Benedikt Buchner

Big Data und Datenschutz im Gesundheitswesen

Zusammenfassung

Big Data zeichnet sich dadurch aus, dass große Datenmengen frei von jeder Ziel- und Zwecksetzung gesammelt werden, um diese dann ergebnisoffen auf Verknüpfungsmuster und Korrelationen hin zu analysieren. Diese Philosophie von Big Data gerät in vielerlei Hinsicht in Konflikt mit grundlegenden Prinzipien des Datenschutzrechts, egal ob es um das datenschutzrechtliche Verbotsprinzip, den Grundsatz der Zweckbindung oder auch um mittelbare Schutzziele des Datenschutzrechts wie den Schutz vor Diskriminierung und den Schutz vor automatisierten Entscheidungen geht. Jedoch hält das Datenschutzrecht auch Lösungsansätze parat, wie sich Big Data und Datenschutz miteinander in Einklang bringen lassen. Von zentraler Bedeutung sind insoweit vor allem die datenschutzrechtliche Privilegierung der Datenverarbeitung zu Forschungszwecken, die Möglichkeit der Anonymisierung und Pseudonymisierung von Daten sowie das Instrument des broad consent.

Abstract/Summary

Big Data is characterised by the collection of large quantities of data without definition of an objective or purpose, in order to analyse these data sets with regard to patterns and correlations without any preconceived outcome. This philosophy of Big Data conflicts in many respects with the fundamental principles of data protection law, whether it is the principle of lawfulness, the principle of purpose limitation, or indirect objectives of data protection law such as protection against discrimination or protection against automated decision-making. However, data protection law also provides solutions for a balancing of Big Data and data protection. In particular the privileged treatment of data processing for research purposes, the techniques of anonymisation and pseudonymisation of data and the instrument of broad consent are of central importance in this context.

Schlüsselwörter

Big Data; Datenschutz; Zweckbindungsgrundsatz; Forschungsprivilegierung; Broad Consent.

Keywords

Big Data; data protection; purpose limitation; research exemption; broad consent.

DOI: 10.14623/zfme.2018.2.131-146

0. Einführung

Big Data ist auch im Gesundheitswesen ein schillernder Begriff. Big Data steht für große Erwartungen, die vor allem auf der Hoffnung basieren, dass dank der Verknüpfung unterschiedlichster Datenarten und ständig wachsender Datenmengen neue Erkenntnisse für die wissenschaftliche Forschung und medizinische Behandlung gewonnen werden können.[1] Stetig befeuert wird diese Hoffnung nicht zuletzt durch die Global Player der Datenverarbeitung wie etwa Google mit Versprechen wie dem, pro Jahr 100.000 Menschenleben zu retten, wenn erst einmal Big Data im Gesundheitswesen Fuß gefasst hat.[2] Zugleich wird Big Data aber auch immer wieder als Gefahr für Würde und (informationelle) Selbstbestimmung des Einzelnen angesehen, wenn Individuen nur noch auf bloße Datenpakete reduziert werden, die es bestmöglich zu analysieren und gewinnbringend zu nutzen gilt. Im Folgenden soll zunächst dargelegt werden, wo und wie der Konflikt zwischen Big Data und Datenschutz rechtlich zu verorten ist, dann aber auch aufgezeigt werden, wie sich dieser Konflikt – zumindest in Teilen – rechtlich wieder auflösen lässt.

1. Big Data: Versuch einer Definition

(Datenschutz-)Rechtlich ist Big Data schon deshalb schwer zu greifen, weil damit ein Phänomen beschrieben wird, das sich als äußerst vielgestaltig präsentiert, und daher die üblichen Definitionen bislang auch mehr oder weniger allgemein ausfallen.

1.1 Big Data und die drei »V«

Gemeinhin stützt sich eine Beschreibung von Big Data auf die sog. drei »V«: Volume, Variety und Velocity.[3] Es geht darum, riesige Datenmengen (Volume), die in unterschiedlichen Formaten vorliegen (Variety), in hoher Geschwindigkeit (Velocity) zu nutzen. Im Lauf der Zeit sind von den verschiedensten Seiten zunehmend mehr »V« an diese Definition angehängt worden, an erster Stelle Value (verstanden als die Möglichkeit einer *gewinnbringenden* Datennutzung),[4] aber auch Validity, Veracity, Vulnerability oder Volatility. Schon diese willkürlich anmutende Vermehrung der Definitionsmerkmale deutet darauf hin, dass Big Data möglicherweise mehr geduldiges Modewort als belastbares Begriffsmerkmal ist. Unabhängig davon stellt sich aber auch schon bei den ursprünglichen drei »V« die Frage, inwieweit hier Erscheinungsformen beschrieben werden, die Big Data von einer »normalen« Datenverarbeitung im Gesundheitswesen unterscheiden, oder ob es sich nicht vielmehr um Merkmale handelt, die auch schon vor dem Big-Data-Zeitalter die Datenverarbeitung im Gesundheitswesen geprägt haben.

Velocity: Geschwindigkeit ist ohne Zweifel ein Charakteristikum, welches Big Data besonders auszeichnet, insbesondere wenn Big Data in Echtzeit Daten verarbeitet. Eines der noch immer prominentesten Beispiele von Big Data im Gesundheitswesen zeichnet sich eben durch eine solch schnelle Datenverarbeitung aus: die Ermittlung der räumlichen Ausbreitung der H1N1-Epidemie (Schweinegrippe) durch Google, welches bestimmte Suchanfragen als Indikatoren für die Virus-Ausbreitung identifizieren konnte. Anders als herkömmliche Beobachtungsverfahren, die auf der zeitaufwendigen Samm-

lung und Analyse von Daten bei praktischen Ärzten beruhten und lediglich die (rückblickende) Erkenntnis brachten, wo das Virus in der Vergangenheit aufgetreten war, konnte Google mit der Auswertung der Suchanfragen den Grippeverlauf (fast) in Echtzeit verfolgen.[5] Velocity ist damit einerseits sicherlich ein typisches Merkmal von Big Data, andererseits jedoch auch kein Alleinstellungsmerkmal, da die Schnelligkeit in der Datenverarbeitung ganz allgemein als selbstverständliche Begleiterscheinung jeder weiteren Technisierung und Digitalisierung des Alltags im Gesundheitswesen eingeordnet werden kann.

Volume: Riesige Datenmengen sind ein Charakteristikum, welches die Datenverarbeitung im Gesundheitswesen schon seit jeher prägt, angefangen bei Behandlungsdaten über Abrechnungs- und Verwaltungsdaten bis hin zu Forschungsdaten. Seitdem es die ärztliche Dokumentationspflicht gibt, ist die Medizin eine datenintensive Disziplin, weil sich jeder Behandlungsprozess stets auch in einem entsprechenden Datenverarbeitungsprozess widerspiegelt. System und Recht der GKV sind weitere Quellen für einen Datenreichtum im Gesundheitssektor, wenn hier Daten zu den verschiedensten Zwecken erhoben und verarbeitet werden – zur Abrechnung, zu Zwecken der Wirtschaftlichkeitskontrolle oder Qualitätssicherung oder auch für besondere Versorgungsformen.[6] Die Liste lässt sich beliebig fortsetzen, verwiesen sei hier nur auf Forschungsdatenbanken und epidemiologische Studien, die ebenso wie etwa auch die sog. Omics-Forschung oder bildgebende Verfahren allesamt riesige Datenmengen produzieren.

Variety: Kennzeichnend für all diese Datenmengen im Gesundheitssektor ist schließlich deren Variety – im weiten Sinne verstanden als Vielfalt von Datentypen und Datenquellen. Und schon seit jeher kreisen die (rechtlichen und praktischen) Überlegungen darum, wie diese Datenvielfalt gebündelt und verknüpft werden kann, um sie – im Sinne von >Value< – möglichst gewinnbringend nutzen zu können, sei es für eine effizientere Verwaltung, sei es zum Wohle des Patienten. Dies mag mittels Instrumenten geschehen, die gemeinhin als typische Big-Data-Instrumente eingeordnet werden, z. B. die elektronische Gesundheitskarte, mit deren Hilfe Medienbrüche überwunden sowie die Behandlungssicherheit und Patientenautonomie verbessert werden sollen. An sich weist aber auch schon ein klassischer regulatorischer Ansatz wie der in § 73 SGB V, wonach der Hausarzt in der vertragsärztlichen Versorgung eine zentrale Stellung einnehmen und die gesamte Behandlungsdokumentation bündeln soll, typische Merkmale von Big Data auf; denn auch bei diesem Modell des Hausarztes als »Datensammelstelle«[7] geht es darum, die umfangreiche Behandlungsdokumentation (Volume), die sich bei den verschiedensten Leistungserbringern findet (Variety), im Sinne der Behandlungsqualität an einer Stelle zu bündeln.

Zunächst einmal lässt sich daher festhalten, dass Big Data im Sinne der drei »V« nicht mehr als ein Sammelbegriff ist, der für die verschiedensten Varianten des Umgangs mit großen und heterogenen Datenmengen herangezogen wird. Aus (datenschutz-)rechtlicher Sicht gibt der Begriff daher nur wenig her, er ist so vielgestaltig, dass sich von vornherein nicht die *eine* datenschutzrechtliche Herausforderung identifizieren bzw. Antwort geben ließe. Dementsprechend sind auch die für Big Data relevanten datenschutzrechtlichen Vorgaben den verschiedensten bereichsspezifischen Regelungen zu entnehmen – in erster Linie abhängig von der jeweiligen Datenquelle und verstreut über alle Rechtsgebiete, vom Recht der GKV über medizinrechtliche Gesetze wie das Gendiagnostikgesetz

bis hin zu bereichsspezifischen Regelungen für Forschungsdaten etwa in den Krebsregister- oder den Krankenhausgesetzen.

1.2 Big Data als zweckfreie Suche nach Korrelationen

Im Unterschied zur obigen Definition von Big Data im Sinne der drei »V« lässt sich Big Data allerdings auch noch anders und zwar abstrakter fassen, indem auf zwei Besonderheiten von Big Data abgestellt wird, die die Art und Weise des Umgangs mit Daten sowie die Erwartungshaltung bei der Datenverarbeitung betreffen. Herkömmlicherweise ist die Verarbeitung von Daten im Gesundheitswesen stets mit einer bestimmten Zielvorstellung oder Fragestellung verknüpft: Patientendaten werden zum Zwecke der ärztlichen Behandlung verarbeitet, Sozialdaten zum Zweck der Abrechnung im GKV-System, Forschungsdaten zum Zweck der Überprüfung von Hypothesen usw. Unter Big Data fehlt es dagegen an einer solchen Ziel- oder Zwecksetzung. Vielmehr zeichnet sich Big Data dadurch aus, dass Daten *ohne Ziel und Zweck* gesammelt und ausgewertet werden. Big Data steuert nicht zielgerichtet auf einen bestimmten Erkenntnisgewinn, auf die Schaffung einer problembezogenen Wissensgrundlage oder auf die Überprüfung einer bestimmten Hypothese zu. Stattdessen wird es unter Big Data Algorithmen überlassen, frei von jeder Ziel- und Zwecksetzung ergebnisoffen nach Verknüpfungsmustern zu suchen, Korrelationen in Datenbeständen zu finden und auf diese Weise Informationen zu gewinnen, deren Art und Umfang im Ausgangspunkt überhaupt nicht absehbar sind.[8]

Stellt man auf ein solches Verständnis von Big Data ab, ändert sich auch die datenschutzrechtliche Perspektive. Im Fokus stehen dann nicht mehr konkrete Big-Data-Szenarien, die unter ein bestimmtes bereichsspezifisches Regelungsregime fallen. Stattdessen geht es um die grundsätzliche Frage einer Vereinbarkeit der so skizzierten Big-Data-Philosophie mit zentralen Wertungen des Datenschutzrechts wie etwa dem Verbots- und dem Zweckbindungsgrundsatz oder dem Grundsatz der Datenminimierung und Speicherbegrenzung. Diese Grundsätze stehen in einem diametralen Gegensatz zur ergebnisoffenen und zweckfreien Vorratsdatenspeicherung, auf die Big Data angewiesen ist, und es stellt sich die Frage, ob und wie Datenschutz und Big Data trotz dieser Gegensätze miteinander in Einklang gebracht werden können.

2. Der Ausgangskonflikt: Big Data versus Datenschutz

Der Konflikt zwischen Big Data und Datenschutz lässt sich an zahlreichen (und grundlegenden) datenschutzrechtlichen Prinzipien festmachen: am datenschutzrechtlichen Verbotsprinzip mit Erlaubnisvorbehalt, am Grundsatz der Zweckbindung, an den Grundsätzen der Datenminimierung und Speicherbegrenzung sowie der Richtigkeit der Datenverarbeitung sowie schließlich auch an zwei Schutzzielen, die zwar nicht unmittelbar den Datenschutz im engeren Sinne betreffen, jedoch mittelbar mit diesem zusammenhängen, nämlich am Schutz vor Diskriminierung und vor einer automatisierten Entscheidungsfindung.

2.1 Verbotsprinzip

Big Data ist darauf angewiesen, dass Daten möglichst frei verfügbar sind. Umgekehrt hat jedoch das Datenschutzrecht deutscher und europäischer Prägung nicht die Freiheit der Verarbeitung personenbezogener Daten zum Ausgangspunkt, sondern deren Verbot. Jede Datenverarbeitung, auch die Weiterverarbeitung von bereits erhobenen Daten, ist nach Art. 6 Abs. 1 Datenschutz-Grundverordnung (DS-GVO) zunächst einmal unzulässig, es sei denn, die von der Datenverarbeitung betroffene Person hat in diese wirksam eingewilligt (Art. 6 Abs. 1 lit. a DS-GVO) oder die Datenverarbeitung lässt sich auf einen der sonstigen in Art. 6 Abs. 1 DS-GVO normierten Erlaubnistatbestände (lit. b – f) stützen. Für die Verarbeitung besonders sensibler Daten i.S.d. Art. 9 Abs. 1 DS-GVO ist das Verbotsprinzip noch einmal ausdrücklich in Art. 9 DS-GVO normiert. Art. 9 Abs. 1 DS-GVO »untersagt« eine Datenverarbeitung u. a. dann, wenn es sich um genetische Daten, Gesundheitsdaten oder Daten zum Sexualleben handelt.

Es überrascht vor diesem Hintergrund nicht, dass das Datenschutzrecht immer wieder als Hemmschuh für und als unvereinbar mit Big-Data-Anwendungen kritisiert wird, bis hin zu der Überzeugung, dass durch Big Data datenschutzrechtliche Grundsätze wie das Verbotsprinzip »ad absurdum geführt« würden.[9] Jedoch ist das datenschutzrechtliche Verbotsprinzip sowohl im deutschen als auch im europäischen Verfassungsrecht fest verankert: Nach Art. 8 Abs. 2 GRCh bedarf jede Verarbeitung personenbezogener Daten einer Einwilligung des Betroffenen oder einer gesetzlichen Legitimation, der Einzelne soll grundsätzlich »selbst Herr seiner Daten« sein.[10] Gewährleistet ist dies nur, wenn zumindest im Ausgangspunkt nicht das Prinzip der Verarbeitungsfreiheit, sondern das des Verarbeitungsverbots gilt.

Selbstverständlich trägt aber auch ein strenges Datenschutzrecht dem Umstand Rechnung, dass die Ziele und Interessen, die mit einer Verarbeitung personenbezogener Daten verfolgt werden, so gewichtig sein können, dass sie die Schutzbedürftigkeit personenbezogener Daten überwiegen. Eben deshalb sieht das Datenschutzrecht diverse Erlaubnistatbestände für eine Datenverarbeitung vor, u. a. eine allgemeine Interessenabwägungsklausel, nach der eine Datenverarbeitung auch dann zulässig ist, wenn diese zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich ist und nicht die Datenschutzinteressen der betroffenen Person überwiegen (Art. 6 Abs. 1 lit. f DS-GVO). Und auch Art. 9 Abs. 2 DS-GVO sieht für die Verarbeitung von besonders schutzwürdigen Daten eine Vielzahl von Ausnahmen vor, in denen das Verarbeitungsverbot nach Art. 9 Abs. 1 DS-GVO nicht gilt, unter anderem auch Ausnahmen für den Gesundheitssektor und für Zwecke der wissenschaftlichen Forschung (s. Art. 9 Abs. 2 lit. h, i und j DS-GVO). Das datenschutzrechtliche Verbotsprinzip ist damit also keineswegs gleichbedeutend mit einer unüberwindbaren Schranke für eine Datenverarbeitung, auch nicht für eine Datenverarbeitung im »großen« Stil. Wozu das Verbotsprinzip allerdings führt, ist eine Art von Verschiebung der Rechtfertigungslast – rechtfertigen muss sich nicht der Einzelne für den Schutz seiner personenbezogenen Daten, rechtfertigen (in Form eines Erlaubnistatbestandes) muss sich vielmehr der Verantwortliche für die Verarbeitung personenbezogener Daten.

2.2 Zweckbindung

Unter der DS-GVO ist der Grundsatz der Zweckbindung in Art. 5 Abs. 1 lit. b DS-GVO normiert. Personenbezogene Daten müssen danach »für festgelegte, eindeutige und rechtmäßige Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden«. Der Grundsatz der Zweckbindung ist nicht neu, sondern hierzulande schon seit dem Volkszählungsurteil des Bundesverfassungsgerichts als zentraler datenschutzrechtlicher Grundsatz fest etabliert. Insbesondere eine Datenverarbeitung »auf Vorrat« für unbestimmte Zwecke, also das, was Big Data in seinem Kern kennzeichnet, ist mit dem Zweckbindungsgrundsatz nicht vereinbar. Vielmehr muss der Zweck der Datenverarbeitung bereits im Moment der Datenerhebung feststehen und so präzise wie möglich bestimmt werden.

Der Zweckbindungsgrundsatz geht nicht so weit, dass eine Datenverarbeitung zu einem anderen als dem ursprünglich verfolgten Zweck überhaupt nicht zulässig ist. Der Verantwortliche darf die bei ihm bereits vorhandenen Daten vielmehr auch zu einem anderen als dem ursprünglichen Erhebungszweck weiterverarbeiten, allerdings nur unter zwei Voraussetzungen, die bei Big Data regelmäßig nicht erfüllt sein dürften: 1. Der Verantwortliche kann die Weiterverarbeitung von Daten zu einem anderen Zweck ebenfalls auf einen Erlaubnistatbestand stützen (Einwilligung der betroffenen Person oder gesetzlich normierter Erlaubnistatbestand). 2. Der Zweck der Weiterverarbeitung ist nicht unvereinbar mit dem ursprünglichen Erhebungszweck.

Was erstere Voraussetzung angeht, muss also auch im Fall einer zweckändernden Datenverarbeitung wieder die Hürde des datenschutzrechtlichen Verbotsprinzips genommen werden (s. soeben 2.1). Aber auch wenn erstere Voraussetzung erfüllt ist, ist des Weiteren auch noch Voraussetzung, dass der Zweck der Weiterverarbeitung *nicht unvereinbar* mit dem ursprünglichen Erhebungszweck ist. Für diese Frage der Vereinbarkeit bzw. Unvereinbarkeit führt dann Art. 6 Abs. 4 DS-GVO – nicht abschließend – eine Reihe von Beurteilungskriterien an. Gerade bei Big Data werden diese Kriterien überwiegend eher gegen als für eine Vereinbarkeit sprechen. Das gilt zum Beispiel für das Kriterium der »Verbindung« zwischen dem ursprünglichen und dem neuen Zweck (lit. a): Je weiter der Zweck der ursprünglichen Verarbeitung und der der Weiterverarbeitung auseinanderliegen, desto mehr spricht dies gegen eine Vereinbarkeit der beiden Zwecksetzungen.[11] Bei Big Data geht es aber gerade auch darum, überraschende und nicht erwartete Korrelationen zu finden. Besonders im Gesundheitsbereich problematisch ist auch das Kriterium der Art der personenbezogenen Daten (lit. c): Strenge Maßstäbe an die Vereinbarkeit sind vor allem dann anzulegen, wenn es sich um besondere Kategorien personenbezogener Daten i. S. v. Art. 9 DS-GVO wie Gesundheitsdaten handelt. Und auch die »möglichen Folgen« der beabsichtigten Weiterverarbeitung für den Betroffenen (lit. d) können gegen eine Vereinbarkeit sprechen, jedenfalls wenn Big Data darauf abzielt, einzelne Personen in einer bestimmten Art und Weise zu kategorisieren und daran – möglicherweise auch negative – Konsequenzen zu knüpfen (s. zu diesem Aspekt sogleich näher unter 2.5 und 2.6). In diesem Zusammenhang ist dann vor allem auch noch zu berücksichtigen, dass es für die betroffene Person bei Big Data regelmäßig besonders schwer ist, die Weiterverarbeitung ihrer Daten im Einzelnen nachvollziehen und ihre Folgen abschätzen zu können.

Insgesamt sind damit also die Hürden, die der datenschutzrechtliche Grundsatz der Zweckbindung für eine – zweckfreie – Verarbeitung personenbezogener Daten unter Big Data aufstellt, sehr hoch, weshalb der Zweckbindungsgrundsatz auch regelmäßig als besonders problematisch für eine Realisierung von Big-Data-Anwendungen angesehen wird. Eine Ausnahme gilt allerdings für die Datenverarbeitung zu wissenschaftlichen Forschungszwecken. Art. 5 Abs. 1 lit. b DS-GVO sieht hier eine Privilegierung der Datenverarbeitung vor, indem eine Vereinbarkeit mit den ursprünglichen Verarbeitungszwecken fingiert wird, wenn die Daten für wissenschaftliche Forschungszwecke weiterverarbeitet werden (ausführlicher dazu unten 3.2).

2.3 Datenminimierung und Speicherbegrenzung

Auch die datenschutzrechtlichen Grundsätze der Datenminimierung und der Speicherbegrenzung sind ganz offensichtlich nur wenig kompatibel mit der Philosophie von Big Data. Nach dem Grundsatz der Datenminimierung in Art. 5 Abs. 1 lit. c DS-GVO müssen personenbezogene Daten dem Zweck angemessen und erheblich sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein. Nach Möglichkeit sollen überhaupt keine personenbezogenen Daten verarbeitet werden oder zumindest nur so wenig personenbezogene Daten wie möglich. § 3a S. 2 BDSG (a. F.) hatte das allgemeine Gebot der Datenminimierung noch durch die Zielvorgabe der Anonymisierung und Pseudonymisierung konkretisiert: »Insbesondere sind personenbezogene Daten zu anonymisieren oder zu pseudonymisieren, soweit dies nach dem Verwendungszweck möglich ist und keinen im Verhältnis zu dem angestrebten Schutzzweck unverhältnismäßigen Aufwand erfordert.« Gleiches gilt künftig auch unter der DS-GVO: Lässt sich ein Verarbeitungszweck auch mit anonymisierten oder pseudonymisierten Daten erreichen, wäre eine Datenverarbeitung gerade nicht mehr i. S. d. Art. 5 Abs. 1 lit. c DS-GVO auf das »notwendige Maß beschränkt«, wenn ein Verantwortlicher auf diese Möglichkeiten nicht zurückgreift.

Mit dem Grundsatz der Datenminimierung eng zusammen hängt der Grundsatz der Speicherbegrenzung (Art. 5 Abs. 1 lit. e DS-GVO). Während ersterer Grundsatz auf den Umfang der Datenverarbeitung abzielt und diesen möglichst gering halten will, betrifft der Grundsatz der Speicherbegrenzung die zeitliche Dimension der Datenverarbeitung.[12] Stets soll eine Identifizierung der betroffenen Personen nur so lange möglich sein, wie es für die Verarbeitungszwecke erforderlich ist. Eine Ausnahme sieht der Grundsatz der Speicherbegrenzung allerdings unter anderem für eine Datenverarbeitung zu wissenschaftlichen Forschungszwecken vor. Letztere erfahren insoweit ebenso eine datenschutzrechtliche Privilegierung wie dies schon beim Zweckbindungsgrundsatz der Fall ist (s. soeben 2.2).

Abgesehen von dieser Privilegierung einer Datenverarbeitung zu Forschungszwecken bleibt es jedoch dabei, dass Big Data einerseits und Datenminimierung sowie Speicherbegrenzung andererseits nur schwer miteinander in Einklang zu bringen sind. Big Data ist auf möglichst viele Daten angewiesen, diese sollen nicht möglichst sparsam und zeitlich begrenzt, sondern stattdessen möglichst umfangreich und zeitlich unbefristet genutzt werden können, um einen maximalen Erkenntnisgewinn erzielen zu können.

2.4 *Richtigkeit*

Nach Art. 5 Abs. 1 lit. d DS-GVO müssen personenbezogene Daten »sachlich richtig und erforderlichenfalls auf dem neuesten Stand sein« (Grundsatz der Richtigkeit). Im Unterschied zu den zuvor dargestellten Grundsätzen ist dem Grundsatz der Richtigkeit bislang in der Diskussion um Big Data vergleichsweise wenig Aufmerksamkeit zuteil geworden. Nichtsdestotrotz ist gerade auch die Richtigkeit der Datenbestände unter Big Data eine besondere Herausforderung. Mit jeder Ausweitung des Datenbestands ist stets auch eine erhöhte Fehleranfälligkeit der Datenverarbeitung verbunden. Hinzu kommt das Problem der Intransparenz der Datenverarbeitung bei Big Data, die eine Kontrolle der Richtigkeit der Datenverarbeitung erschwert. Sichtbare Konsequenz einer Datenverarbeitung im Stile von Big Data ist oftmals lediglich eine Bewertung in Form eines bestimmten Scores oder eine Einordnung in eine bestimmte Kategorie (»vertrauenswürdiger Vertragspartner«, »gesunder Versicherungsnehmer« etc.). Bestimmt werden diese Bewertungen und Einordnungen durch hochkomplexe und geheime Algorithmen, die es für den einzelnen Betroffenen regelmäßig unmöglich machen, zu kontrollieren, ob die in den Algorithmus eingespeisten Daten »sachlich richtig« und »auf dem neuesten Stand« i. S. d. Art. 5 Abs. 1 lit. d DS-GVO sind.

2.5 *Diskriminierungsschutz*

Die hochkomplexen und geheimen Algorithmen, die typisch für eine Datenverarbeitung à la Big Data sind, machen Big Data nicht nur anfällig für Fehler, sondern auch für Diskriminierung. Die Kategorisierung von Personen auf Basis undurchsichtiger Datenbestände bringt die Gefahr mit sich, dass algorithmenbasierte Entscheidungen auch auf Merkmale gestützt werden, die ein besonderes Diskriminierungspotential bergen. Zweifelhafte Berühmtheit hat in der jüngeren Vergangenheit vor allem das Big-Data-Programm *Compas* erlangt. Dieses soll in den USA algorithmenbasiert die Wahrscheinlichkeit für die Rückfälligkeit oder auch Gefährlichkeit eines Straftäters berechnen und dementsprechend Richter bei ihren Entscheidungen beraten. Jedoch hat sich gezeigt, dass diese datengestützten »Berechnungen« mindestens ebenso diskriminierungs- und fehleranfällig wie menschliche Entscheidungen sind[13] – was dann aber vor allem deshalb problematisch ist, weil Big-Data-basierte Entscheidungen regelmäßig besonders intransparent sind und ihnen oftmals der Nimbus als besonders präzise und unbestechlich arbeitende Programme zukommt.

Zwar nicht unmittelbar, wohl aber mittelbar zielt auch der Datenschutz auf einen Schutz vor solcherlei diskriminierenden Effekten von Big Data ab – und das sogar besonders effektiv, indem das Datenschutzrecht mittels Verboten sicherstellt, dass Informationen, die ein besonderes Diskriminierungspotential bergen, erst gar nicht zur Kenntnis genommen werden können. Möglichen Diskriminierungen wird damit von vornherein die erforderliche Informationsgrundlage entzogen. Eingang in das Datenschutzrecht findet dieser Diskriminierungsschutz über die Brücke des Art. 9 DS-GVO. Regelungsgegenstand dieser Vorschrift sind die sog. besonderen Kategorien personenbezogener Daten, die vom Gesetzgeber als besonders schutzwürdig (»sensitiv«, »sensibel«) angesehen werden und deren Verarbeitung deshalb nochmals strengeren datenschutzrechtlichen

Anforderungen genügen muss. Art. 9 Abs. 1 DS-GVO zählt dabei zu den besonders schutzwürdigen Daten unter anderem auch genetische Daten und Gesundheitsdaten. Generell werden all diejenigen Merkmale als besonders schutzwürdige Daten eingeordnet, die auch als besonders diskriminierungsanfällig einzustufen sind und die sich daher weitestgehend identisch so auch in Art. 21 GRCh (Nichtdiskriminierung) wiederfinden.[14] Eine Verarbeitung dieser besonders schutzwürdigen Daten ist nach Art. 9 Abs. 1 DS-GVO grundsätzlich »untersagt« und dieses Verbotsprinzip fungiert insoweit also auch als »informationelles Diskriminierungsverbot«[15], Datenschutz wird zu einer Art von »präventivem Diskriminierungsschutz«.[16]

Die strengen Regeln, die das Datenschutzrecht für eine Verarbeitung besonders schutzwürdiger Daten vorsieht, haben in der datenschutzrechtlichen Diskussion zu der Frage geführt, ob und inwieweit dann eine Datenverarbeitung im Stile von Big Data überhaupt noch zulässig ist.[17] Zutreffend ist sicherlich, dass für den Fall der Verarbeitung besonders schutzwürdiger Daten das Verbotsprinzip des Datenschutzrechts nochmals akzentuiert wird (»Die Verarbeitung personenbezogener Daten [die i. S. v. Art. 9 besonders schutzwürdig sind] ist untersagt«). Allerdings sieht auch Art. 9 Abs. 2 DS-GVO (ebenso wie Art. 6 Abs. 1 DS-GVO für »normale« Daten) dann wieder zahlreiche Legitimationstatbestände vor, die eine Datenverarbeitung erlauben, u. a. auch für den Gesundheitssektor und für Zwecke der wissenschaftlichen Forschung (s. Art. 9 Abs. 2 lit. h, i und j DS-GVO). Entscheidend ist daher letztlich, inwieweit sich diese Erlaubnistatbestände auch für eine Datenverarbeitung im Stile von Big Data heranziehen lassen.

2.6 Automatisierte Entscheidungen

Problematisch ist Big Data schließlich vor allem auch dann, wenn automatisierte Datenverarbeitungs- und Entscheidungsprozesse dazu führen, das der Einzelne mittels kaum nachvollziehbarer Kriterien auf einzelne Persönlichkeitsmerkmale reduziert wird und hieran bestimmte – v. a. nachteilige – Konsequenzen geknüpft werden: Ein schlechter *Credit Score* führt dazu, dass jemand keinen oder nur einen Kredit zu ungünstigeren Konditionen bekommt, ein schlechter *Compas Score* (s. o. 2.5) führt dazu, dass jemand zu einer höheren Haftstrafe verurteilt wird, und möglicherweise wird irgendwann ein ungünstiger Lebenserwartungs-Score dazu führen, dass eine weitere Behandlung nicht mehr stattfindet, weil sich diese nicht mehr »lohnt«.[18]

All diese und ähnliche Szenarien sind für das Recht in vielerlei Hinsicht eine Herausforderung. Der Einzelne darf unter keinen Umständen zu einem bloßen Objekt computergestützter Programme werden, daher muss das Recht sicherstellen, dass Entscheidungen über eine Person nicht allein von Maschinen und Algorithmen getroffen werden – egal ob es um den Zugang zum Kreditmarkt oder die Inanspruchnahme ärztlicher Behandlung geht. Obwohl an sich keine spezifisch datenschutzrechtliche Herausforderung, hat diese – unter Big Data zentrale – Aufgabe für das Recht auch in das datenschutzrechtliche Regelungsregime Eingang gefunden und zwar unter dem Stichwort der automatisierten Entscheidung. Art. 22 Abs. 1 DS-GVO normiert insoweit ein grundsätzliches Verbot automatisierter Entscheidungen. Unzulässig ist es danach, dass Entscheidungen, die für den Einzelnen eine rechtliche Wirkung entfalten oder ihn erheblich beeinträchtigen, »ausschließlich« auf einer automatisierten Datenverarbeitung beruhen. *Zur Unter-*

stützung einer Entscheidung kann also durchaus auf Big Data zurückgegriffen werden, jedoch darf Big Data nicht *alleinige Grundlage* für eine Entscheidungsfindung sein. Rechtlich relevante oder nachteilige Entscheidungen dürfen also niemals allein durch Algorithmen »ohne jegliches menschliche Eingreifen« erfolgen (EG 71 der DS-GVO), im Ergebnis muss vielmehr die getroffene Entscheidung stets von einer natürlichen Person (statt von einer »Maschine«) zu verantworten sein.[19]

3. *Lösungsansätze: Big Data und Datenschutz*

Mit Blick auf die Vielzahl von datenschutzrechtlichen Regelungsprinzipien, die allesamt der Philosophie von Big Data mehr oder weniger diametral entgegenstehen, stellt sich die Frage, ob und wie unter dem geltenden Datenschutzrecht realisiert werden kann, was der Deutsche Ethikrat in seiner Stellungnahme zu Big Data und Gesundheit als idealtypische Zielsetzung formuliert hat: die Chancen von Big Data für die medizinische Forschung, die klinische Anwendung und das individuelle Gesundheitsverhalten zu nutzen und gleichzeitig die damit einhergehenden Risiken für die informationelle Freiheitsgestaltung auf ein Minimum zu reduzieren.[20] Welche Möglichkeiten insoweit bestehen, soll im Folgenden anhand einiger ausgewählter datenschutzrechtlicher Regelungsansätze beispielhaft aufgezeigt werden.

3.1 *Privilegierung der Forschung*

Die Chancen von Big Data gilt es im Gesundheitsbereich zuallererst für die medizinische Forschung zu nutzen. Die Zusammenführung und Analyse von immer größeren Datenmengen aus verschiedensten Quellen – von genetischen Daten über Bild- und klinische Daten bis hin zu Daten über Umwelteinflüsse und individuelle Lebensgewohnheiten – können dabei helfen, Korrelationen aufzudecken, biologische und medizinische Zusammenhänge zu verstehen und so die Prävention, Diagnose und Therapie in der medizinischen Praxis zu verbessern.[21] Datenschutz darf hier nicht zum Selbstzweck werden und den Schutz informationeller Selbstbestimmung einseitig überhöhen. Es darf gerade nicht der Eindruck erweckt werden, dass – wie es kritisch pointiert formuliert wird – unter dem Datenschutz »der Menschenwürde besser gedient zu sein [scheint], wenn Daten, die für die medizinische Forschung von hohem Wert sein können, vernichtet werden.«[22]

Eine solche Einseitigkeit zugunsten des Datenschutzes und zulasten der Forschung ist im Datenschutzrecht allerdings auch nicht angelegt. Im Gegenteil findet sich für die Datenverarbeitung zu wissenschaftlichen Forschungszwecken eine Vielzahl von Regelungsansätzen, die allesamt darauf abzielen, eine solche Datenverarbeitung datenschutzrechtlich zu privilegieren, angefangen bei der Lockerung des Zweckbindungsgebots über Ausnahmen vom grundsätzlichen Verbot einer Verarbeitung besonders schutzwürdiger Daten bis hin zu zahlreichen Ausnahmen bei den Betroffenenrechten.

Von all diesen Privilegierungen profitiert dann ebenso auch Big Data, wobei jedoch zu beachten ist, dass Big Data nicht ohne Weiteres ausnahmslos auch als privilegierte »Forschung« im Sinne des Datenschutzrechts einzuordnen ist. Allein der Umstand, dass Big Data mit einer Zusammenführung und Analyse von Daten stets auch auf eine wie auch

immer im Einzelnen ausgestaltete »wissenschaftliche Vorgehensweise« verweisen kann, reicht jedenfalls nicht aus.[23] Die DS-GVO spricht bewusst von wissenschaftlichen *Forschungs*zwecken in Abgrenzung zum allgemeineren Begriff der »wissenschaftlichen Zwecke«, um sicherzustellen, dass nicht jede Analyse und Aufbereitung von Daten im Zusammenhang mit Big Data schon eine Sonderbehandlung als »Wissenschaft« beanspruchen kann.[24] Die »wissenschaftliche Forschung« als solche ist dann allerdings weit zu verstehen und umfasst die Grundlagen- und angewandte Forschung ebenso wie die privat finanzierte Forschung (s. EG 159 der DS-GVO).

3.2 Lockerung des Zweckbindungsgrundsatzes

Eine Privilegierung der Datenverarbeitung zu wissenschaftlichen Forschungszwecken stellt vor allem die Lockerung des Zweckbindungsgrundsatzes dar. Art. 5 Abs. 1 lit. b Hs. 2 DS-GVO stellt für die Weiterverarbeitung von Daten zu wissenschaftlichen Forschungszwecken die Fiktion auf, dass diese Zwecke stets mit den ursprünglich verfolgten Zwecken vereinbar sind. Die (oben unter 2.2 aufgeführten) Voraussetzungen für die Annahme einer Zweckvereinbarkeit – die so bei Big Data regelmäßig nicht erfüllt sind – sind damit bei einer Datenverarbeitung zu wissenschaftlichen Forschungszwecken von vornherein nicht von Relevanz.

Auch im nationalen bereichsspezifischen Recht finden sich Vorschriften, die auf einen Ausgleich zwischen Datenschutz und Forschungsfreiheit abzielen. Zu nennen ist hier etwa § 75 SGB X, der eine Übermittlung von Sozialdaten erlaubt, soweit dies für ein bestimmtes Vorhaben der wissenschaftlichen Forschung im Sozialleistungsbereich erforderlich ist; auf diese Weise wird insbesondere auch der riesige »Datenschatz« der gesetzlichen Krankenversicherung für die wissenschaftliche Forschung eröffnet. Ein weiteres Beispiel sind die Forschungsklauseln im Landeskrankenhausrecht, die eine Verarbeitung von Patientendaten auch für wissenschaftliche medizinische Forschungsvorhaben unter bestimmten Voraussetzungen erlauben (s. z. B. § 7 BremKHDSG oder § 25 LKG-Bln).

3.3 Ausnahmen vom Verbot der Verarbeitung besonders sensibler Daten

Art. 9 Abs. 2 lit. j DS-GVO erlaubt dem mitgliedstaatlichen Gesetzgeber, Ausnahmen vom grundsätzlichen Verbot einer Verarbeitung von besonders schutzwürdigen Daten i. S. d Art. 9 zu normieren, wenn eine Datenverarbeitung für wissenschaftliche Forschungszwecke erforderlich ist. § 27 Abs. 1 S. 1 BDSG füllt diesen Regelungsspielraum in Form einer Interessenabwägungsklausel aus und erlaubt eine Datenverarbeitung für Forschungszwecke, »wenn die Verarbeitung zu diesen Zwecken erforderlich ist und Interessen des Verantwortlichen an der Verarbeitung die Interessen der betroffenen Person an einem Ausschluss der Verarbeitung erheblich überwiegen«. Die Vorschrift soll einen Freiraum für die Datenverarbeitung zu Forschungszwecken schaffen, von dem ebenso auch Big Data profitiert – vorausgesetzt, es fällt unter die wissenschaftliche Forschung und die mit der Forschung verfolgten Interessen überwiegen die Datenschutzinteressen der betroffenen Personen erheblich. Bei Big-Data-Forschung, die ausschließlich auf eine

Verbesserung von Prävention, Diagnose und Therapie in der medizinischen Praxis abzielt, ist davon regelmäßig auszugehen.

Davon zu unterscheiden ist die Big-Data-»Forschung« großer IT-Unternehmen wie Google, Amazon oder Apple, soweit diese Gesundheitsdaten verarbeiten.[25] Zwar mag auch deren Datenverarbeitung auf einen wissenschaftlichen Erkenntnisgewinn abzielen, damit einher geht jedoch regelmäßig auch eine kommerzielle Nutzung personenbezogener Daten, die entweder bereits von vornherein eine datenschutzrechtliche Privilegierung als »Forschung« ausschließt oder aber zumindest eine Interessenabwägung zuungunsten der datenverarbeitenden Stelle ausgehen lässt. Damit ist nicht gesagt, dass Google und Co. auf der Suche nach mehr Wissen und im Bemühen um bessere Gesundheit nicht auch Big Data einsetzen dürfen, jedoch müssen sie ihre Datenverarbeitung insoweit auf andere Art und Weise legitimieren, sei es durch eine Anonymisierung der Daten, sei es durch Einholung einer (wirksamen) Einwilligung der betroffenen Personen (s. zu diesen Alternativen sogleich).

3.4 Anonymisierung

Unproblematisch ist die Vereinbarkeit von Big Data und Datenschutz im Fall der Verarbeitung von anonymen Informationen. EG 26 der DS-GVO definiert anonyme Informationen als Informationen, die sich »nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, oder personenbezogene Daten, die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann.« Dass derlei Daten nicht in den Anwendungsbereich des Datenschutzrechts fallen, folgt schon daraus, dass ihnen jeglicher Personenbezug fehlt und damit gar keine »personenbezogenen« Daten vorliegen. Anonyme Informationen bzw. Daten sind also frei – und damit bestehen dann auch keine Schranken für die Verarbeitung dieser Daten unter Big Data.

Fraglich ist allerdings, ob es in Zeiten von Big Data überhaupt noch anonyme Daten im Sinne der obigen Definition geben kann. Regelmäßig ist im Zusammenhang mit Big Data von einem »Ende der Anonymität« die Rede, weil Analyse-Algorithmen jede Anonymisierung von Daten auf Dauer unmöglich machten.[26] Nie könne ausgeschlossen werden, dass mit Hilfe von Big-Data-Mechanismen anonyme Daten so kombiniert werden, dass im Ergebnis doch wieder ein Personenbezug hergestellt werden kann.[27]

Begründet sind solcherlei Befürchtungen sicherlich in all den Konstellationen, in denen eine Anonymisierung von Daten dergestalt aussieht, dass ein bestimmter Datensatz zwar keine klassischen personenidentifizierenden Merkmale aufweist, gleichwohl aber noch in der Form individualisiert ist, als er sich auf eine (wenngleich auch anonyme) *Einzel*person bezieht: das anonymisierte Nutzungsprofil eines Facebook-Profils, die anonymisierten Gesundheitsdaten eines Fitnesstrackers, das anonymisierte Suchprofil etc. Insoweit kann niemals sicher ausgeschlossen werden, dass bei einer Kombination von solcherlei Datensätzen und der Analyse mittels Big-Data-Mechanismen sich Muster erkennen lassen, die trotz an sich erfolgter Anonymisierung die Daten früher oder später doch wieder zu personenbezogenen Daten werden lassen.[28]

In vielen Fällen bedürfen Big-Data-Anwendungen allerdings ohnehin nicht solcher Einzel-Datensätze, sondern können sich vielmehr auf aggregierte Daten beschränken,

die gänzlich losgelöst von irgendwelchen (auch anonymen) Einzelpersonen sind. Verwiesen sei hier an dieser Stelle nochmals auf das eingangs erwähnte Beispiel der suchwortbasierten Prognose einer Grippeepidemie durch Google: Für diese prominente Big-Data-Anwendung reicht es vollkommen aus, die Suchworte als solche zu erheben und auszuwerten. Einer Rückkoppelung dieser Begriffe an irgendwelche (auch anonymisierten) Nutzer bedarf es hier von vornherein nicht. Wenn Google aber gleichwohl in diesem Zusammenhang anonymisierte, pseudonymisierte oder auch personenbezogene Suchprofile von Einzelpersonen erstellt, so geschieht dies nicht im Dienste der Gesundheit oder der Forschung, sondern zu Zwecken der kommerziellen Verwertung von Nutzerdaten. Insoweit besteht dann aber auch keinerlei Anlass, mit Verweis auf die Vorteile von Big Data irgendwelche Abstriche beim Datenschutz hinzunehmen.

3.5 Pseudonymisierung

§ 27 Abs. 3 BDSG, der für die Datenverarbeitung zu Forschungszwecken bestimmte flankierende Maßnahmen zur Wahrung des Datenschutzes normiert, schreibt grundsätzlich vor, dass personenbezogene Daten zu anonymisieren sind, sobald dies nach dem Forschungszweck möglich ist. In den Fällen jedoch, in denen der Forschungszweck einer Anonymisierung entgegensteht, sind die Daten zumindest zu pseudonymisieren. Definiert ist die Pseudonymisierung in Art. 4 Nr. 5 DS-GVO als die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können. Anders als bei der Anonymisierung von Daten existiert bei der Pseudonymisierung eine Zuordnungsregel, die es zumindest dem Kenner dieser Regel ermöglicht, die Pseudonymisierung wieder rückgängig zu machen und den Personenbezug der Daten wieder herzustellen.[29]

Auch die Pseudonymisierung personenbezogener Daten kann also ein Weg sein, Big Data und Datenschutz miteinander in Einklang zu bringen. So sind Daten beispielsweise dann zu pseudonymisieren, wenn im Zuge von Langzeitstudien eine fortlaufende Zuordnung neuer Daten zu bereits vorhandenen Daten erforderlich ist und daher eine Anonymisierung der Daten mit dem Forschungszweck nicht vereinbar wäre.[30] Eine Pseudonymisierung statt einer Anonymisierung von Daten bietet sich auch dann an, wenn die Möglichkeit erhalten bleiben soll, die betroffene Person erneut zu kontaktieren.[31] Und schließlich ist eine Anonymisierung von Daten von vornherein unmöglich, wenn es sich um Biomaterialien oder genetische Daten handelt, denen aufgrund der enthaltenen Erbinformationen ein Personenbezug inhärent ist.[32]

3.6 Einwilligung

Legitimieren lässt sich eine Datenverarbeitung unter Big Data darüber hinaus stets auch mit einer entsprechenden Einwilligung der betroffenen Person – vorausgesetzt, die Einwilligung erfüllt alle Wirksamkeitsvoraussetzungen, die das Datenschutzrecht aufstellt. Problematisch sind insoweit bei Big-Data-Anwendungen vor allem die Voraussetzungen der Informiertheit und Zweckbestimmtheit der Einwilligung. Beide Anforderungen las-

sen sich bei Big Data mit Blick auf die Zweckfreiheit und Ergebnisoffenheit der Datenverarbeitungsprozesse oftmals nur schwer umsetzen.

Einen Ausweg bietet hier jedoch der sogenannte *broad consent* – zumindest soweit es um eine Datenverarbeitung zu wissenschaftlichen Forschungszwecken geht. Nach der Rechtsfigur des *broad consent* kann eine Einwilligung im Fall einer Datenverarbeitung zu Forschungszwecken auch »breiter«, d.h. unbestimmter ausfallen und sich auch allgemein auf bestimmte Bereiche wissenschaftlicher Forschung erstrecken.[33] Auch in der DS-GVO findet sich dieser Ansatz wieder. Nach EG 33 der DS-GVO soll eine – entsprechend allgemeine – Einwilligung auch »für bestimmte Bereiche wissenschaftlicher Forschung« zulässig ein, wenn dabei die anerkannten ethischen Standards der wissenschaftlichen Forschung eingehalten werden. Voraussetzung ist lediglich, dass es für die betroffene Person möglich ist, ihre Einwilligung »auf bestimmte Forschungsbereiche oder Teile von Forschungsprojekten in dem vom verfolgten Zweck zugelassenen Maße« zu beschränken (EG 33 a.E.).

Mit der Rechtsfigur des *broad consent* wird der Tatsache Rechnung getragen, dass sich bei neuen Forschungsprojekten die genaue Zielsetzung nicht von vornherein festlegen lässt, und auch dem Umstand, dass in der Forschung personenbezogene Daten oftmals in großem Umfang, beispielsweise in Biobanken, gesammelt werden, um diese für künftige Studien vorzuhalten. Die Einwilligung der betroffenen Personen in eine Datenverarbeitung zu Forschungszwecken muss hier notwendigerweise allgemeiner ausfallen. Gleiches gilt aber auch für Big Data, weshalb auch hier auf den *broad consent* als Legitimationsgrundlage für eine Datenverarbeitung zurückgegriffen werden kann, vorausgesetzt, die konkrete Big-Data-Anwendung lässt sich als wissenschaftliche Forschung einordnen.

4. Fazit

Big Data und Datenschutz mögen sich auf den ersten Blick als nur schwer oder gar nicht miteinander vereinbar präsentieren. Bei näherem Hinsehen ist das Verhältnis zwischen beiden jedoch differenzierter zu beurteilen. Die eine »große Lösung«, wie sich Big Data und Datenschutz miteinander in Einklang bringen lassen, gibt es sicherlich nicht. Und in Anbetracht der Tatsache, dass sich das Datenschutzrecht mit der DS-GVO gerade eben erst neu formiert hat, scheint es auch wenig realistisch, dass in absehbarer Zeit ein neues Big-Data-kompatibles Datenschutzrecht kommen wird. Die Herausforderung wird daher primär darin zu sehen sein, in punktueller Feinarbeit an den verschiedensten Stellschrauben des Datenschutzrechts zu drehen, um je nach konkreter Big-Data-Anwendung mittels entsprechender Auslegung der einschlägigen datenschutzrechtlichen Regelungen zu einem interessengerechten Ausgleich zwischen Big Data und Datenschutz zu kommen. Dass dies möglich ist, wurde hier beispielhaft anhand einer Reihe von datenschutzrechtlichen Regelungsansätzen aufgezeigt. Diese lassen für Big-Data-Anwendungen durchaus einen weiten Spielraum, zumindest dann, wenn mit Big Data wissenschaftliche Forschungszwecke verfolgt werden. Der oben angesprochenen Zielsetzung, zum einen die Chancen von Big Data für die medizinische Forschung, die klinische Anwendung und das individuelle Gesundheitsverhalten zu nutzen und zum anderen aber auch

das Recht auf informationelle Selbstbestimmung so weit wie möglich zu wahren,[34] kommt man damit auch schon nach geltendem Recht ziemlich nahe.

ANMERKUNGEN

1 Vgl. DEUTSCHER ETHIKRAT, *Stellungnahme: Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung*, 30.11.2017, hier: 48.

2 So jedenfalls das Versprechen von Larry Page, ehemals Vorstandschef von Google und nunmehr von Googles Muttergesellschaft Alphabet; zit. nach M. MÜHL, *Sensible Gesundheitsdaten – Die Vermessung des Körpers*, in: FAZ v. 17. Juli 2014.

3 Siehe etwa T. WEICHERT, *Big Data, Gesundheit und der Datenschutz*, in: Datenschutz und Datensicherheit 38 (2014) 831–838.

4 Vgl. WEICHERT (Anm. 3); vgl. P. RICHTER, *Big Data, Statistik und die Datenschutz-Grundverordnung*, in: Datenschutz und Datensicherheit 40 (2016) 581–586.

5 Vgl. V. MAYER-SCHÖNBERGER, *Big Data – Eine Revolution, die unser Leben verändern wird*, in: Bundesgesetzblatt 58 (2015) 788–793.

6 S. den Überblick bei WEICHERT (Anm. 3).

7 H.-D. SPROLL, in: D. Krauskopf, Soziale Krankenversicherung, Pflegeversicherung, München, Stand: Nov. 2017 § 73 SGB V Rn. 18.

8 Vgl. K.-H. LADEUR, *Wissenserzeugung im Sozialrecht und der Aufstieg von »Big Data«*, in: B. Buchner/K.-H. Ladeur, Wissensgenerierung und -verarbeitung im Gesundheits- und Sozialrecht, Tübingen 2016, 89–105, hier: 93; vgl. I. PIGEOT/S. JACOBS/U. KOCH-GROMUS, *Große Datensammlungen im Gesundheitswesen – Chance oder Risiko?*, in: Bundesgesetzblatt 58 (2015) 785–787.

9 Vgl. J. SCHNEIDER, *Datenschutz nach der EU-Datenschutz-Grundverordnung*, München 2017, 164.

10 Vgl. W. FRENZ, *Handbuch Europarecht*, Berlin/Heidelberg 2009, Bd. 4, Rn. 1380; vgl. ähnlich R. STREINZ/W. MICHL, *Die Drittwirkung des europäischen Datenschutzgrundrechts (Art. 8 GRCh) im deutschen Privatrecht*, in: Europäische Zeitschrift für Wirtschaftsrecht (2011) 384–388, hier: 385.

11 Art.-29-Datenschutzgruppe, *Opinion 03/2013 on purpose limitation*, WP 203 (2013), 23.

12 Vgl. T. HERBST, in: J. Kühling/B. Buchner (Hrsg.), DS-GVO, München 2017, Art. 5 Rn. 64.

13 Vgl. Bertelsmann Arbeitspapier, *Wenn Maschinen Menschen bewerten. Internationale Fallbeispiele für Prozesse algorithmischer Entscheidungsfindung*, Gütersloh 2017, hier: 10 f.

14 Vgl. dazu auch T. WEICHERT, *»Sensitive Daten« revisited*, in: Datenschutz und Datensicherheit 41 (2017) 538–543, hier: 539.

15 T. WEICHERT, in: J. Kühling/B. Buchner (Hrsg.), DS-GVO, München 2017, Art. 9 Rn. 2.

16 V. RIEBLE, Anm. zu BAG (11.11.1993), EzA Nr. 40 zu § 123 BGB, 13.

17 Vgl. etwa J. SCHNEIDER, *Schließt Art. 9 DS-GVO die Zulässigkeit der Verarbeitung bei Big Data aus?*, in: Zeitschrift für Datenschutz 7 (2017) 303–308.

18 S. zum *Compas Score*: Bertelsmann Arbeitspapier (Anm. 13), hier: 9; zur Berechnung der Lebenserwartung mittels Algorithmen s. A. LOBE, *Lebenserwartung – Der Algorithmus schlägt die letzte Stunde*, in FAZ v. 8.1.2017.

19 Vgl. B. BUCHNER, in: J. Kühling/B. Buchner (Hrsg.), DS-GVO, München 2017, Art. 22 Rn. 15.

20 Vgl. DEUTSCHER ETHIKRAT (Anm. 1), hier: 41.

21 Ausführlich zum Einsatz von Big Data in der biomedizinischen Forschung: DEUTSCHER ETHIKRAT (Anm. 1), hier: 90 ff.

22 LADEUR (Anm. 8).

23 Vgl. ausführlich dazu P. RICHTER, *Datenschutz zwecklos? – Das Prinzip der Zweckbindung im Ratsentwurf der DSGVO*, in: Datenschutz und Datensicherheit 39 (2015) 735–740, hier: 737f.; s.a. B. BUCHNER, *Grundsätze und Rechtmäßigkeit der Datenverarbeitung unter der DS-GVO*, in: Datenschutz und Datensicherheit 40 (2016) 155–161, hier: 157; vgl. A. ROSSNAGEL/M.

Nebel/P. Richter, *Was bleibt vom Europäischen Datenschutzrecht? – Überlegungen zum Ratsentwurf der DS-GVO*, in: Zeitschrift für Datenschutz 5 (2015) 455–460, hier: 457 f.

[24] Vgl. J. P. Albrecht/F. Jotzo, *Das neue Datenschutzrecht der EU*, Baden-Baden 2017, Teil 3 Rn. 71.

[25] Siehe dazu Deutscher Ethikrat (Anm. 1), hier: 113 ff.

[26] Vgl. V. Boehme-Nessler, *Das Ende der Anonymität*, in: Datenschutz und Datensicherheit 40 (2016) 419–423.

[27] Vgl. M. Sarunski, *Big Data – Ende der Anonymität?*, in: Datenschutz und Datensicherheit 40 (2016) 424–427.

[28] Vgl. dazu Sarunski (Anm. 27), hier: 426 f.

[29] Vgl. M. Karg, *Anonymität, Pseudonyme und Personenbezug revisited?*, in: Datenschutz und Datensicherheit 41 (2015) 520–526, hier: 521.

[30] Vgl. M. Krawczak/T. Weichert, *Vorschlag einer modernen Infrastruktur für die medizinische Forschung in Deutschland*, Thesenpapier, 2017, hier: 7.

[31] Vgl. zu Biobanken T. Herbst, *Rechtliche und ethische Probleme des Umgangs mit Proben und Daten bei großen Biobanken*, in: Datenschutz und Datensicherheit 40 (2016) 371–375, hier: 373.

[32] Vgl. M. Krawczak/T. Weichert (Anm. 30), hier: 7.

[33] Vgl. Herbst (Anm. 31), hier: 373 zum *broad consent* bei Biodatenbanken.

[34] Vgl. Deutscher Ethikrat (Anm. 1), hier: 41.