen, 2014, 31 ff.; zur Kohärenz der Methodik OLG München, ZD 2014, 570 (572); vgl. auch OLG Frankfurt a. M., Urt. v. 7.4.2015 – 24 U 82/14 –, BeckRS 2015, 06846, Rn. 46.

124 BT-Drucks. 16/10529, S. 17: »[...] die Score-Formel, an deren Geheimhaltung die Unternehmen ein überwiegendes schutzwürdiges Interesse haben«.

125 BGHZ 200, 38 ff. zu dem Urt. vgl. bspw. *Moos*, Update Datenschutz, in: Taeger (Hrsg.), Big Data & Co, 2014, S. 525 (533 ff.).

126 BT-Drucks. 16/10529, S. 17; urheberrechtlichen Schutz genießen Algorithmen als solche jedoch grundsätzlich ebenso wenig (§ 69 Abs. 2 S. 2 UrhG) wie die Daten, die die Grundlage der Analyse bilden (§ 4 Abs. 1 Hs. 2 i. V. m. § 2 Abs. 2 UrhG); dazu auch *Dorner*, CR 2014, 617 (620 f.).

127 BVerfGE 65, 1 (43).

128 Vgl. dazu auch unten V. 1., S. 36 ff.

nicht unerhebliches Diskriminierungs- und Steuerungspotenzial, das Überprüfungsbedarf auslöst. Deshalb ist nicht nur Betroffenen ein Einblick in die grundsätzliche Gewichtung der in das Scoring einfließenden Faktoren zu gewähren. Es ist auch eine behördliche Algorithmenkontrolle jedenfalls in den Fällen geboten, die sensible Auswirkungen in persönlichkeits sensitiven Bereichen zeitigen. Ebenso wie ein blindes Vertrauen fehl am Platze ist, Google schaffe bei seinen Suchalgorithmen Chancengleichheit seiner eigenen und fremder Angebote oder Facebook richte seine Nachrichtenalgorithmen alleine am Sachgehalt der Nachrichten aus (und nicht an seinem Interesse, die Nutzer auf seiner Plattform zu halten sowie Werbeeinnahmen zu generieren), ist nicht per se sichergestellt, dass die Kriterien der SCHUFA entscheidungsrelevant sind und nicht – direkt oder indirekt – unzulässig diskriminieren.

Dass diese Sorge nicht allein bunter Fantasie entspringt, macht beispielhaft ein Patent deutlich, das sich Google im Jahr 2011 sicherte. Es ermöglicht *Dynamic Pricing* (bzw. *Cookie-Pricing*) im Onlinehandel, also die personenspezifische Preisdifferenzierung, die unter Verzicht auf allgemeine Preise¹²⁹ Zahlungsfähigkeits- und Profilaspekte in die individuelle Preisbildung einfließen lässt, etwa finanzkräftigen und zahlungsbereiten Apple-Kunden für das gleiche Produkt höhere Preise als anderen Kunden abringt.

Dass in einer auf der Analyse großer Datenmengen aufbauenden digitalen Welt viele weitere algorithmisch gesteuerte Differenzierungsmodelle umsetzbar werden, macht deutlich: Es braucht eine Neutralitäts- und eine Persönlichkeitskompatibilitätskontrolle von Algorithmen. Die bestehenden datenschutzrechtlichen Vorschriften bedürfen daher einer Konkretisierung und Ausdifferenzierung, die den Herausforderungen algorithmischer Steuerungsmacht eine rechtliche Kontrollmacht entgegensetzt. Um den gleichzeitig erforderlichen Schutz von Geschäftsgeheimnissen sicherzustellen, ist eine Ausgestaltung der Algorithmenkontrolle als In-camera-Verfahren denkbar und geboten. (Entsprechend technisch und personell zu rüstende) Datenschutzbeauftragte bzw. andere geeignete Kontrollinstanzen erlangen dann exklusiven Einblick in die Algorithmenstruktur. Sie können diese dadurch auf ihre Vereinbarkeit mit den Grundrechten und dem geltenden Datenschutzrecht überprüfen. Grundsätzlich ist insoweit eine Befugnis zur Ex-post-Kontrolle ausreichend. Nur in sensiblen Bereichen, etwa Analysen, die religiöse Überzeugungen, Gesundheitsdaten oder sonstige besondere Arten personenbezogener Daten betreffen (§ 3 Abs. 9 BDSG), ist eine Ex-ante-Kontrolle angezeigt. Dass ein solches Überwachungsregime nicht leicht realisierbar ist, entbindet den Staat nicht von seiner Verantwortung, wirksamen Persönlichkeitsschutz zu verbürgen.

2. Allgemeine Zulässigkeitstatbestände

Jenseits Big-Data-spezifischer Tatbestände, insbesondere der §§ 6a und 28b BDSG, richtet sich die Zulässigkeit von Big-Data-Analysen nach den allgemeinen Regeln: Es gilt das Verbotsprinzip des Datenschutzrechts. Eine Verarbeitung personenbezogener Daten bedarf daher entweder einer informierten Einwilligung (a) oder einer gesetzlichen Verarbeitungserlaubnis (b) – § 12 Abs. 1 TMG, § 4 Abs. 1 BDSG.

129 Der allgemeine Preis ist im historischen Vergleich allerdings (auch in Deutschland) eine eher junge Erscheinung. Erst mit dem Aufstieg von Kaufhäusern zum Ende des neunzehnten Jahrhunderts haben sich einheitliche, klar ausgewiesene Preise allenthalben durchgesetzt. Bis dahin waren unterschiedliche Preise für unterschiedliche Gesellschaftsschichten Teil des Einkaufsalltags.

a) Einwilligung

Eine informierte Einwilligung muss (sowohl nach dem bestehenden BDSG als auch nach der geplanten Datenschutz-Grundverordnung¹³⁰) für den konkreten Fall und in Kenntnis der Sachlage »für einen oder mehrere genau festgelegte Zwecke« erteilt werden (Art. 6 Abs. 1 lit. a Datenschutz-Grundverordnung-E; ähnlich § 4a Abs. 1 S. 2 BDSG). Denn erst dann kann der Betroffene das Risiko einschätzen, wie sich eine spätere Verarbeitung der Daten auf seine Persönlichkeit auswirkt und ob die Vorteile die Nachteile überwiegen.

Zu dem Zeitpunkt, zu dem eine Einwilligung in die Verarbeitung von Massendaten durch Big-Data-Algorithmen erfolgt, sind die Folgen in der Regel jedoch nicht absehbar und die Erhebungszwecke nicht ohne Weiteres bestimmt.¹³¹ Schließlich zeichnen sich Big-Data-Analysen ihrem Wesen nach gerade dadurch aus, Daten für vorher nicht ausdrücklich bestimmte Zwecke zu verwenden.

Zwar setzt eine wirksame Einwilligung nicht voraus, dass das Ergebnis der *Auswertung* bereits vor der Erlaubniserteilung feststeht. Denn zur Auswertung sollen die Daten ja verarbeitet werden.¹³² Feststehen muss aber der Zweck der *Erhebung* (§ 4a Abs. 1 S. 1 BDSG). Eine Einwilligung in Big-Data-Analysen muss daher den Analysezweck hinreichend klar benennen.¹³³ Allgemeine Zweckangaben, wie z.B. »zum Zwecke der Werbung«, sind zu unbestimmt.¹³⁴ Eine Einwilligung in die Zusammenführung unterschiedlicher Datenbestände zur Generierung eines Profilbildes ist – als Teil allgemeiner Nutzungs- oder Datenschutzbestimmungen – regelmäßig eine unzulässige überraschende Klausel im Sinne des § 305c Abs. 1 BGB.¹³⁵

Um die Einwilligung zu erlangen, darf der Anbieter auch keinen unangemessenen Druck ausüben. Er darf sein Leistungsangebot insbesondere nicht ohne Weiteres von der Zustimmung des Nutzers zur weiteren Datenanalyse abhängig machen. Das gilt jedenfalls für Anbieter nicht substituierbarer Dienste, insbesondere der Daseinsvorsorge, z.B. der Energieversorgung oder der Telekommunikation. Sie sind einem Koppelungsverbot unterworfen, das Anbietern den Missbrauch ihrer Macht versagen soll. § 95 Abs. 5 TKG und § 28 Abs. 3b BDSG verankern das ausdrücklich¹³⁶ (vgl. auch Art. 7 Abs. 4 S. 2 Datenschutz-Grundverordnung-E). Die Vorschriften sind zwar normativ als Ausnahmetatbestand konzipiert, drücken aber einen allgemeinen, analogiefähigen Rechtsgedanken des Datenschutzrechts aus: Notwendige Voraussetzung der Einwilligung sind Entscheidungsfreiheit und Freiwilligkeit, die nicht unter dem Damoklesschwert einer »coactus volui«-Situation stehen dürfen.¹³⁷

130 Erwägungsgrund Nr. 25, Art. 7 Abs. 1 Datenschutz-Grundverordnung-E.

131 Roßnagel (Fn. 91), 564; ähnlich auch Baeriswyl, *digma* 2013, 14 (16); Weichert (Fn. 14), 256.

132 In diesem Sinne zu Recht Ulmer (Fn. 12), 229.

133 Vgl. dazu auch oben III. 2., S. 20.

134 Roßnagel (Fn. 91), 564.

135 Zur Unwirksamkeit von Datenschutzerklärungen wegen unangemessener Benachteiligung des Kunden nach § 307 Abs. 1 i. V. m. Abs. 2 Nr. 1 BGB, §§ 4, 4a BDSG, § 12 f. TMG; LG Berlin, NJW 2013, 2605; MMR 2014, 563 ff.

136 Martini (Fn. 85), 232 f.; Weichert (Fn. 12), 139.

137 A. A. Ohrtmann/Schwiering, NJW 2014, 2984 (2988) m. w. N.; Rogosch, Die Einwilligung im Datenschutzrecht, 2013, S. 186. Rogosch weist zu Recht darauf hin, dass der systematisch passende Ort für ein allgemeines Koppelungsverbot die Vorschrift des § 4a BDSG gewesen wäre. Daraus ergibt sich aber noch nicht zwingend das Fehlen einer Regelungslücke. Vielmehr hat der Gesetzgeber in Fällen der Datenerhebung und -speicherung für eigene Zwecke die größte Missbrauchstypik unzulässiger Kopplung in der Praxis vermutet und daher andere Sachbereiche nicht ausdrücklich einer Regelung zugeführt. Nichts spricht dafür, dass der Gesetzgeber öffentliche Stellen von einem solchen Koppelungsverbot bewusst freistellen wollte. Die besseren Gründe streiten insoweit für eine planwidrige, im Wege eines Analogieschlusses schließungsbedürftige Regelungslücke.

De lege ferenda sollte sich die Einwilligung explizit auch auf die Zusammenführung unterschiedlicher Daten aus verschiedenen Quellen und ihre Auswertung *mithilfe von Algorithmen* beziehen. So lassen sich den Nutzern der erhöhte Gefährdungsgrad und das Ausmaß der Einwilligung bewusst machen. Einwilligungen sollten (jedenfalls) für Big-Data-Analysen auch grundsätzlich zeitlich befristet sein, um dem Risiko der Entkopplung von aktuellen Persönlichkeitspräferenzen zu begegnen.

b) Gesetzliche Verarbeitungsbefugnis

Soweit eine Big-Data-Analyse nicht von der Einwilligung des Betroffenen getragen ist, ist sie auf eine gesetzliche Verarbeitungserlaubnis als Rechtfertigungsgrundlage angewiesen. Die Rechtsordnung steckt der Erhebung, Verarbeitung und Nutzung personenbezogener Daten insoweit einen engen Rahmen. Dessen Spannweite hängt von der Art der Daten ab, die in die Analyse einfließen sollen. Zu unterscheiden ist zwischen Daten, die zur technischen Übertragung von Telekommunikationsinhalten generiert werden, insbesondere *Standortdaten* (§ 3 Nr. 19 TKG, aa), telemedienrechtlichen *Bestands- und Nutzungsdaten*, welche die Inanspruchnahme von Telemedien im Verhältnis zwischen Diensteanbieter und Nutzer ermöglichen oder abrechnen sollen (bb) sowie *Inhaltsdaten* eines Kommunikationsprozesses, etwa eines E-Mail-Austausches, eines Chats oder eines Postings (cc). Den Zugriff auf Standortdaten regelt das TKG, denjenigen auf Bestands- und Nutzungsdaten das TMG und auf Inhaltsdaten das BDSG.¹³⁸

aa) Standortdaten

Für Standortdaten, etwa von Smartphones, limitieren die §§ 96 und 98 TKG die Verarbeitungszwecke. Sie machen das Datengold digitaler Mobilität nur zum Zwecke der Nachrichtenübermittlung, deren Abrechnung und zur Erbringung eines angeforderten Dienstes mit Zusatznutzen, etwa eines Lokalisationsdienstes, zugänglich – und dies auch nur so lange, wie es zeitlich erforderlich ist (§ 96 Abs. 1 S. 2 und Abs. 3; § 98 Abs. 1 S. 1 und Abs. 4 TKG¹³⁹). Im Anschluss sind die Daten zu löschen (§ 96 Abs. 1 S. 3 TKG). Eine allgemeine und breite personenbezogene Datenauswertung, etwa für Zwecke der Werbung oder des Direktmarketings, das eine umgreifende Auswertung mit dem Ziel der Konturierung eines Personenprofils vornimmt, ist unzulässig.¹⁴⁰

138 Vgl. zur Abgrenzung zwischen diesen Regelungsmaterien auf der Grundlage des Schichtenmodells etwa *Martini*, Datenschutz und Sicherheit bei der elektronischen Rechnung, in: Rogall-Grothe (Hrsg.), Leitfaden Elektronische Rechnung in der öffentlichen Verwaltung, 2014, S. 51 (72 f.); *Spindler/Nink*, in: Spindler/Schuster (Hrsg.), Recht der elektronischen Medien, 3. Aufl., 2015, § 15 TMG, Rn. 3.

139 Zur Abgrenzung zwischen den drei Normen *Martini/Weiß/Ziekow*, Rechtliche Zulässigkeit flächendeckender Alarmierungen der Bevölkerung in Katastrophenfällen per SMS (KatWarn), 2013, S. 96 ff.

140 *Ulmer* (Fn. 12), 229.

bb) Telemedienrechtliche Bestands- und Nutzungsdaten

Vordergründig größeren Auswertungs- und Verarbeitungsspielraum gesteht das TMG den Anbietern von Telemediendiensten, etwa Facebook, Google oder XING, zu. Sie dürfen für Zwecke der Werbung und der bedarfsgerechten Gestaltung der Telemedien pseudonymisierte Nutzungsprofile erstellen (§ 15 Abs. 3 S. 1 TMG). Davon machen die Diensteanbieter umfänglich Gebrauch. Zulässig sind aber nur *Nutzungsprofile*,¹⁴¹ keine übergreifenden *Persönlichkeitsprofile*.¹⁴² In die Nutzungsprofile dürfen auch nur die Daten einfließen, welche für die Inanspruchnahme des Telemediums oder der Abrechnung seiner Leistungen erforderlich sind (vgl. § 15 Abs. 1 S. 1, Abs. 4 S. 1 TMG).

Personenbezogene Daten sind nach Ablauf des Zugriffs zu löschen (§ 13 Abs. 4 S. 1 Nr. 2 TMG). Eine Zusammenführung mit anderen Daten über den Nutzer – wie sie Big Data konzeptionell immanent ist – ist ausdrücklich ausgeschlossen (§ 15 Abs. 3 S. 3, § 13 Abs. 4 S. 1 Nr. 6 TMG).

Die Erhebung und Verwendung von *Bestandsdaten*, etwa Name, Anschrift, E-Mail, Zahlungsdaten und Geburtsdatum, sind nur so weit zulässig, wie sie für die Begründung, inhaltliche Ausgestaltung oder Änderung des Vertragsverhältnisses (§ 14 Abs. 1 TMG) erforderlich sind oder eine staatliche Stelle auf der Grundlage des § 14 Abs. 2 TMG an den Telemedienanbieter aus Gründen der Gefahrenabwehr, etwa zur Unterbindung von Anschlägen oder der Strafverfolgung, ein Auskunftsverlangen richtet.¹⁴³

cc) Inhaltsdaten

Für Inhaltsdaten, also den Inhalt der Kommunikation (z.B. den Text einer E-mail oder einer Facebook-Nachricht), bestimmt sich die Zulässigkeit der Verarbeitung grundsätzlich nach §§ 28 ff. BDSG, für öffentliche Stellen nach den §§ 13 ff. BDSG bzw. dem jeweiligen LDSG.

(1) Verarbeitung für eigene Geschäftszwecke – § 28 BDSG

Sowohl das deutsche (§ 28 Abs. 1 Nr. 2 BDSG) als auch das unionale Recht (Art. 6 Abs. 1 lit. f der Datenschutz-Grundverordnung-E) machen die Zulässigkeit der Verarbeitung im Wesentlichen davon abhängig, ob das Auswertungsinteresse gegenüber dem schutzwürdigen Interesse des Betroffenen überwiegt.¹⁴⁴

141 Nutzungsprofile beschränken sich auf die Auswertung von Nutzungsdaten i. S. d. § 15 Abs. 1 S. 2 TMG. Sie ermitteln insbesondere typische Nutzungszeiten und Verweildauern, sind aber nicht darauf gerichtet (und nicht in der Lage), die Persönlichkeit des Nutzers umfassend zu kategorisieren.

142 *Dix/Schaar*, in: Roßnagel (Hrsg.), *Recht der Telemediendienste*, 2013, § 15 TMG, Rn. 62. A. A. wohl *Ohrtmann/Schwiering* (Fn. 137), 2986 f. Zur Speicherung dynamischer IP-Adressen und ihrer Vereinbarkeit mit dem unionalen Datenschutzrecht siehe den Vorlagebeschluss des BGH, GRUR 2015, 192 ff.

143 Zu verfassungsrechtlichen Bedenken an der Vorschrift *Dix*, in: Roßnagel (Hrsg.), *Recht der Telemediendienste*, 2013, § 14 TMG, Rn. 47 ff.

144 Vgl. *Weichert* (Fn. 12), 139.

(α) Abwägung zwischen Auswertungsinteresse und Persönlichkeitsinteresse

Die Interessen des Betroffenen überwiegen umso eher, je heterogener die verarbeiteten Daten, je systematischer und je umfänglicher die Auswertung ist.¹⁴⁵ Das Interesse an einer Auswertung über verschiedene, den ursprünglichen Kontext verlassende Datentypen hinweg hat grundsätzlich hinter dem Persönlichkeitsrecht zurückzustehen. Das gilt insbesondere bei Auswertungen, die auf eine Entscheidung mit erheblichen rechtlichen Folgen oder erheblichem Beeinträchtigungspotenzial, z.B. ein Persönlichkeitsprofil, (§ 6a BDSG) sowie auf besondere Arten personenbezogener Daten, wie ethnische Herkunft oder politische Meinung (§ 3 Abs. 9 BDSG i. V. m. § 4d Abs. 5 S. 2 Nr. 1 BDSG), gerichtet sind.¹⁴⁶ Verantwortlichen Stellen ist es daher untersagt, Angaben für nicht näher definierte spätere Nutzungsmöglichkeiten in eine beliebig verwendbare Datensammlung einzubringen. Kundenprofile dürfen sie nur mit Einwilligung des Betroffenen erstellen.¹⁴⁷

Ein überwiegendes Auswertungsinteresse vermutet das Gesetz demgegenüber grundsätzlich bei allgemein zugänglichen Daten (§ 28 Abs. 1 S. 1 Nr. 3 BDSG, § 29 Abs. 1 S. 1 Nr. 2 BDSG, § 30a Abs. 1 S. 1 Nr. 2 BDSG). Darunter fallen zwar keine Facebook-Nachrichten an Freunde, sehr wohl aber viele klassische Big-Data-Internetquellen, etwa Tweets, in öffentlichen Internetforen verfügbare Daten und allgemein zugängliche Pinnwandeinträge bei Facebook sowie öffentliche Daten der Facebook-Profile,¹⁴⁸ wenn die Privatsphäreneinstellungen den Leserkreis solcher Informationen und Mitteilungen also nicht einschränken.

(β) Schutz des Vergessenwerdens

Sobald die Kenntnis personenbezogener Daten für den Zweck, zu dem sie gespeichert wurden, nicht mehr erforderlich ist, sind die Daten zu löschen (§ 35 Abs. 2 Nr. 3 BDSG). Für die Verarbeitung zum Zwecke der Übermittlung, etwa zum Abruf durch Dritte, trifft den Verantwortlichen eine periodische Prüfungspflicht nach drei bzw. vier Jahren (§ 35 Abs. 2 S. 2 Nr. 4 BDSG).

Der automatisierten Verarbeitung im Wege von Big-Data-Analysen kann der Betroffene nach § 35 Abs. 5 BDSG auch widersprechen. Zwei Hürden gilt es dafür allerdings zu überwinden: Der Betroffene muss von der Verarbeitung erfahren und sein schutzwürdiges Lösungsinteresse muss gerade wegen seiner *besonderen persönlichen Situation* (nicht nur wegen eines Allgemeininteresses an der Geheimhaltung von Daten) überwiegen (§ 35 Abs. 5 S. 1 BDSG). An beidem wird es häufig mangeln. § 35 Abs. 5 BDSG ist insoweit für den Schutz gegen Big-Data-Analysen ein relativ stumpfes Schwert.

Auch für Daten, die *bei Dritten* gespeichert sind, helfen dem Betroffenen weder § 35 Abs. 2 noch Abs. 5 BDSG. Insoweit hat der Gesetzgeber zwar § 35 Abs. 7 BDSG als Schutzinstrument angelegt: Er begründet (in den Grenzen der Verhältnismäßigkeit und der Schutzwürdigkeit der Betroffeneninteressen) eine Unterrichtungspflicht der datenverarbeitenden verantwortlichen Stelle gegenüber den Dritten. Sein Schutz erschöpft sich aber in deren Unterrichtung über eine *erfolgte eigene* Berichtigung, Sperrung sowie Löschung oder Sperrung wegen Unzulässigkeit der Speicherung. Eine Pflicht zur Einwirkung auf Dritte (damit

145 Ebenso etwa Venzke-Caprese, DuD 2013, 775 (779); Koch (Fn. 104), 16.

146 Weichert, in: Däubler/Weichert/Wedde (Hrsg.), BDSG, 4. Aufl., 2014, Einl., Rn. 45 f.

147 BITKOM (Fn. 12), S. 44; Roßnagel (Fn. 91), 565.

148 Venzke-Caprese (Fn. 145), 776.

diese insbesondere eine Löschung von Daten vornehmen) löst sie aber ebenso wenig aus wie Handlungspflichtigen Dritter.

Weiter geht demgegenüber insoweit Art. 17 der geplanten Datenschutz-Grundverordnung. Auf seiner Grundlage sollen Betroffene die Löschung personenbezogener Daten sowie diesbezüglicher Querverweise und die Unterlassung der weiteren Verbreitung nicht nur von dem für die Verarbeitung Verantwortlichen, sondern auch von Dritten u.a. dann verlangen können, wenn die Daten für die ursprünglichen Erhebungs- und Verarbeitungszwecke nicht mehr notwendig sind.¹⁴⁹

In gleichsam voraussetzendem Gehorsam zum Entwurf der Datenschutz-Grundverordnung hat der EuGH bereits aus der bestehenden Gesetzeslage (in dogmatisch kühner Weise) ein Recht auf Vergessenwerden herausgelesen.¹⁵⁰ Er verlangt von Suchmaschinenbetreibern als Big-Data-Generatoren auf Antrag die Entfernung solcher – ursprünglich rechtmäßiger – Einträge, deren Verarbeitung Inhalte (insbesondere in Anbetracht der inzwischen verstrichenen Zeit) nach Abwägung aller betroffenen Interessen das Recht auf Achtung des Privatlebens (Art. 7 GrCh) oder das Recht auf Schutz personenbezogener Daten (Art. 8 GrCh) verletzen.¹⁵¹ Der EuGH reagiert damit auf den Befund, dass Suchmaschinenbetreiber die Informationsströme im Netz immer stärker kanalisieren; sie nehmen eine Rolle als Gatekeeper des Informationszugangs wahr. Dass das Gericht die Entscheidung über die Kollision zwischen dem Informationsinteresse der Allgemeinheit und dem Persönlichkeitsschutz des Einzelnen primär in die Obhut eines privaten Anbieters – hier der Google Inc. – legt, ist jedenfalls rechtspolitisch unbefriedigend. Zwar ist eine solche Abwägungsentscheidung zwischen Persönlichkeitsinteresse und Informationsinteresse der Öffentlichkeit im Äußerungsrecht, etwa bei der Veröffentlichung von Presseartikeln, nichts Ungewöhnliches. Die Konfliktentscheidung ist bei dem Suchmaschinenbetreiber allerdings nicht in den richtigen Händen. Entsprechend dem Gebot effizienter Konfliktbewältigung sollte die Austragung der Auseinandersetzung vielmehr primär in die Hände derjenigen gelegt (und bei ihnen konzentriert) werden, welche der Konflikt unmittelbar berührt und die über die insoweit maßgeblichen Informationen verfügen: des Äußernden und des Betroffenen. So hat die Rechtsordnung die grundrechtliche Kollisionslage auch in der Vergangenheit bewältigt. Im Internet tritt zwar mit den Suchmaschinenbetreibern als Informationsintermediären zu den bereits bestehenden Kommunikationsebenen eine neue hinzu. Vorrangig sollte aber der Inhaltsanbieter (mit entsprechenden Einwirkungspflichten) dafür Sorge tragen müssen, dass persönlichkeitsverletzende Informationen (z.B. mit Hilfe des Ausschlussprotokolls „robot.txt“) über Suchmaschinen nicht auffindbar sind. Den Suchmaschinenbetreiber trifft dann nur eine subsidiäre Verantwortung, soweit ein Vorgehen gegen den Inhalteanbieter leerläuft, der Betroffene also sonst vollständig schutzlos gestellt wäre.

149 Art. 17 Abs. 1 lit. a Datenschutz-Grundverordnung-E. Etwas anderes gilt aber dann, wenn die Speicherung zur Ausübung des Rechts auf freie Meinungsäußerung, aus Gründen des öffentlichen Interesses im Bereich der öffentlichen Gesundheit, für historische und statistische Zwecke oder zum Zwecke der wissenschaftlichen Forschung oder zur Erfüllung einer legitimen gesetzlichen Pflicht zur Vorhaltung der personenbezogenen Daten erforderlich ist, Art. 17 Abs. 3 Datenschutz-Grundverordnung-E.

150 EuGH, NVwZ 2014, 857 ff. Dazu insbesondere *Boehme-Neßler*, NVwZ 2014, 825 ff.; *Nolte*, NJW 2014, 2238 ff.; *Piltz*, PinG 2 (2014), 180 ff. sowie bereits *Ambrose*, Stanford Technology Law Review 16 (2013), 369 ff.; *Hornung/Hofmann*, JZ 2013, 163 ff.; *Kodde*, ZD 2013, 115 ff.; *Mayer-Schönberger*, Delete, 2. Aufl., 2011; *Walker*, Hastings Law Journal 64 (2012), 257 (269 ff.). Aus der nationalen Rechtsprechung: LG Hamburg, NJW 2015, 796 ff.; LG Heidelberg, MMR 2015, 348 ff.; LG Berlin, ZUM-RD 2015, 276 f.

151 EuGH, NVwZ 2014, 857 (864, Rn. 92 f.).

Der Suchmaschinenbetreiber ist als Informationsintermediär weder dazu bestimmt noch dazu legitimiert, komplexe, für die demokratischen Grundlagen einer Gesellschaft wesentliche Abwägungsentscheidungen zwischen Informationsinteresse der Allgemeinheit und Persönlichkeitsinteresse des Einzelnen als Clearingstelle der digitalen Identität unbefangen und sachgerecht zu treffen. Als wirtschaftliches Unternehmen, das durch die Löschung auch in seinem (durch Art. 16 GrCh grundrechtlich geschützten) wirtschaftlichen Interesse betroffen ist und insoweit immer auch in eigener Sache entscheidet, ist es geneigt, die Interessen der Allgemeinheit am Informationszugang und derer, deren Inhalte aufgrund einer Löschung unter einer bestimmten Begriffssuche nicht mehr automatisch verlinkt werden, systematisch unterzugewichten. Die Idee einer plural zusammengesetzten, gerichtsähnlichen Schlichtungsstelle, welche das komplexe Geflecht der konfligierenden Interessen unparteiisch abbildet – und damit auch die Richtigkeit der zu treffenden Abwägungsentscheidung eher Gewähr bietet, hat deshalb Charme. Solche Mittlungsinstanzen haben sich in anderen Rechtsbereichen, etwa im Presse- und Verbraucherschutzrecht, bewährt.

Zum Schutz des Einzelnen davor, an überholten Persönlichkeitsmustern festgehalten zu werden, sollte der Gesetzgeber darüber hinaus einer Auswertung offener, allgemein zugänglicher Daten zu Big-Data-Zwecken entweder eine zeitliche Verwertungsschranke ziehen (Verfallsdatum)¹⁵² oder der Verarbeitung älterer Daten einen abdiskontierten Gewichtungsfaktor auferlegen. Jedenfalls sensible personenbezogene Daten, deren Generierung so lange zurückliegt, dass ihre Aussagekraft eingeschränkt ist – denken lässt sich (in Analogie zu § 12 Abs. 3 BVerfSchG oder § 35 Abs. 2 S. 1 Nr. 4 BDSG) an eine Zeitraum von z.B. drei bis fünf Jahren –, sollten grundsätzlich keine oder nur noch schwächer gewichtete Verwendung für Big-Data-Anwendungen finden dürfen. Absichern lässt sich das durch entsprechende Schlüssel als Metadaten, welche den Daten als zweckbestimmende »tags« mit auf den Weg gegeben werden. Das stellt eine Abnabelung des digitalen Alter Egos von der realen Person – insbesondere unabhängig von der individuellen Kenntnis oder Ausübung von Antragsrechten – am ehesten sicher.

(2) Datenverarbeitung öffentlicher Stellen

Noch größeren Einschränkungen der Verarbeitungsbefugnis als Private sind öffentliche Stellen unterworfen.¹⁵³ Zwar kann die Auswertung von Massendatensätzen substanziellen Gemeinwohlertrag spenden. Das gilt für die öffentliche Infrastrukturplanung, etwa von Bildungseinrichtungen, Nahverkehrsleistungen oder Energieversorgungsprojekten, ebenso wie für die Wirtschaftsförderung, die Bekämpfung des Missbrauchs von Sozialleistungen oder die Korruptionsbekämpfung.¹⁵⁴

Staatlichen Stellen setzen jedoch nicht nur die Integrität und Vertraulichkeit informationstechnischer Systeme besondere verfassungsrechtliche Handlungsgrenzen,¹⁵⁵ sondern auch die Bindung an die Wahrnehmung eigener Aufgaben und der Grundsatz der Zweckbindung (vgl. auch einfachgesetzlich § 13 Abs. 1 und § 14 Abs. 1 BDSG sowie die entsprechenden Landesdatenschutzgesetze): Sie koppeln

152 Dazu *Mayer-Schönberger* (Fn. 150), S. 201 ff. sowie in der bestehenden Rechtsordnung bspw. § 882e Abs. 1 ZPO, § 3 InsoBekV und § 46 BZRG.

153 Die geplante Datenschutz-Grundverordnung differenziert in ihrem bisherigen Entwurf demgegenüber in der Strenge der Datenverarbeitungsbefugnis nicht zwischen öffentlichen und privaten Stellen. Das wird den unterschiedlichen Bindungen nicht gerecht.

154 Vgl. dazu im Einzelnen bereits oben S. 11 ff.

155 BVerfGE 120, 274 ff.

die Datenspeicherung, -veränderung und -nutzung an den jeweils eigenen gesetzlichen Aufgabenbereich und den ursprünglichen Erhebungszweck. Das limitiert die Instrumentalisierung von Daten für Big-Data-Zwecke nachhaltig. So wäre es zwar verlockend, Straßenmaut-Daten auch für die Verbrechensaufklärung einzusetzen. Zu diesem Zweck wurden die Daten aber nicht erhoben, sondern ausschließlich für die Mautabwicklung und -kontrolle (§ 7 Abs. 2 S. 2 Bundesfernstraßenmautgesetz). Nur wenn der Gesetzgeber ihre Verwertung für andere Zwecke ausdrücklich und hinreichend klar zulässt, ist der Weg für entsprechende Nutzungen frei. Wenn der Staat Informationen, die er zur Verfügung hat, nicht nutzt, sondern ggf. noch einmal erheben muss, wirkt das prima facie zwar geradezu anachronistisch. So zeichnete beispielsweise die fehlende Koordination der Verfassungsschutzbehörden dafür mitverantwortlich, der NSU-Gruppe unbehelligtes Wirken zu ermöglichen. Einem allgemeinen, zuständigkeitsüberschreitenden Informationsaustausch, der jegliche Informationsgrenzen abbaut, will die Trennung und Begrenzung von Zuständigkeiten aber gerade einen Riegel vorschieben. Die föderale und fachliche Aufgliederung von Behörden entfaltet insoweit eine grundrechtliche Nebenwirkung.¹⁵⁶ Daraus erwächst ein informationelles Trennungsprinzip, etwa für den Austausch zwischen offen arbeitenden Polizeibehörden und verdeckt arbeitenden Nachrichtendiensten. Der Verfassungsschutz darf sich gewünschte Informationen nicht über den Umweg der Polizei beschaffen. So hat es das BVerfG in seiner Entscheidung zur Antiterrordatei ausdrücklich verfügt.¹⁵⁷ In der Literatur ist der Topos einer »informationellen Gewaltenteilung« schon lange bekannt.¹⁵⁸ Auch die Landesdatenschutzgesetze sehen ausdrücklich eine getrennte Verarbeitung zu unterschiedlichen Zwecken erhobener Daten vor (vgl. etwa § 9 Abs. 2 S. 2 Nr. 8 RhPflDSG).

3. Anonymisierung als Königsweg

In vielen Fällen bietet das Datenschutzrecht für Big-Data-Anwendungen gegenwärtig keine gesetzliche Verarbeitungsgrundlage. Es eröffnet ihren Einsatz dann nur in zwei Gestaltungsformen: Auf der Grundlage einer Einwilligung¹⁵⁹ oder durch Auflösung des Personenbezugs im Wege der Anonymisierung. Diese ist datenschutzrechtlich der Königsweg des Einsatzes von Big-Data-Techniken.¹⁶⁰ Sie verschließt dem BDSG den normativen Geltungsanspruch (§ 1 Abs. 1 und 2 BDSG). Eine Pseudonymisierung (§ 3 Abs. 6a BDSG) genügt insoweit nicht. Denn sie lässt die Möglichkeit der Zuordnung zu einer Person unangetastet, ermöglicht namentlich eine Rekonstruktion des Personenzusammenhangs.

156 BVerfGE 133, 277 (323, Rn. 113) – Antiterrordatei.

157 BVerfGE 133, 277 (329, Rn. 123) – Antiterrordatei; dazu *Fremuth*, AöR 139 (2014), 32 (34 ff.). Zum Trennungsgebot zwischen Polizei und Geheimdiensten bereits *Bull*, PinG 1 (2013), 6 f.; *Gusy*, Die Verwaltung 24 (1991), 467 ff.

158 Dazu etwa *Forgó/Krügel/Rapp*, Zwecksetzung und informationelle Gewaltenteilung, 2006.

159 Dazu oben S. 27.

160 In diesem Sinne auch ausdrücklich *Ulmer* (Fn. 12), 229. Die anonymisierte Verwendung bereits erhobener Daten ist auch kein neuer Verwendungszweck im Sinne des BDSG, der einer eigenen Verarbeitungsgrundlage bedürfte. Das entspricht auch Schutzrationalität des Datenschutzrechts, Beeinträchtigungen des Persönlichkeitsrechts durch den Umgang mit personenbezogenen Daten nach Möglichkeit zu vermeiden (§ 1 Abs. 1 BDSG). Vgl. auch *Dorner* (Fn. 126), 628 m. w. N. Allerdings herrscht auch unter Fachleuten keine Einigkeit, ab wann ein Datum anonym oder pseudonym ist, vgl. *Stöcker* (Fn. 31), 9.

Anonymisierungsmethoden müssen allerdings zuverlässig sicherstellen, dass eine Zuordnung zu einer Person mit nach menschlichem Ermessen hinreichender Wahrscheinlichkeit (wenn auch nicht notwendig absoluter Sicherheit)¹⁶¹ ausgeschlossen ist. Der Datensatz darf also eine inhaltliche Aussage über eine bestimmte natürliche Person entweder gar nicht zulassen (absolute Anonymisierung) oder nur mit unverhältnismäßigem Aufwand an Zeitkosten und Arbeitskraft (faktische Anonymisierung) – § 3 Abs. 6 BDSG. Entscheidend sind insoweit das Wissen und die Möglichkeiten der verantwortlichen Stelle. Die rein hypothetische Möglichkeit, die Person identifizieren zu können, reicht für die Anwendbarkeit des BDSG zwar noch nicht; sehr wohl ist aber Zusatzwissen zu berücksichtigen, über das die verantwortliche Stelle – ggf. unter Rückgriff auf Dritte – mit »vernünftigerweise« eingesetzten Mitteln verfügen kann (vgl. Erwägungsgrund Nr. 26 S. 2 der EG-Datenschutz-RL 95/46/EG).

a) Anonymisierung unter Big-Data-Bedingungen

Anonymisierte Daten können unter Umständen durch ihre massenhafte Verknüpfung und Filterung mit Hilfe von Big-Data-Technologien Rückschlüsse auf konkrete Personen zulassen, individuelle Lebensgewohnheiten offen legen und Lebensentwürfe vermessbar machen. Je größer die Gruppen, umso geringer grundsätzlich zwar das Risiko einer Reidentifizierung. Je umfangreicher und detaillierter die Merkmalsdaten sind und je länger Bewegungsmuster gespeichert werden, umso stärker steigt die Wahrscheinlichkeit, dass ein Abgleich der Merkmalsdaten eine Reidentifizierung ermöglicht.¹⁶² Es genügen häufig wenige Merkmale, um eine Person zu reidentifizieren,¹⁶³ wenn sie »Ausdruck ihrer physischen, physiologischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität sind«.¹⁶⁴ Dies gilt insbesondere, wenn die Aggregation auf der Zugehörigkeit zu identitätsstiftenden Gruppen, etwa politischer, religiöser, sozialer oder ethnischer Art, aufsetzt.¹⁶⁵ Schon die Zusammenführung von vier Datenpunkten in einem Konvolut von Kreditkartendaten kann genügen, um eine Person eindeutig bestimmbar zu machen.¹⁶⁶ Ein sicherer Weg der Anonymisierung ist unter Big-Data-Bedingungen insofern immer schwerer auffindbar.

161 Das Gesetz akzeptiert insoweit ein Restrisiko. Es stellt eine Verhältnismäßigkeitsprüfung zwischen dem Risiko der Personenzuordnung und dem damit verbundenen Aufwand als Teil einer Risikoanalyse an, vgl. *Wójtowicz*, PinG 1 (2013), 65 (68).

162 *Bitter/Buchmüller/Uecker*, Datenschutzrecht, in: Hoeren (Hrsg.), *Big Data und Recht*, 2014, S. 58 (79 f.); *Katko/Babaei-Beigi*, MMR 2014, 360 (361 f.); *Weichert* (Fn. 14), 258; *Wójtowicz* (Fn. 161), 67.

163 Vgl. etwa *Bornemann* (Fn. 10), 233; *Koch* (Fn. 104), 18.

164 Art. 2 lit. a Richtlinie 95/46/EG (Datenschutz-Richtlinie).

165 Bedenklich ist daher aus deutscher Perspektive der Plan Großbritanniens, eine Online-Datenbank aufzubauen, welche Informationen zu den Krankheiten der Briten in pseudonymisierter Form vorhält und diese Versicherungs- sowie Pharmaunternehmen zum Ankauf zur Verfügung stellt. Vgl. *Anonymous*, »Patientendaten aller Briten werden verkauft«, FAZ vom 21.1.2014, S. 9.

166 Siehe dazu *Montjoye/Laura Radaelli/Singh et al.*, *science* 347 (2015), 536 ff.

b) Lösungswege

Die Kunst, eine sachgerechte Antwort auf diese Herausforderungen von Big-Data-Technologien zu finden, liegt daher darin, geeignete Anonymisierungsstrategien zu entwickeln, welche einerseits die Analysezwecke nicht torpedieren, andererseits aber den Personenbezug sicher auflösen. Zielführend kann es sein, Datensätze in einer Weise zu zerlegen und zu verwürfeln, dass sie für statistische Zwecke ausgewertet, aber nicht einer bestimmten Person zugeordnet werden können. Der Anonymisierungsschlüssel bedarf eines Austauschs nach kurzen Zeitabständen.¹⁶⁷ Big Data impliziert insofern besondere Anforderungen an ein »Vergessenwerden«. Denkbar sind auch Methoden der Randomisierung, insbesondere der *differential privacy*, bei der Datenbankabfragen und -antworten einen Filter passieren, der Unschärfen hinzufügt und Daten auf statistische Kenngrößen herunterbricht, um einen Rückschluss auf bestimmte Personen auszuschließen.¹⁶⁸

V. Regulierungsstrategien und Fazit

Der Modebegriff »Big Data« steht für nachhaltige Umwälzungen unserer herkömmlichen Formen der Informationsanalyse.¹⁶⁹ Längst hat die algorithmengestützte Massendatenauswertung sich zu einem Megatrend gemausert. Big Data verspricht, bisher brachliegende Weinberge voller fruchtbarer Datenrebstöcke abzuernten, um daraus ein wertvolles Cuvée zu komponieren. Die Analyse liefert die Erntemaschine, welche die mühselige und kostspielige Handarbeit durch intelligente Algorithmen ersetzt.¹⁷⁰ Sie macht die »Datenlese« immer leichter und ertragreicher.

Vollständig neu sind ihre Möglichkeiten nicht. Bei genauerem Hinsehen ist Big Data weniger eine radikale Innovation als eine technische Transformation, weniger eine Revolution als eine Evolution, die eine neue Stufe erklimmt.¹⁷¹ Bei seinen technischen Möglichkeiten handelt es sich zugleich nicht lediglich um alten Wein in neuen Schläuchen. Seine DNA trägt enormes Wertschöpfungs- und gesellschaftliches Problemlösungspotenzial in sich – von der Medizin über den Umweltschutz bis hin zur Infrastruktursteuerung. Durch eine umfassende Exploration des Rohstoffs »Daten« generieren Big-Data-Analysen entscheidungsrelevantes Wissen. Sie ergänzen die menschliche Rationalität und überwinden hergebrachte Grenzen menschlicher Problemverarbeitungskapazität. In einer datengetränkten Welt wird die Qualität der Datenverarbeitung insbesondere zu einem zentralen Produktionsfaktor der digitalen Infrastruktur: Der Wert von Daten hängt immer mehr von der Schnelligkeit und Qualität ihrer Verarbeitung ab. Diejenigen Akteure, denen es

167 Ulmer (Fn. 12), 330.

168 Diese Methoden gelten als Goldstandard der Privacy-Sicherung; *Artikel-29-Datenschutzgruppe*, Stellungnahme 5/2014 zu Anonymisierungstechniken, WP 216, 2014, S. 14 ff.; *Müller-Quade*, Kursbuch 177: Privat 2.0 2014, 130 (143 f.); *Weichert* (Fn. 14), 258; aus technischer Sicht siehe auch *Dwork*, *Differential Privacy: A Cryptographic Approach to Private Data Analysis*, in: Lane/Stodden/Bender et al. (Hrsg.), *Privacy, Big Data, and the Public Good*, 2014, S. 296 ff.

169 Das gilt vor allem dann, wenn aus »Big Data« »Smart Data« wird, *Kreibich* (Fn. 46), 22.

170 Vgl. auch *Weichert* (Fn. 12), 132.

171 Was im Zeitalter der Aktenbände ein findiger Ermittler noch in langwieriger Kleinstarbeit aus vielen Quellen zusammenführen musste, lässt sich nun in Millisekunden per Mausklick ergründen.

gelingt, relevante von irrelevanten Daten in Echtzeit zu unterscheiden, können auf entscheidende Vorteile im politischen, gesellschaftlichen und wirtschaftlichen Wettbewerb hoffen.¹⁷²

Big Data ist aber nicht nur Big Business. Es kann auch einer neuen Überwachungsarchitektur den Weg ebnen. Big Data ist insoweit Fluch und Segen zugleich.¹⁷³ Denn noch nie zuvor war es so leicht und unbemerkt möglich, letzte Hüllen auf dem Weg zum intimen Innersten der Persönlichkeit unversehens abzustreifen. Big Data und Big Brother gehen schnell eine gefährliche Liaison ein. Der Informationsvorsprung, den die Auswertungsmöglichkeiten von Big-Data-Technologien verschaffen, paart sich mit einer Informations- und Steuerungsmacht, die auf der undurchsichtigen Eigenrationalität von Algorithmen aufbaut. Strukturelle Asymmetrien von Interessendurchsetzungschancen mit korrespondierenden Gefährdungen des informationellen Selbstbestimmungsrechts sind deren Frucht. Facebook & Co steuern durch ihre Algorithmen nicht alleine, welche Informationen der im Datensilo seines sozialen Netzwerks sitzende Nutzer zu Gesicht bekommt (und welche nicht). Vielmehr treffen die algorithmischen Logiken marktmächtiger Big-Data-Kollektoren Entscheidungen mit erheblicher Tragweite für die öffentliche Wahrnehmung und steuern in bisher unbekanntem Umfang den Zugang zu Märkten und Teilhabechancen in einer digitalen Welt.

1. Kollektivierung des Risikos für die informationelle Selbstbestimmung durch Big-Data-Algorithmen

Der Zauber der Algorithmen, den Big-Data-Technologien versprühen, verleitet zu einem unreflektierten, gar blinden Glauben in die Richtigkeit ihrer Resultate.¹⁷⁴ Algorithmen als Handlungsanweisungen zur Aufgabenbewältigung nach einem vordefinierten Schema sprechen unser Grundvertrauen in die Objektivität von Zahlen an. Sie suggerieren eine Unfehlbarkeit der Analyse, die unter den Bedingungen großer Datengrundgesamtheiten den Beweis des ersten Anscheins für sich in Anspruch nimmt. Maschinell berechnete Wahrscheinlichkeiten maskieren sich als Wahrheiten (bzw. werden als solche wahrgenommen), ohne solche sein zu können.

Sie behandeln das individuelle Verhalten vorrangig als Ausdruck kollektiver Handlungsmuster. Aus der Zuordnung zu einer mit besonderen Risikofaktoren behafteten Gruppe erwächst dann schnell ein Pauschalverdacht. Die Analyseinstrumente verwenden die Ergebnisse einer Auswertung nämlich auch für solche Personen, deren persönliche Daten nicht zur Verfügung stehen, die Gruppenmerkmale aber erfüllen.¹⁷⁵ Sie lösen die Suche nach Ursachen für bestimmte Verhaltensweisen durch die Suche nach Korrelationen ab. Die Korrelationsanalyse lässt eine überwiegende Wahrscheinlichkeit ausreichen, um eine Handlungsanleitung zu geben. Dadurch können Big-Data-Analysen die persönliche Freiheit bedrohen. Wer in die Verarbeitung seiner Daten ausdrücklich einwilligt, zeitigt damit auch externe Effekte für andere, die sich in einer vergleichbaren Situation befinden. Der faktischen Datenvorhaltung wächst so eine normative Kraft

172 Vgl. auch *Bachmann/Kemper/Gerzer* (Fn. 9), 49 ff.

173 Damit kann sich auch ein Scheideweg »zwischen Sicherheit und Vorhersehbarkeit einerseits und Freiheit und Risiko andererseits« verbinden, *Mayer-Schönberger* (Fn. 44), 19.

174 *Mayer-Schönberger/Cukier* (Fn. 3), S. 209.

175 *Roßnagel* (Fn. 91), 566.

für unbeteiligte Dritte zu. Big Data kollektiviert insofern das Risiko für die informationelle Selbstbestimmung.¹⁷⁶ Es macht dann das Individuum zum Gefangenen von Wahrscheinlichkeiten, die spekulative Rückschlüsse vom Gruppenverhalten auf den Einzelnen ziehen, auf diesen aber nicht zutreffen müssen.¹⁷⁷ Die Zuordnung zu statistisch ermittelten Gruppen mündet schnell in eine individuell unzutreffende Bewertung auf der Grundlage verzerrter Persönlichkeitsbilder sowie in eine Selektivität der Realitätserfassung, die zu stereotypisierenden Vorverurteilungen einlädt.¹⁷⁸ Mit den Analysetools verbindet sich insoweit das besondere Risiko einer Gruppendiskriminierung.¹⁷⁹

Wer in einem Gebiet wohnt, das für eine schlechte Zahlungsmoral bekannt ist, muss damit rechnen, von manch einem Anbieter nicht oder nur gegen Vorkasse beliefert zu werden. Gruppen, die aufgrund ihrer Kaufhistorie demgegenüber für besondere Zahlungskraft bekannt sind, erhalten bei Einkäufen im Internet womöglich unversehens ein höheres Preisangebot als der Durchschnittskunde und sehen sich damit einer Preisdiskriminierung ausgesetzt. Wer zum Shoppen nach New York möchte, muss in einer Big-Data-Welt gewärtigen, dass er diesen Plan durch Eingabe von Verdacht erregenden Suchbegriffen im Internet, wie z.B. »Al Qaida«, gefährden kann.

Bei allen Risiken, die Big-Data-Technologien in sich bergen: Die Forderung, personenbezogene Daten deshalb generell dem Regime von Big Data zu entziehen,¹⁸⁰ schösse freilich über das Ziel hinaus. Auf statistische Größen rekurrierende, automatische Entscheidungen sollte die Rechtsordnung jedoch zum einen einer speziellen Risikokontrolle unterwerfen – jedenfalls dann, wenn eine Auswertung in eine Zuordnung zu einer Risikogruppe mündet, die eine Diskriminierung in einem sensiblen Merkmal nach sich ziehen kann.

Big-Data-Technologie weist in mancher Hinsicht strukturelle Ähnlichkeit mit risikotechnologischen Innovationen, wie etwa der Atomkraft, auf: Ihr Gemeinwohlpotenzial ist enorm, solange sie unter Kontrolle sind. Geraten sie in die falschen Hände oder werden sie falsch eingesetzt, richten sie allerdings kaum beherrschbaren Schaden an. Die Delegation immer weiterreichender Entscheidungen an immer komplexer konfigurierte, intransparente und autonom agierende Systeme birgt bei unzureichender Risikovorsorge die Gefahr eines folgenschweren Kontrollverlustes. Algorithmen sind gleichsam die »Atomkraft des 21. Jahrhunderts«. Als Schaltstellen in der digitalen Infrastruktur der Zukunft ist die regulatorische Umhegung ihres Einsatzes in besonders kritischen, weil diskriminierungsanfälligen oder persönlichkeitsrechtlich sensiblen Entscheidungszusammenhängen geboten. Algorithmen als neuer Machtfaktor der digitalen Welt sind insbesondere nicht frei von Wertungen; sie spiegeln die Wertmuster derer wider, die sie programmieren, und bedürfen daher einer Kontrolle. Sachgerechte Big-Data-Regulierung ist insofern Algorithmenkontrolle.¹⁸¹ Es braucht ein Algorithmen-Risikotechnologierecht. Seine Aufgabe ist es, eine ausgewogene

176 Roßnagel (Fn. 91), 566; Weichert (Fn. 14), 254.

177 Mayer-Schönberger/Cukier (Fn. 3), S. 206.

178 Roßnagel (Fn. 91), 566; vgl. dazu auch Leopold (Fn. 35), 80; zu eng aber seine Schlussfolgerung, dass Anonymisierungsverfahren der geeignete Hebel zur Problembehebung sind. Denn die Problematik beruht gerade darauf, dass die Zuordnung einer Person zu einer Gruppe Diskriminierungspotenzial in sich trägt.

179 Optimistischer insoweit Mayer-Schönberger/Cukier (Fn. 3), S. 203, welche die Chance sehen, der Zwangsjacke von Gruppenidentitäten im Gefolge von Big Data zu entkommen, da Big Data ein besseres, weniger diskriminierendes und stärker individualisierendes Profiling verspreche.

180 So aber Ohm, UCLA Law Review 57 (2010), 1701 (1742 f.).

181 Vgl. dazu auch oben S. 26.

Balance zwischen dem notwendigen Schutz personenbezogener Daten und dem Interesse der Gesellschaft an innovativen Nutzungsformen herzustellen. Hierbei darf nicht die Technik das Recht, sondern muss das Recht die Technik bestimmen: In der digitalen neuen Welt darf nicht alles, was technisch möglich ist, auch rechtlich zulässig sein. Technikentwicklung und Recht sollten stattdessen (z.B. im Rahmen von Standardsetzungsprozessen) in einen fruchtbaren Dialog treten.

Das Datenschutzrecht setzt Big-Data-Analysen bereits heute Grenzen. Diese sind allerdings noch nicht hinreichend auf das digitale Zeitalter und die Chancen sowie Herausforderungen abgestimmt, die von Big-Data-Analysen und ihren Algorithmen ausgehen. Big Data braucht insofern keine Neuerfindung des Datenschutzes – wohl aber eine zeitgerechte Anpassung.

2. Transparenzsichernde Handlungsinstrumente

Der Staat ist aufgerufen, sichere Rahmenbedingungen für den Umgang mit den technologischen Möglichkeiten algorithmengesteuerter Massendatenauswertung zu entwickeln – Rahmenbedingungen, die Vertrauen in ihren Nutzen sowie ihre Verantwortbarkeit begründen und die individuellen sowie kollektiven Risiken umfassender Persönlichkeitsausforschung und einer Gruppendiskriminierung durch vorsorgenden Datenschutz verringern. Flankierende Sicherungsinstrumente bilden insoweit insbesondere Ansprüche Betroffener auf Zugang zu über sie gespeicherte Informationen, feste Löschungsfristen, die Daten im Zeitlauf verwischen wie Fußspuren am Sandstrand, Big-Data-adäquate Anonymisierungstechniken und staatliche Kontrolle bei potenziell diskriminierenden algorithmischen Steuerungen.¹⁸² Ein zentraler Schlüssel zum verantwortungsvollen Umgang mit Big Data ist die Herstellung von Transparenz in der Analysepraxis. Big-Data-Verfahren müssen offenlegen, welche Daten erhoben und verwertet, wie sie verschmolzen und wie ihre Resultate verwendet werden. Die Zertifizierung und Auditierung entsprechender Analyseinstrumente, insbesondere durch die staatlichen Datenschutzbeauftragten, sowie Datenschutz-Folgeabschätzungen für besonders persönlichkeitsensitive Verarbeitungsvorgänge¹⁸³ können insoweit ein wirksames präventives Kontrollinstrument sein.¹⁸⁴

Datenschutzregulierung darf kein *race to the bottom* auslösen, in dem niedrige Datenschutzstandards einen Wettbewerbsvorteil versprechen. Sie sollte vielmehr einen Wettbewerb um Techniken entfachen, welche die digitale Unversehrtheit des Einzelnen im Internet sichern. Diese sollten die Menschen in die Lage versetzen, aufgeklärt darüber zu entscheiden, was mit ihren Daten geschieht. Die Bürger erhalten dann im Idealfall, etwa durch standardisierte Schnittstellen zu ihren Online-Nutzungsdaten bei Big-Data-Kollektoren, durch »Online-Kontoauszüge« einen Einblick in ihr Daten-Konto (und Portabilitätsrechte). Das eröffnet ihnen einen einfachen Weg, zu kontrollieren und zu erfahren, welche Daten verantwortliche Stellen über sie vorhalten, in welchem Kontext ihre Daten eingesetzt werden und welche Daten gelöscht oder mit einem Verfallsdatum versehen werden sollen, um ihre digitale Souveränität und Autonomie

182 Roßnagel (Fn. 91), 566.

183 In diese Richtung denkt auch Art. 33 Datenschutz-Grundverordnung-E.

184 Roßnagel (Fn. 91), 566.

wahren zu können.¹⁸⁵ *Privacy by design* ist ein wichtiges Hilfsmittel zur Sicherung der informationellen Selbstbestimmung gegenüber Big-Data-Anwendungen. Es stellt sicher, dass technologische Herausforderungen für das Selbstbestimmungsrecht bereits bei der Konzeption neuer Technologien Berücksichtigung finden. Viel spricht für eine Vorverlagerung des Schutzes in die Technikenebene.¹⁸⁶

Ob es gelingt, die Nutzungschancen der Big-Data-Technologie unter Zuhilfenahme solcher Techniken mit dem Persönlichkeitsschutz Betroffener zu versöhnen, stellt die Weichen für ihr gesellschaftliches und ökonomisches Entfaltungspotenzial. Ohne das Vertrauen der Datenspende in die missbrauchsfreie, gesellschaftlich akzeptierte Nutzung¹⁸⁷ gibt es nicht zuletzt kein Big Data. Sein Wesen besteht im Idealfall nicht nur darin, Daten zu sammeln, sondern diese auch vertrauenswürdig zu beherrschen und einzusetzen. Eine ausgewogene Balance zwischen wirtschaftlicher Wertschöpfung und Persönlichkeitsschutz herzustellen, darin liegt auch die zentrale Herausforderung der Datenschutzverfassung einer freiheitlichen Gesellschaft. Geboten ist eine Diskussion um den verantwortungsvollen und gesellschaftlich akzeptierten Rahmen der neuen Möglichkeiten.¹⁸⁸ Nur dann schöpft der Bürger Vertrauen, durch Big-Data-Analyse nicht lediglich zum Objekt eines informatorischen Panoptikums zu werden. Anderenfalls wird er aufseufzen wie einer der *Precogs* in Film *Minority Report*: »Ich brauche Ruhe, ich brauche Ruhe vor der Zukunft.«

Literaturverzeichnis

- Ambrose, Meg Meta, *It's about time: Pivacy, Information, Life Cycles, and the Rgight to be forgotten*, Stanford Technology Law Review 16 (2013), S. 369–421.
- Andelfinger, Volker P./Hänisch, Till, *Internet der Dinge, Technik, Trends und Geschäftsmodelle*, Wiesbaden, 2015.
- Andersen, Chris, *The Long Tail*, http://archive.wired.com/wired/archive/12.10/tail.html?pg=1&topic=tail&topic_set (25.2.2015).
- Anderson, Chris, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, Wired Magazine vom 23.6.2008, archive.wired.com/science/discoveries/magazine/16-07/pb_theory (10.1.2015).
- Anonymous, »Patientendaten aller Briten werden verkauft«, Staatliche Gesundheitsbehörde NHS baut im Internet eine neue Datenbank auf, FAZ vom 21.1.2014, S. 9.
- Anonymous, »Big Data gegen böse Buben«, FAZ vom 1.12.2014, S. 26.
- Anonymous, 17 Telekommitarbeiter und ihr Blick in die Zukunft, FAZ vom 21.3.2015, S. 26.
- Armbrüster, Christian, *Auf dem Weg zum gläsernen Versicherten?*, NJW 2015, Editorial Heft 6.
- Arning, Marian/Moos, Fleming, *Big Data bei verhaltensbezogener Online-Werbung, Programmatic Buying und Real Time Advertising*, ZD 2014, S. 242–248.
- Artikel-29-Datenschutzgruppe, *Stellungnahme 5/2014 zu Anonymisierungstechniken*, WP 216, Brüssel, 2014.

185 Greenwood/Stopczynski/Sweatt et al., *The New Deal on Data: A Framework for Institutional Controls*, in: Lane/Stodden/Bender et al. (Hrsg.), *Privacy, Big Data, and the Public Good*, 2014, S. 192 (195 ff.) sprechen einem »New Deal« zum Schutz der Souveränität über Daten.

186 Leopold (Fn. 35), 80.

187 Sie wird unter dem Stichwort »datability« – einem der Leitbegriffe der CeBIT 2014 – diskutiert. Das Kunstwort kombiniert Big Data mit den englischen Begriffen für Nachhaltigkeit (sustainability) und Verantwortung (responsibility). Er steht für die Herausforderung, große Datenmengen nachhaltig und verantwortungsvoll zu verarbeiten.

188 Vgl. auch Morozov, APuZ 2015, 3 (7): »Letztendlich können wir nur dann eine alternative Vision für die Nutzung von Kommunikationstechnologien artikulieren, wenn wir hinterfragen, ob es angemessen ist, dass es die großen Unternehmen wie Google und Facebook sind, die letztgültig vermitteln, wie wir leben, heilen, studieren oder reisen.«

- Austin, Peter C./Mamdani, Muhammad M./Juurlink, David N./Hux, Janet E., Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health, *Journal of Clinical Epidemiology* 59 (2006), S. 964–969.
- Bachmann, Ronald/Kemper, Guido/Gerzer, Thomas, *Big Data – Fluch oder Segen?*, Unternehmen im Spiegel gesellschaftlichen Wandels, Heidelberg, Neckar, 2014.
- Baeriswyl, Bruno, «Big Data» ohne Datenschutz-Leitplanken, Die Herausforderungen für den Datenschutz bei «Big Data» sind enorm – eine Lösung ist (noch) nicht absehbar, *digma* 2013, S. 14–17.
- Bauernhansl, Thomas, *Industrie 4.0 in Produktion, Automatisierung und Logistik, Anwendung, Technologien und Migration*, Wiesbaden, 2014.
- BITKOM, *Big Data im Praxiseinsatz, Szenarien, Beispiele, Effekte*, Berlin, 2012.
- BITKOM, *Weltmarkt für Big Data wächst rasant*, 2014.
- BITKOM, *Big Data und Geschäftsmodell-Innovationen in der Praxis: 40+ Beispiele*, Berlin, 2015.
- Bitter, Till/Buchmüller, Christoph/Uecker, Philip, Datenschutzrecht, in: Hoeren, Thomas (Hrsg.), *Big Data und Recht*, Bd. 83, München, 2014, S. 58–93.
- Blaß, Katharina, *Bewegungsstatistik: Verkehrsbetrieb wertet Mobilfunkdaten aus*, www.spiegel.de/netzwelt/netzpolitik/daten-schutz-vag-in-nuernberg-greift-telekom-daten-ab-a-1024001.html.
- Boehme-Neßler, Volker, Das Recht auf Vergessenwerden: Ein neues Internet-Grundrecht im Europäischen Recht, *NVwZ* 2014, S. 825–830.
- Bornemann, Dirk, Big Data - Chancen und rechtliche Hürden, *RDV* 2013, S. 232–235.
- Boyd, Danah/Crawford, Kate, Big Data als kulturelles, technologisches und wissenschaftliches Phänomen – Sechs Provokationen, in: Geiselberger, Heinrich (Hrsg.), *Big Data, Das neue Versprechen der Allwissenheit*, Berlin, 2013, S. 187–218.
- Brandenburg, Arne/Leuthner, Christian, Location based Services und Local Commerce, Rechtliche Betrachtung des Einsatzes aktueller Technologien, in: Taeger, Jürgen (Hrsg.), *Big Data & Co, Neue Herausforderungen für das Informationsrecht*, Ede- wecht, 2014, S. 651–665.
- Bräutigam, Peter/Klindt, Thomas, *Industrie 4.0, das Internet der Dinge und das Recht*, *NJW* 2015, S. 1137–1142.
- Brücher, Cornel, *Rethink Big Data*, Heidelberg et al., 2013.
- Bull, Hans Peter, Die Verfassung stützen – aber richtig, Aufgaben und Befugnisse der Sicherheitsbehörden müssen neu verteilt werden, *PinG* 1 (2013).
- Buschauer, Regine, (Very) nervous systems. Big Mobile data, in: Reichert, Ramón (Hrsg.), *Big Data, Die Gesellschaft als digitale Maschine*, Bielefeld, 2014, S. 405–436.
- Commerzbank, *Management im Wandel, Digitaler, effizienter, flexibler!*, Frankfurt a. M., 2015.
- Dapp, Thomas F., *Big Data – die ungezähmte Macht*, Frankfurt, 2014.
- Däubler, Wolfgang/Weichert, Thilo/Wedde, Peter (Hrsg.), *BDSG, 4. Aufl.*, Frankfurt a. M., 2014.
- Dehmer, Matthias/Holzinger, Andreas/Emmert-Streib, Frank, Personalized medicine by means of complex networks: a Big Data challenge, in: Weber, Rolf H./Thouvenin, Florent (Hrsg.), *Big Data und Datenschutz - Gegenseitige Herausforderungen*, Bd. 59, Zürich, 2014, S. 37–44.
- Dorner, Michael, Big Data und „Dateneigentum“, *Grundfragen des modernen Daten- und Informationshandelns*, *CR* 2014, S. 617–628.
- Dugas, Andrea Freyer/Hsieh, Yu-Hsinag/Levin, Scott R./Pinesm Jesse M/Mareiniss/Mohareb/Gaydos, Charlotte A./Perl, Trish M, Rothman, Richard E., Google Flu Trends: Correlation with Emergency Department Influenza Rates and Crowding Metrics, *CID* 2015, S. 463–469.
- Dwork, Cynthia, Differential Privacy: A Cryptographic Approach to Private Data Analysis, in: Lane, Julia/Stodden, Victoria/Bender, Stefan u.a. (Hrsg.), *Privacy, Big Data, and the Public Good, Frameworks for Engagement*, New York, NY, 2014, S. 296–322.
- Dyson, George, *Turing's cathedral, The origins of the digital universe*, New York, 2012.
- Evans, Dave, *Das Internet der Dinge, So verändert die nächste Dimension des Internet die Welt*, Jan Jose, 2011.
- Fessler, Sven, Die Datenflut steigt - wie können wir sie nutzen?, *datareport* 1/2013, S. 11–14.
- Forgó, Nikolaus/Krügel, Tina/Rapp, Stefan, *Zwecksetzung und informationelle Gewaltenteilung*, Baden-Baden, 2006.
- Fraunhofer-IAIS, *Big Data - Vorsprung durch Wissen, Innovationspotenzialanalyse*, 2012.

- Fremuth, Michael Lysander, Wächst zusammen, was zusammengehört?, Das Trennungsgebot zwischen Polizeibehörden und Nachrichtendiensten im Lichte der Reform der deutschen Sicherheitsarchitektur, AöR 139 (2014), S. 32–79.
- Geiselberger, Heinrich (Hrsg.), Big Data, Das neue Versprechen der Allwissenheit, Berlin, 2013.
- Glaser, Peter, Erinnerung der Zukunft, in: Geiselberger, Heinrich (Hrsg.), Big Data, Das neue Versprechen der Allwissenheit, Berlin, 2013, S. 281–294.
- Greenwood, Daniel/Stopczynski/Sweatt/Hardjono, Thomas/Pentland, Alex, The New Deal on Data: A Framework for Institutional Controls, in: Lane, Julia/Stodden, Victoria/Bender, Stefan u.a. (Hrsg.), Privacy, Big Data, and the Public Good, Frameworks for Engagement, New York, NY, 2014, S. 192–210.
- Gusy, Christoph, Das gesetzliche Trennungsgebot zwischen Polizei und Verfassungsschutz, Die Verwaltung 24 (1991), S. 467.
- Hallermann, Ulrich, Vorabkontrollen nach dem BDSG, Handlungsempfehlungen für die praktische Umsetzung, RDV 2015, S. 23–26.
- Härtling, Niko, Profiling: Vorschläge für eine intelligente Regulierung, Was aus der Zweistufigkeit des Profiling für die Regelung des Nicht-öffentlichen Datenschutzbereichs folgt, CR 2014, S. 528–536.
- Haun, Matthias, Handbuch Robotik, Programmieren und Einsatz intelligenter Roboter, 2. Aufl., Berlin, Heidelberg, 2013.
- Heller, Piotr, Zeigt mir, wie fit du bist, FAS vom 6.4.2014, S. 55.
- Heuer, Steffan, Kleine Daten, große Wirkung, Düsseldorf, 2013.
- Hill, Hermann, Business Intelligence/Business Analytics im öffentlichen Sektor, DÖV 2010, S. 789–797.
- Hill, Hermann, Aus Daten Sinn machen: Analyse- und Deutungskompetenzen in der Datenflut, DÖV 2014, S. 213–222.
- Hill, Hermann/Schliesky, Utz (Hrsg.), Neubestimmung der Privatheit, Baden-Baden, 2014.
- Hirsch-Kreinsen, Hartmut, Digitalisierung industrieller Arbeit, Die Vision Industrie 4.0 und ihre sozialen Herausforderungen, Baden-Baden, 2015.
- Hoeren, Thomas (Hrsg.), Big Data und Recht Bd. 83, München, 2014.
- Hofmann, Olaf, Methoden des Social-Media-Monitoring, in: König, Christian/Stahl, Matthias/Wiegand, Erich (Hrsg.), Soziale Medien, Gegenstand und Instrument der Forschung, Wiesbaden, 2014, S. 161–170.
- Hofstetter, Yvonne, Verkannte Revolution: Big Data und die Macht des Marktes, APuZ 2015, S. 33–38.
- Hornung, Gerrit/Hofmann, Kai, Ein „Recht auf Vergessenwerden“?: Anspruch und Wirklichkeit eines neuen Datenschutzrechts, JZ 2013, S. 163–170.
- Hume, David, Eine Untersuchung über den menschlichen Verstand, Übersetzt von Raoul Richter, 12. Aufl., Hamburg, 1993.
- Kaku, Michio, Die Physik des Bewusstseins, Über die Zukunft des Geistes, Reinbek, 2014.
- Katko, Peter/Babaei-Beigi, Ayda, Accountability statt Einwilligung? : Führt Big Data zum Paradigmenwechsel im Datenschutz?, MMR 2014, S. 360–364.
- Kester, George W., What Happened to the Super Bowl Stock Market Predictor?, The Journal of Investing 19 (2010), S. 82–87.
- Klausnitzer, Rudi, Das Ende des Zufalls, Salzburg, 2013.
- Koch, Frank A., Über die Unvereinbarkeit des deutschen und europäischen Datenschutzrechts mit Big Data, ITRB 2015, S. 13–20.
- Kodde, Claudia, Die „Pflicht zu Vergessen“, ZD 2013, S. 115–118.
- König, Christian/Stahl, Matthias/Wiegand, Erich (Hrsg.), Soziale Medien, Gegenstand und Instrument der Forschung, Wiesbaden, 2014.
- Körber, Torsten, Google im Fadenkreuz der Kartellbehörden, NZKart 2014, S. 293–294.
- Kreibich, Rolf, Von Big zu Smart - zu Sustainable?, APuZ 2015, S. 20–26.
- Kremp, Matthias, Durchbruch bei künstlicher Intelligenz: Der unheimlich menschliche Eugene Goostmann, Spiegel online vom 9.6.2014, www.spiegel.de/netzwelt/gadgets/eugene-goostman-computer-besteht-erstmal-turing-test-a-974131.html (20.10.2014).
- Lane, Julia/Stodden, Victoria/Bender, Stefan/Nissenbaum, Helen (Hrsg.), Privacy, Big Data, and the Public Good, Frameworks for Engagement, New York, NY, 2014.
- Langkafel, Peter, Auf dem Weg zum Dr. Algorithmus? Potentiale von Big Data in der Medizin, APuZ 2015, S. 27–32.
- LDI NRW, 22. Datenschutzbericht, Für die Zeit vom 1.1.2013 bis zum 31.12.2014, Düsseldorf, 2015.

- Leopold, Nils, Big Data – eine neue Herausforderung für den Datenschutz, vorgänge 2012, S. 74–82.
- Liedke, Bernd, BIG DATA - small information : muss der datenschutzrechtliche Auskunftsanspruch reformiert werden?, K&R 2014, S. 709–714.
- Lüdemann, Volker/Sengstacken, Christin/Vogelpohl, Kerstin, Pay as you drive: Datenschutz in der Telematikversicherung, RDV 2014, S. 302–306.
- Mantz, Reto, Verwertung von Standortdaten und Bewegungsprofilen durch Telekommunikationsdiensteanbieter, K&R 2013, S. 7–11.
- Martini, Mario, Datenschutz und Sicherheit bei der elektronischen Rechnung, in: Rogall-Grothe, Cornelia (Hrsg.), Leitfaden Elektronische Rechnung in der öffentlichen Verwaltung, Grundlagen, Umsetzungsempfehlungen, Best practices, Frankfurt am Main, 2014, S. 51–83.
- Martini, Mario, Vom heimischen Sofa in die digitale Agora: E-Partizipation als Instrument einer lebendigen Demokratie?, in: Hill, Hermann/Schliesky, Utz (Hrsg.), Neubestimmung der Privatheit, Baden-Baden, 2014, S. 193–247.
- Martini, Mario/Weiß, Wolfgang/Ziekow, Jan, Rechtliche Zulässigkeit flächendeckender Alarmierungen der Bevölkerung in Katastrophenfällen per SMS (KatWarn), Baden-Baden, 2013.
- Mayer-Schönberger, Victor, Delete, Die Tugend des Vergessens in digitalen Zeiten, 2. Aufl., Berlin, 2011.
- Mayer-Schönberger, Victor, Neue Erkenntnisse über die Wirklichkeit, FuL 2014, S. 706–707.
- Mayer-Schönberger, Victor, Was ist Big Data? Zur Beschleunigung des menschlichen Erkenntnisprozesses, APuZ 2015, S. 14–19.
- Mayer-Schönberger, Viktor/Cukier, Kenneth, Big Data, Die Revolution, die unser Leben verändern wird, München, 2013.
- Meinecke, Dirk, Big Data und Data Mining: Automatisierte Strafverfolgung als neue Wunderwaffe der Verbrechensbekämpfung?, in: Taeger, Jürgen (Hrsg.), Big Data & Co, Neue Herausforderungen für das Informationsrecht, Edewecht, 2014, S. 183–202.
- Montjoye, Yves-Alexandre de/Laura Radaelli/Singh, Vivek Kumar/Pentland, Alex “Sandy”, Unique in the shopping mall: On the reidentifiability of credit card metadata, science 347 (2015), S. 536–539.
- Moos, Flemming, Update Datenschutz, in: Taeger, Jürgen (Hrsg.), Big Data & Co, Neue Herausforderungen für das Informationsrecht, Edewecht, 2014, S. 525–538.
- Morozov, Evgeny, „Ich habe doch nichts zu verbergen“, APuZ 2015, S. 3–7.
- Mühl, Melanie, Korrekturen ausgeschlossen, FAZ vom 14.2.2014, S. 35.
- Müller-Quade, Jörn, Privatsphäre gesucht!, Neue Big-Data-Techniken auf dem Vormarsch, Kursbuch 177: Privat 2.0 2014, S. 130–144.
- Nolte, Norbert, Das Recht auf Vergessenwerden – mehr als nur ein Hype?, NJW 2014, S. 2238–2242.
- Offenhuber, Dietmar/Ratti, Carlo, Drei Mythen über Smart Citys und Big Data, in: Geiselberger, Heinrich (Hrsg.), Big Data, Das neue Versprechen der Allwissenheit, Berlin, 2013, S. 149–155.
- Ohm, Paul, Broken Promise of Privacy: responding to the Surprising Failure of Anonymization, UCLA Law Review 57 (2010), S. 1701–1777.
- Ohrtmann, Jan-Peter/Schwiering, Sebastian, Big Data und Datenschutz – Rechtliche Herausforderungen und Lösungsansätze, NJW 2014, S. 2984–2990.
- Paal, Boris, Vielfaltssicherung im Suchmaschinen-Sektor, ZRP 2015, S. 34–38.
- Paal, Boris P., Immaterialgüter, Internetmonopole und Kartellrecht, GRUR-Beilage 2014, S. 69–77.
- Pentland, Alex, Social physics, How good ideas spread - the lessons from a new science, 2014.
- Peters, Falk/Kersten, Heinrich/Wolfenstetter, Klaus-Dieter (Hrsg.), Innovativer Datenschutz, Berlin, 2012.
- Piltz, Carlo, Recht auf Vergessenwerden: das Google-Urteil in der Praxis, PinG 2 (2014), S. 180–182.
- Pinnow, Carsten/Schäfer, Stephan, Industrie 4.0 (R)Evolution für Wirtschaft, Politik und Gesellschaft, Berlin, 2015.
- President's Council of Advisors on Science and Technology, Report to the President: Big Data and Privacy: A technological Perspective, 2014.
- Reichert, Ramón (Hrsg.), Big Data, Die Gesellschaft als digitale Maschine, Bielefeld, 2014.
- Rogall-Grothe, Cornelia (Hrsg.), Leitfaden Elektronische Rechnung in der öffentlichen Verwaltung, Grundlagen, Umsetzungsempfehlungen, Best practices, Frankfurt am Main, 2014.

- Rogosch, Patricia, Die Einwilligung im Datenschutzrecht, Baden-Baden, 2013.
- Roßnagel, Alexander (Hrsg.), Beck'scher Kommentar zum Recht der Telemediendienste, Telemediengesetz, Jugendmedienschutz-Staatsvertrag (Auszug), Signaturgesetz, Signaturverordnung, Vorschriften zum elektronischen Rechts- und Geschäftsverkehr, München, 2013.
- Roßnagel, Alexander, Big Data – Small Privacy?, ZD 2013, S. 562–567.
- Rudder, Christian, Dataclysm, Who we are(Wenn We Think No One's Looking, 2014.
- Schaar, Peter, Zwischen Big Data und Big Brother: zehn Jahre als Bundesbeauftragter für den Datenschutz und die Informationsfreiheit, RDV 2013, S. 223–227.
- Schirmmayer, Frank, Der verwettete Mensch, in: Geiselberger, Heinrich (Hrsg.), Big Data, Das neue Versprechen der Allwissenheit, Berlin, 2013, S. 273–280.
- Schürmann, Kathrin/von der Heide, (i)Beacons- technische Hintergrund und Datenschutz – rechtliche Anforderungen, in: Taeger, Jürgen (Hrsg.), Big Data & Co, Neue Herausforderungen für das Informationsrecht, Edewecht, 2014, S. 637–649.
- Schwichtenberg, Simon, „Pay as you drive“ - neue und altbekannte Probleme, DuD 2015, S. 378–382.
- Spies, Axel, USA: Verkehrsüberwachung in New York (Midtown in Motion) und Big Data, ZD-Aktuell 2013, S. 03734.
- Spindler, Gerald/Schuster, Fabian (Hrsg.), Recht der elektronischen Medien, Kommentar, 3. Aufl., München, 2015.
- Stöcker, Christian, Politikfeld Big Data: Hoffnungen, Vorhaben und viele offene Fragen, APuZ 2015, S. 8–13.
- Taeger, Jürgen (Hrsg.), Big Data & Co, Neue Herausforderungen für das Informationsrecht, Edewecht, 2014.
- Taleb, Nassim Nicholas, Der schwarze Schwan, Konsequenzen aus der Krise, München, 2012.
- Thouvenin, Florent, Erkennbarkeit und Zweckbindung, Grundprinzipien des Datenschutzrechts auf dem Prüfstand von Big Data, in: Weber, Rolf H./Thouvenin, Florent (Hrsg.), Big Data und Datenschutz - Gegenseitige Herausforderungen, Bd. 59, Zürich, 2014, S. 61–83.
- TNS Infratest, Quo vadis Big Data, Herausforderungen – Erfahrungen – Lösungsansätze, 2012.
- Türpe, Sven, Denkverbote für Star-Trek-Computer?: Big Data, statistische Modelle und lernende Maschinen, DuD 38 (2014), S. 31–35.
- ULD Schleswig-Holstein, Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen, Kiel, 2014.
- Ulmer, Claus-Dieter, BIG DATA - neue Geschäftsmodelle, neue Verantwortlichkeiten?, RDV 2013, S. 227–232.
- Venzke-Caprese, Sven, Social Media Monitoring, Analyse und Profiling ohne klare Grenzen?, DuD 2013, S. 775–779.
- Wagner, Frank, Datenschutz in Connected Homes, in: Peters, Falk/Kersten, Heinrich/Wolfenstetter, Klaus-Dieter (Hrsg.), Innovativer Datenschutz, Berlin, 2012, S. 205.
- Walker, Robert Kirk, The Right to be forgotten, Hastings Law Journal 64 (2012), S. 257–286.
- Weber, Rolf H./Thouvenin, Florent (Hrsg.), Big Data und Datenschutz - Gegenseitige Herausforderungen Bd. 59, Zürich, 2014.
- Weichert, Thilo, Big Data und Datenschutz, ZD 2013, S. 251–259.
- Weichert, Tilo, Big Data – eine Herausforderung für den Datenschutz, in: Geiselberger, Heinrich (Hrsg.), Big Data, Das neue Versprechen der Allwissenheit, Berlin, 2013, S. 131–148.
- Wiekens, Carina J./Stapel, Diederik A., Self-Awareness and Saliency of Social Versus Individualistic Behavioral Standards, Social Psychology 41 (2010), S. 10–19.
- Wójtowicz, Monika, Wirksame Anonymisierung im Kontext von Big Data, PinG 1 (2013), S. 65–69.
- Wolfram, Stephen, A new kind of science, Champaign, IL, 2002.
- Zeidler, Simon Alexander/Brüggemann, Sebastian, Die Zukunft personalisierter Werbung im Internet, CR 2014, S. 248–257.
- Zieger, Christoph/Smirra, Nikolas, Fallstricke bei Big Data-Anwendungen, Rechtliche Gesichtspunkte bei der Analyse fremder Datenbestände, MMR 2013, S. 418–421.
- Zuboff, Shoshana, Die neuen Massenausforschungswaffen, FAZ vom 13.2.2014, S. 33.

