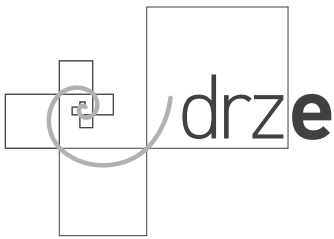
Ethik in den Biowissenschaften –
Sachstandsberichte des DRZE

Band 22: Big Data in der Medizin



Im Auftrag des
Deutschen Referenzzentrums für Ethik in den Biowissenschaften

herausgegeben von
Dieter Sturma und Dirk Lanzerath

www.drze.de

VERLAG KARL ALBER

Laura Summa / Ulrich Mansmann /
Benedikt Buchner / Maximilian Schnebbe

Big Data in der Medizin

Konzeptionelle, rechtliche und ethische Aspekte

Verlag Karl Alber Freiburg/München

Diese Publikation wird als Vorhaben der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste im Rahmen des Akademienprogramms von der Bundesrepublik Deutschland und dem Land Nordrhein-Westfalen gefördert.

DOI: 10.23769/vka-2020-49191

Originalausgabe

www.verlag-alber.de

Redaktion: Aurélie Halsband

Satz und PDF-E-Book: SatzWeise, Bad Wünnenberg
Herstellung: CPI books GmbH, Leck

Printed in Germany

ISBN 978-3-495-49191-1

Inhalt

II. Big Data in der Medizin: Rechtliche Aspekte

Benedikt Buchner, Maximilian Schnebbe

1. Einleitung

Big Data in der Medizin wirft eine Vielzahl von rechtlichen Fragestellungen auf, etwa welche Maßstäbe unter den Bedingungen einer auf Big Data basierenden Systemmedizin für die Festlegung des medizinischen Behandlungsstandards gelten sollen oder wer unter welchen Voraussetzungen haftungsrechtlich verantwortlich ist, wenn es beim Einsatz von Big-Data-Systemen zu Behandlungsfehlern kommt.[1] Auch dem Recht auf Nichtwissen kommt im Fall von Big Data noch einmal eine besondere Bedeutung zu, wenn sich die bisherige »Informationshoheit des Patienten«[2] im Behandlungskontext angesichts der Kombination von Big Data und algorithmengestützten Programmen zunehmend schwerer aufrechterhalten lässt. So vielgestaltig diese und andere rechtliche Problemstellungen sind, steht jedoch ganz im Zentrum der rechtlichen Diskussion um Big Data und Medizin die Frage des Datenschutzes. Big Data in der Medizin basiert in weitem Umfang auf *personenbezogenen* Daten; es geht um die Daten von Patientinnen und Patienten, Probandinnen und Probanden oder Versicherten, denen noch dazu in ihrer Ausprägung als Gesundheitsdaten eine besondere Sensibilität im Datenschutzrecht beigemessen wird (Art. 9 Abs. 1 DS-GVO: sogenannte besondere Kategorien personenbezogener Daten).

Im Ausgangspunkt handelt es sich bei Big Data und Datenschutz um zwei Zielsetzungen, wie sie unterschiedlicher kaum ausfallen könnten. Während Big Data darauf ausgerichtet ist, möglichst viel an Daten frei von jeder Ziel- und Zwecksetzung zu sammeln und zu analysieren, zielt Datenschutz darauf ab, die Verarbeitung von personenbezogenen Daten möglichst streng zu reglementieren und dem Einzelnen die Befugnis zu sichern, selbst über das Ob und Wie einer Verarbeitung »seiner« per-

[1] Vgl. Katzenmeier 2019: 259 ff.
[2] Hahn 2019: 197.

sonenbezogenen Daten zu bestimmen – ganz im Sinne des Rechts auf informationelle Selbstbestimmung, wie es seit der Volkszählungsentscheidung des Bundesverfassungsgerichts das Verständnis von Datenschutz hierzulande prägt.[3]

Festmachen lässt sich das Spannungsverhältnis zwischen Big Data und Datenschutz schon am Grundprinzip des deutschen und europäischen Datenschutzrechts: dem Verbotsprinzip mit Erlaubnisvorbehalt. Datenschutzrechtlicher Ausgangspunkt ist nicht die Freiheit der Verarbeitung personenbezogener Daten, sondern deren Verbot. Jede Datenverarbeitung, auch die Weiterverarbeitung von bereits erhobenen Daten, ist nach Art. 6 Abs. 1 der Europäischen 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. Mit Blick auf die Verarbeitung von Gesundheitsdaten als besonders sensible Daten wird das Verbotsprinzip nochmals in Art. 9 Abs. 1 DS-GVO betont, indem dort eine Verarbeitung von Gesundheitsdaten (ebenso wie etwa auch von genetischen Daten oder Daten zum Sexualleben) ausdrücklich »untersagt« wird.

Mit dem Ansinnen von Big Data, möglichst viele Daten möglichst frei nutzen zu können, ist dieser datenschutzrechtliche Ausgangspunkt offensichtlich nur schwer vereinbar. Entsprechend wird das Datenschutzrecht in geltender Prägung auch immer wieder als fortschrittsfeindlich kritisiert, bis hin zu der Überzeugung, dass spätestens durch Big Data althergebrachte datenschutzrechtliche Grundsätze wie das Verbotsprinzip »ad absurdum geführt« würden.[4] Wenn das Datenschutzrecht hier teils unter erheblichen Rechtfertigungsdruck gerät, liegt das vor allem auch an den großen Erwartungen, die mit Big Data verbunden werden. Die Hoffnung ist, dass mittels der Verknüpfung unterschiedlichster Datenarten und ständig wachsender Datenmengen neue Erkenntnisse für die wissenschaftliche Forschung und medizinische Behandlung gewonnen werden können.[5] Dabei gerät leicht aus dem Blick, dass Big Data ungeachtet aller Chancen auch erhebliche Risiken birgt – aus datenschutz- und persönlichkeitsrechtlicher Perspektive in erster Linie für die Würde und informationelle Selbstbestimmung des Einzelnen, wenn dieser nur noch auf

[3] Vgl. Bundesverfassungsgericht 1983: 1ff.
[4] Schneider 2017: 164.
[5] Vgl. Deutscher Ethikrat 2018: 48.

bloße Datenpakete reduziert wird, die es bestmöglich zu analysieren und gewinnbringend zu nutzen gilt.[6]

2. Ausgangsfrage: Big Data als Verarbeitung *personenbezogener* Daten?

Die Ziele von Big Data geraten nicht ausnahmslos mit den Grundprinzipien des Datenschutzrechts in Konflikt, sondern nur dann, wenn überhaupt der datenschutzrechtliche Anwendungsbereich eröffnet ist, also konkret dann, wenn die Daten, die verarbeitet werden, einen Personenbezug aufweisen. Nach Art. 4 Nr. 1 DS-GVO liegen *personenbezogene* Daten immer dann vor, wenn sich Informationen auf eine »identifizierte oder identifizierbare« natürliche Person beziehen, wenn also zumindest die Möglichkeit besteht, anhand bestimmter Identifizierungsmerkmale die Identität einer Person festzustellen, auf die sich die fraglichen Daten beziehen.

Datenschutzrechtlich irrelevant sind damit von vornherein all diejenigen Big-Data-Anwendungen, die sich auf die Auswertung von anonymen Informationen ohne Personenbezug beschränken. Dazu zählen insbesondere Big-Data-Anwendungen, die keiner Einzel-Datensätze zu bestimmten Personen bedürfen, sondern sich auf aggregierte Daten stützen können, die keine Verbindungen mehr zu (bestimmten oder bestimmbaren) Einzelpersonen aufweisen. Erwägungsgrund (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.« Solche anonymen Informationen bzw. Daten sind datenschutzrechtlich betrachtet »Allgemeingut« und unterliegen keinerlei Schranken für eine Nutzung mittels Big Data.

2.1 *Beispiel: Schweinegrippe einerseits, Covid-19 andererseits*

Eine dem Grunde nach anonym ausgestaltete Big-Data-Anwendung ist etwa das prominente Beispiel der suchwortbasierten Prognose einer Influenzapandemie durch Google: Im Fall der H1N1-Pandemie von 2009/10 (Schweinegrippe) gelang es Google, die räumliche Ausbreitung dieser

[6] Vgl. Buchner 2018: 132.

Grippe nachzuverfolgen, indem das Unternehmen bestimmte Suchanfragen als Indikatoren für die Virus-Ausbreitung identifizierte. Im Unterschied zu den bis dato üblichen Beobachtungsverfahren konnte Google diese Virus-Ausbreitung so gut wie in Echtzeit analysieren, weil diese Methode des Big Data nicht auf das zeitaufwendige Sammeln und Analysieren von Daten vor Ort angewiesen war (die zudem auch nur rückblickende Erkenntnisse zum Verbreitungsgrad zuließen).[7]

Aus Perspektive des Datenschutzes ist solch eine Big-Data-Anwendung irrelevant – bzw. kann zumindest so konzipiert werden, dass sie irrelevant ist. Google war für die Echtzeitbeobachtung der Virus-Ausbreitung zu keinem Zeitpunkt darauf angewiesen, irgendwelche personenbezogenen Daten zu erheben. Ausreichend war vielmehr allein die Erhebung und Auswertung der Suchworte als solche, ohne dass es darüber hinaus noch einer Verknüpfung dieser Suchworte mit einer bestimmten oder bestimmbaren Person bedurft hätte. Falls Google gleichwohl in diesem Zusammenhang möglicherweise anonymisierte, pseudonymisierte oder auch personenbezogene Suchprofile von Einzelpersonen erstellt haben sollte, hatte dies mit der eigentlichen Zwecksetzung der Big-Data-Anwendung nichts mehr zu tun, sondern geschah allein zu Zwecken einer auch noch kommerziellen Verwertung von Nutzerdaten.[8]

An sich eine vergleichbare Konstellation betreffend präsentiert sich Big Data im aktuellen Fall des Coronavirus Sars-CoV-2 gleichwohl in einem ganz anderen Licht und vor allem mit erheblicher datenschutzrechtlicher Sprengkraft. Anders als im Fall von H1N1 beschränken sich die Big-Data-Instrumentarien bei Sars-CoV-2 keineswegs darauf, auf Grundlage aggregierter Daten allgemeine Aussagen über den Verlauf der Pandemie zu gewinnen. Diskutiert – und in China bereits praktiziert – wird stattdessen die lückenlose individuelle Überwachung von Millionen Einzelpersonen, um auf diese Weise die Ausbreitung von Sars-CoV-2 mit Hilfe von Big Data zu bekämpfen.[9] Mittels Apps wie »Close Contact Detector« oder »Health Code« werden sämtliche Bewegungen der Nutzerin oder des Nutzers aufgezeichnet, um auf dieser Grundlage Bewegungsspuren mit den Aufenthaltsorten von Infizierten zu vergleichen und dementsprechend die Nutzerin oder den Nutzer in die Kategorien grün (freie Bewegung), gelb (7-Tage-Quarantäne) oder rot (2-Wochen-Isolation) einzuordnen.[10]

---

[7] Vgl. Mayer-Schönberger / Cukier 2013: 789.
[8] Vgl. Buchner 2018: 14.
[9] Vgl. Böge 2020: 7.
[10] Vgl. Heller 2020: 72.

Vor- und Nachteile dieser Nutzung von Big Data werden unterschiedlich beurteilt. Nach Einschätzung der WHO hat sie dazu beigetragen, Quarantäne-Maßnahmen erfolgreich umzusetzen.[11] Auf der anderen Seite stehen erhebliche Eingriffe in die Rechte und Freiheiten der betroffenen Personen: die umfassende Vollzeitüberwachung auf Grundlage besonders sensibler Daten sowie die Ungenauigkeit und Fehleranfälligkeit von automatisierten Entscheidungen, die noch dazu ein erhebliches Diskriminierungs- und Stigmatisierungspotenzial haben.

2.2 *Generell: Anonyme Daten in Zeiten von Big Data?*

Obiges Beispiel der Rundumüberwachung bei Sars-CoV-2 ist aus Datenschutzperspektive das Worst-Case-Szenario einer Big-Data-Anwendung, andererseits aber auch nicht das ›typische‹ Beispiel für eine Datenverarbeitung im Dienste von Big Data. Mit Big Data in der Medizin mag zwar grundsätzlich ein Paradigmenwechsel einhergehen – weg von einem indikationsorientierten Ansatz hin zu einer personalisierten Medizin: Ziel ist nicht mehr die beste Therapie für eine bestimmte Indikation, sondern die beste Therapie für eine bestimmte Person.[12] Gleichwohl führt aber auch ein solcher Paradigmenwechsel nicht dazu, dass medizinische Datenverarbeitung im Dienste von Big Data stets auf die individuelle Überwachung von Einzelpersonen abzielt. Ziel ist vielmehr zuallererst der *medizinische* Erkenntnisgewinn. Auch wenn es hierfür im Ausgangspunkt einer Verarbeitung personenbezogener Daten bedarf, so ist es doch in vielen Konstellationen möglich, diesen Personenbezug im weiteren Prozess der Wissensgenerierung wieder aufzuheben und so auch den Datenschutzbelangen der betroffenen Personen Rechnung zu tragen.

Eben diesen Weg gibt auch das geltende Datenschutzrecht insbesondere für die Datenverarbeitung zu Forschungszwecken vor. Art. 89 Abs. 1 DS-GVO verlangt hier technisch-organisatorische Maßnahmen, die den Grundsatz der sogenannten Datenminimierung gewährleisten sollen. Beispielhaft führt Art. 89 Abs. 1 DS-GVO die Pseudonymisierung von Daten an. Jedoch ist vorgeschaltet stets zunächst einmal zu überprüfen, ob der Forschungszweck nicht auch schon mit anonymisierten Daten erreicht werden kann. Eben in diesem Sinne regelt im nationalen Recht § 27

[11] Vgl. ibid.

[12] Vgl. Katzenmeier 2019: 260. Vgl. auch Abschnitt 2 (»Digitalisierung und das lernende Gesundheitssystem«) des Teils 1 (Konzeptionelle, organisatorische und technische Aspekte) des vorliegenden Sachstandsberichts.

Abs. 3 Bundesdatenschutzgesetz (BDSG) für die Forschung mit besonderen Kategorien personenbezogener Daten (also etwa mit Gesundheitsdaten), dass diese Daten zu anonymisieren sind, sobald dies nach dem Forschungszweck zulässig ist.

Fraglich ist in diesem Zusammenhang allerdings, ob es in Zeiten von Big Data überhaupt noch möglich ist, Daten so zu anonymisieren, dass im Sinne von EG 26 DS-GVO »die betroffene Person nicht oder nicht mehr identifiziert werden kann«. Je mehr Daten verarbeitet werden, desto größer ist die Wahrscheinlichkeit einer Re-Identifizierung, wenn (an sich anonymisierte) Daten mit anderen Datenbeständen zusammengeführt werden.[13] So ist für das Big-Data-Zeitalter auch schon ein »Ende der Anonymität« ausgerufen worden, weil die Kombination von immer mehr Daten und immer leistungsfähigeren Algorithmen jede Anonymisierung von Daten auf Dauer unmöglich machten.[14] Niemals könne vollständig ausgeschlossen werden, dass mit Hilfe von Big-Data-Mechanismen anonyme Daten so kombiniert werden, dass im Ergebnis ein Personenbezug wieder hergestellt werden kann.[15]

Zu befürchten ist eine solche Re-Identifizierung allerdings in erster Linie in den Konstellationen, in denen bei einem bestimmten Datensatz zwar die personenidentifizierenden Merkmale entfernt worden sind, der Datensatz gleichwohl aber noch in der Form individualisiert ist, dass er sich auf eine (wenn auch anonyme) Einzelperson bezieht. Hier kann niemals sicher ausgeschlossen werden, dass sich mittels einer Zusammenführung von Datenbeständen und des Einsatzes von Big-Data-Analysen bestimmte Muster erkennen lassen, die den Personenbezug von an sich einmal anonymisierten Daten wieder aufleben lassen.[16] ›Sicherer‹ ist die Anonymisierung hingegen bei all den Big-Data-Anwendungen, die sich allein auf aggregierte Daten beschränken, die keinerlei Bezug mehr zu irgendwelchen *Einzel*personen haben.

Die rechtliche Vorgabe, so weit wie möglich mit anonymisierten Daten zu arbeiten, ist vor allem auch eine technische Herausforderung. Ulrich Mansmann stellt in seinem Beitrag zu den konzeptionellen, technischen und organisatorischen Aspekten von Big Data in der Medizin eine Vielzahl von technischen Verfahren vor, mittels derer sichergestellt werden soll, dass es nicht zu einer Re-Identifikation von Einzelpersonen kommt. Zu diesen Verfahren zählt er u. a. *Differential Privacy*, Verteiltes

[13] Vgl. Watteler / Kinder-Kurlanda 2015: 518.
[14] Vgl. Boehme-Neßler 2016: 423.
[15] Vgl. Sarunski 2016: 427.
[16] Vgl. Buchner 2018: 142.

Rechnen und *Secure Multi-Party Calculations*. Auch Mansmann weist jedoch ausdrücklich auf das konzeptionelle Problem hin, dass selbst bei globalen Statistiken durch ein »geschicktes Abfragen« möglicherweise wieder personenidentifizierende Informationen zu Tage treten können.[17]

2.3 *Pseudonymisierung*

Ist im konkreten Fall eine Anonymisierung nach dem Forschungszweck nicht möglich, gibt das nationale Datenschutzrecht in § 27 Abs. 3 S. 2 BDSG vor, dass personenbezogene Daten zumindest zu pseudonymisieren sind: Die Merkmale, mittels derer »Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren Person zugeordnet werden können«, sind nach der Vorschrift »gesondert zu speichern« und dürfen des Weiteren nach § 27 Abs. 3 S. 3 BDSG »mit den Einzelangaben nur zusammengeführt werden, soweit der Forschungs- oder Statistikzweck dies erfordert«. Ein Ausweichen auf die Pseudonymisierung von Daten anstelle einer Anonymisierung ist etwa dann erforderlich, wenn im Zuge von Langzeitstudien eine fortlaufende Zuordnung neuer Daten zu bereits vorhandenen Daten möglich sein muss.[18]

In der DS-GVO ist die Pseudonymisierung von personenbezogenen Daten in Art. 4 Nr. 5 DS-GVO definiert: Es handelt sich danach um eine Verarbeitung personenbezogener Daten in der Weise, dass diese Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können. Die Pseudonymisierung zielt damit zwar ebenso wie auch die Anonymisierung darauf ab, eine Personenbezogenheit von Daten auszuschließen oder zumindest zu erschweren. Gleichwohl sind jedoch pseudonymisierte Daten im Unterschied zu anonymisierten Daten weiterhin als personenbezogene Daten einzuordnen (EG 26 DS-GVO) – mit der Konsequenz, dass der datenschutzrechtliche Anwendungsbereich eröffnet ist und Datenverarbeitende in der Nutzung pseudonymisierter Daten gerade nicht »frei« sind, sondern den datenschutzrechtlichen Grenzen unterliegen. Die ganz grundlegende Differenzierung zwischen anonymisierten und pseudonymisierten Daten ist darauf zurückzuführen, dass nur bei der Anonymisierung

[17] Vgl. Abschnitt 8.4 (»Verfahren zum Datenschutz und zur gemeinsamen Datennutzung«) des Teils 1 (Konzeptionelle, organisatorische und technische Aspekte) des vorliegenden Sachstandsberichts.

[18] Vgl. Krawczak / Weichert 2017: 7.

von Daten die Zuordnung zu einer Person dauerhaft gegenüber jedem ausgeschlossen ist. Hingegen existiert bei der Pseudonymisierung eine Zuordnungsregel, mittels derer es möglich ist, eine Pseudonymisierung auch wieder rückgängig zu machen und den Personenbezug pseudonymisierter Daten wieder aufleben zu lassen.[19]

Dessen ungeachtet ist die Pseudonymisierung gerade im Forschungskontext ein zentrales Instrument, um Datenschutz und Forschungsfreiheit miteinander in Einklang zu bringen. Handelt es sich um besonders sensible Daten oder sind die Datenschutzrisiken aus anderen Gründen besonders ausgeprägt, kommt unter Umständen auch eine mehrfache Pseudonymisierung in Betracht.[20] Zu neueren technischen Methoden, die in eine ähnliche Richtung wie die klassische Pseudonymisierung gehen und etwa für die genomische Forschung diskutiert werden, zählen die *Statistical Disclosure Control* oder die Datensynthese.[21]

Im Ergebnis bleibt daher festzuhalten: Datenschutzrechtlich irrelevant ist Big Data (ebenso wie jede andere Form der Datenverarbeitung) nur dann, wenn es sich ausschließlich um die Nutzung von anonymen Informationen handelt, die keiner bestimmten oder bestimmbaren Person mehr zugeordnet werden können. In allen anderen Konstellationen, auch bei der Nutzung von pseudonymisierten Daten, muss sich Big Data an den datenschutzrechtlichen Vorgaben messen lassen.

3. Die spezifischen Herausforderungen von Big Data aus rechtlicher Perspektive

Fällt eine Big-Data-Anwendung in den Anwendungsbereich des Datenschutzrechts, stellt sich in einem nächsten Schritt die Frage, ob und inwieweit der Umstand, dass personenbezogene Daten nicht ›normal‹, sondern im Stile von Big Data verarbeitet werden, zu einer grundsätzlich anderen rechtlichen Einschätzung führt: Sind aus rechtlicher Perspektive die Spezifika von Big Data im Bereich der Medizin so besonders, dass sie es erfordern, eine Big-Data-Anwendung grundsätzlich nach anderen rechtlichen Maßstäben zu bewerten als eine ›normale‹ Datenverarbeitung? Zur Beantwortung dieser Frage bietet es sich an, auf die Merkmale von Big Data abzustellen, die gemeinhin, u. a. vom Deutschen Ethikrat, für eine Definition von Big Data herangezogen werden: *Volume*, *Variety*

[19] Vgl. Karg 2015: 521; Roßnagl / Scholz 2000: 724.
[20] Vgl. Herbst 2016: 374.
[21] Vgl. Hamacher et. al. 2020: 90.

und *Velocity*.[22] Big Data zeichnet sich nach diesem Verständnis also dadurch aus, dass riesige Datenmengen (Volume), die in unterschiedlichen Formaten vorliegen (Variety), in hoher Geschwindigkeit (Velocity) genutzt werden.

3.1 *Geschwindigkeit (Velocity)*

Geschwindigkeit ist sicherlich ein besonderes Charakteristikum vieler Big-Data-Anwendungen, insbesondere wenn Daten in Echtzeit verarbeitet werden, wie etwa auch bei den oben genannten Beispielen der Beobachtung bzw. Überwachung einer Virusausbreitung. Andererseits ist *Velocity* kein Alleinstellungsmerkmal von Big Data, vielmehr ist die immer schnellere Datenverarbeitung typische Begleiterscheinung jeder weiteren Technisierung und Digitalisierung des Gesundheitswesens.[23]

Umgekehrt gibt es gerade im medizinischen Kontext auch zahlreiche Big-Data-Anwendungen, die alles andere als ›schnell‹ (oder sogar in Echtzeit) Ergebnisse produzieren, etwa die Nutzung von Krankenkassendaten zu wissenschaftlichen Erkenntniszwecken. Eine solche Datenverarbeitung ist nicht ›schnell‹, sondern vielmehr langsam bzw. langfristig angelegt. Zu den Big-Data-Anwendungen ist sie gleichwohl zu zählen, weil schon aufgrund der schieren Menge an Daten eine Vielzahl von wertvollen Erkenntnissen für die Gesundheits- und Präventionsforschung gewonnen werden kann, etwa bei der Aufdeckung von Arzneimittelrisiken und einer Vielzahl anderer Fragestellungen der Versorgungsforschung. Ebenso sind zu den ›langsamen‹ bzw. langfristig angelegten Big-Data-Anwendungsszenarien etwa auch die Zusammenführung von Sozialdaten mit anderen Daten (Primärdaten aus Gesundheitsstudien, Daten aus der medizinischen Forschung), die Nutzung von Gesundheitsdaten auf der Grundlage von eHealth-Anwendungen (insbesondere elektronische Gesundheitskarte) sowie der Umgang mit Big-Data-Wissen im Zusammenhang mit Anreizsystemen und Beitragsdifferenzierungen in der Krankenversicherung zu zählen.

Daraus folgt: Das Merkmal *Velocity* mag zwar ein Standardkriterium für Big Data sein. Im medizinischen Kontext spielt es bei Big-Data-Anwendungen jedoch oftmals keine Rolle und ist schon deshalb auch aus rechtlicher Perspektive ohne Relevanz. Davon unabhängig stellt Geschwindigkeit aber ohnehin eine rechtliche Kategorie dar, die für die Ein-

[22] Vgl. Deutscher Ethikrat 2018: 54.
[23] Vgl. Buchner 2018: 133.

ordnung eines Datenverarbeitungsprozesses als zulässig oder unzulässig regelmäßig nicht ausschlaggebend ist.

3.2 *Datenmenge (Volume)*

Bereits der Name »Big Data« steht stellvertretend für das Spezifikum, dass Big-Data-Anwendungen typischerweise riesige Datenmengen (Volume) verarbeiten. So betrachtet ist dann aber die Datenverarbeitung im medizinischen Bereich schon seit jeher als Big-Data-Anwendung einzuordnen. Datenverarbeitung in der Medizin ist typischerweise und war schon immer durch eine Datenfülle geprägt; egal ob es um Behandlungsdaten, um Abrechnungs- und Verwaltungsdaten oder um Forschungsdaten geht. Seitdem es die ärztliche Dokumentationspflicht gibt, ist die Medizin eine datenintensive Disziplin, weil sich jeder Behandlungsprozess stets auch in einem entsprechenden Datenverarbeitungsprozess widerspiegelt.[24] Auch das Sozialversicherungssystem, insbesondere die Gesetzliche Krankenversicherung, produziert riesige Mengen an (Gesundheits-)Daten, wenn hier Daten zu den verschiedensten Zwecken erhoben und verarbeitet werden – zu Zwecken der Abrechnung, der Wirtschaftlichkeitskontrolle oder Qualitätssicherung oder auch für besondere Versorgungsformen.[25] Weitere Quellen des Datenreichtums in der Medizin sind Forschungsdatenbanken und epidemiologische Studien ebenso wie etwa auch die sogenannte Omics-Forschung oder bildgebende Verfahren.[26]

Rechtlich betrachtet steht die Datenfülle unter Big Data in einem offensichtlichen Widerspruch zu zentralen Grundprinzipien, wie sie die DS-GVO (ebenso aber auch schon das frühere Datenschutzrecht) prägen. An erster Stelle steht hier der Grundsatz der Datenminimierung gemäß Art. 5 Abs. 1 lit. c DS-GVO: Personenbezogene Daten müssen danach dem Zweck angemessen und erheblich sowie *auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt* sein. Konkretisiert wird das allgemeine Gebot der Datenminimierung insbesondere durch die bereits angesprochene Zielvorgabe der Anonymisierung und Pseudonymisierung: Lässt sich ein Verarbeitungszweck auch mit anonymisierten oder pseudonymisierten Daten erreichen, wäre es mit dem Grundsatz der Da-

[24] Vgl. Buchner 2018: 133.
[25] Vgl. Weichert 2014: 838.
[26] Vgl. Buchner 2018: 133.

tenminimierung nicht vereinbar, wenn ein Verantwortlicher auf diese Möglichkeit bei der Datenverarbeitung nicht zurückgreifen würde.

Ergänzt wird der Grundsatz der Datenminimierung durch den Grundsatz der sogenannten Speicherbegrenzung: Gemäß Art. 5 Abs. 1 lit. e DS-GVO müssen personenbezogene Daten in einer Form gespeichert werden, die die Identifizierung der betroffenen Personen nur so lange ermöglicht, wie es für die konkreten Verarbeitungszwecke erforderlich ist. Vorgegeben wird damit also auch eine *zeitliche* Grenze für die Datenverarbeitung: Die Speicherung personenbezogener Daten muss beendet werden, sobald deren Kenntnis für die mit der Verarbeitung verfolgten Zwecke nicht mehr erforderlich ist.

Big Data und Recht stehen damit in einem offensichtlichen Spannungsverhältnis: Datenfülle (Volume) einerseits und Datenminimierung sowie Speicherbegrenzung andererseits. Big Data ist auf möglichst viele Daten angewiesen, welche nicht sparsam und zeitlich begrenzt nutzbar sein sollen, sondern stattdessen möglichst umfangreich und zeitlich unbefristet zur Verfügung stehen sollen, um einen maximalen Erkenntnisgewinn zu gewährleisten.

3.3 *Datenvielfalt (Variety)*

Drittes Definitionsmerkmal von Big Data ist schließlich das Merkmal der *Variety* – verstanden als Vielfalt von Datentypen und Datenquellen. Big Data zeichnet sich in diesem Zusammenhang vor allem dadurch aus, dass es diese Datenvielfalt ›unter ein Dach‹ bringen will. Die Daten werden aus ihrem ursprünglichen Verwendungszusammenhang herauslöst und stattdessen frei von irgendeiner Ziel- oder Zwecksetzung genutzt: Verknüpfungsmuster sollen gesucht, Korrelationen in Datenbeständen gefunden und auf diese Weise Erkenntnisse gewonnen werden, deren Art und Umfang im Ausgangspunkt noch nicht absehbar waren.[27]

Entsprechend werden auch im Fall von medizinischen Daten die Verarbeitungsprozesse von ihren ursprünglichen Ziel- und Zwecksetzungen befreit: Patientendaten werden nicht mehr allein zum Zwecke der ärztlichen Behandlung verarbeitet, Sozialdaten nicht allein zum Zweck der Abrechnung im GKV-System und Forschungsdaten nicht allein zum Zweck der Überprüfung von Hypothesen usw. All diese Daten sollen vielmehr ohne konkrete Zweckbestimmung gesammelt und ausgewertet werden. Big Data steuert gerade nicht mehr zielgerichtet auf einen von

[27] Vgl. Ladeur 2016: 93.

vornherein anvisierten Erkenntnisgewinn hin, sondern speichert Daten auf Vorrat und wertet diese zweckfrei und ergebnisoffen aus.

Big Data gerät in dieser Ausprägung einer zweckfreien Vorratsdatenspeicherung in direkten Konflikt mit einem der ganz zentralen Grundsätze des Datenschutzrechts, dem Zweckbindungsgrundsatz. Danach dürfen Daten stets nur zu dem Zweck verarbeitet und genutzt werden, zu dem sie erhoben worden sind. Spätestens seit dem Volkszählungsurteil des Bundesverfassungsgerichts handelt es sich um einen der zentralen datenschutzrechtlichen Grundsätze, der so auch im europäischen Recht anerkannt ist.[28] Auch nach Art. 5 Abs. 1 lit. b DS-GVO müssen personenbezogene Daten »für festgelegte, eindeutige und rechtmäßige Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden«. Eine Datenverarbeitung auf Vorrat für unbestimmte Zwecke, also das, was Big Data in seinem Kern ausmacht, ist mit dem Zweckbindungsgrundsatz nicht vereinbar.

4. Die Antworten des Rechts auf Big Data

Die angesprochenen datenschutzrechtlichen Regelungsprinzipien – das Verbotsprinzip, die Grundsätze der Datenminimierung und Speicherbegrenzung, der Grundsatz der Zweckbindung – sind der Grund dafür, dass das geltende Datenschutzrecht vielfach als unvereinbar mit der ›Philosophie‹ von Big Data bezeichnet wird. Auch aus Sicht des Deutschen Ethikrats ist das aktuelle Datenschutzrecht eben wegen dieser Prinzipien auf das Phänomen Big Data nur »unzureichend eingestellt«.[29] In seiner Stellungnahme sieht der Ethikrat im geltenden Datenschutzrecht keine angemessenen Regelungsstrukturen, die den Herausforderungen von Big Data gerecht werden könnten.[30] Daher schlägt der Ethikrat eine Abkehr von wesentlichen datenschutzrechtlichen Grundsätzen hin zu der Ausrichtung der DS-GVO an einem Konzept einer »Datensouveränität« vor.

Zwingend ist diese kritische Sichtweise allerdings nicht – unter anderem deshalb nicht, weil sie dem forschungsfreundlichen und damit auch innovationsoffenen Charakter der DS-GVO nur unzureichend Rechnung trägt. In der DS-GVO finden sich, ebenso wie auch im nationalen

[28] Vgl. Bundesverfassungsgericht 1983: 1 ff.
[29] Deutscher Ethikrat 2018: 128.
[30] Vgl. ibid.: 22 f.

Recht, zahlreiche Regelungen, die eine Datenverarbeitung zu Forschungszwecken privilegieren, indem sie die angesprochenen datenschutzrechtlichen Regelungsprinzipien entsprechend einschränken und auf diese Weise auch in weitem Umfang Big-Data-Anwendungen ermöglichen.[31] Dies beginnt bei der Lockerung des Zweckbindungsgebots und setzt sich fort mit Ausnahmen vom grundsätzlichen Verbot einer Datenverarbeitung bis hin zu zahlreichen Ausnahmen bei den Betroffenenrechten und der Möglichkeit eines sogenannten *broad consent*. Um einen möglichen rechtlichen Weiterentwicklungsbedarf beurteilen zu können, bedarf es daher zunächst einmal einer differenzierteren Herausarbeitung und Bewertung der einschlägigen Regelungen unter dem Aspekt der Vereinbarkeit von Datenschutzrecht und Big Data.

4.1 *Verbotsprinzip – aber mit Erlaubnisvorbehalt*

Nach Art. 8 Abs. 2 Satz 1 der Charta der Grundrechte der Europäischen Union (GRCh) bedürfen Daten verarbeitende Stellen für jede Form einer Verarbeitung von personenbezogenen Daten stets einer Legitimation, entweder in Form eines gesetzlichen Erlaubnistatbestands für die Datenverarbeitung oder in Form einer Einwilligung des bzw. der Betroffenen selbst. Die datenschutzrechtliche Grundregel, das Verbotsprinzip mit Erlaubnisvorbehalt, ist damit also bereits im Verfassungsrecht verankert: Personenbezogene Daten dürfen grundsätzlich zunächst einmal nicht verarbeitet werden, es sei denn, dies ist ausnahmsweise gesetzlich erlaubt oder durch die betroffene Person selbst mittels Einwilligung legitimiert.

4.1.1 Differenzierter Ansatz

Damit ist jedoch keineswegs die Konsequenz verbunden, dass das Datenschutzrecht in seiner jetzigen Konzeption eine unüberwindbare Hürde für die Verarbeitung personenbezogener Daten darstellen würde; auch nicht für umfangreiche Datenverarbeitungsprozesse im Stile von Big Data. Vielmehr verfolgt das Datenschutzrecht seit jeher einen risikobasierten Ansatz und nimmt ebenso auch die Ziele und Interessen, die mit einer Verarbeitung personenbezogener Daten verfolgt werden, in den Blick.[32] Auch den Grundrechtspositionen der datenverarbeitenden Stel-

[31] Vgl. Buchner 2018: 131.
[32] Vgl. Tinnefeld et. al. 2020: 241.

len wird im datenschutzrechtlichen Regelungsgefüge eine hohe Bedeutung beigemessen – so insbesondere im Rahmen der gesetzlichen Erlaubnistatbestände oder auch durch die Möglichkeit, über den Weg des privatautonomen Interessenausgleichs (Einwilligung, Vertrag) die Datenverarbeitung zu legitimieren. Ebenso differenziert die DS-GVO nach mehr oder weniger schutzwürdigen Daten und berücksichtigt im Rahmen der gesetzlichen Erlaubnistatbestände auch das unterschiedliche Gefährdungspotenzial von Datenverarbeitungsprozessen, etwa über die schutzwürdigen Interessen des Betroffenen (vgl. Art. 6 Abs. 1 lit. f). Aus der Grundidee des Datenschutzrechts, eine Verarbeitung von Daten auf das notwendige Maß zu beschränken, folgt nicht, dass nicht auch Daten im großen Umfang unter Big Data verarbeitet werden dürften – vorausgesetzt es gibt hierfür einen plausiblen Grund und die widerstreitenden Interessen von Betroffenen und Verantwortlichen werden gegeneinander abgewogen und miteinander in Einklang gebracht.[33]

4.1.2 Gesetzliche Erlaubnistatbestände (insbesondere Forschung)

Auch für die Verarbeitung von besonders schutzwürdigen Daten wie Gesundheitsdaten sieht Art. 9 Abs. 2 DS-GVO eine Reihe von Ausnahmen von dem in Art. 9 Abs. 1 DS-GVO normierten Verbotsprinzip vor. Erlaubnistatbestände finden sich unter anderem für den Gesundheitssektor in Art. 9 Abs. 2 lit. h und lit. i DS-GVO. Von besonderer Relevanz für Big-Data-Anwendungen ist darüber hinaus die Norm des Art. 9 Abs. 2 lit. j DS-GVO. Danach können im mitgliedstaatlichen Recht Ausnahmen vom grundsätzlichen Verbot einer Datenverarbeitung vorgesehen werden, wenn die 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«. Auf diese Privilegierung einer Datenverarbeitung zu Forschungszwecken kann sich auch Big Data stützen – vorausgesetzt, die konkrete Big-Data-Anwendung ist als wissenschaftliche Forschung einzuordnen und die mit der Forschung verfolgten Interessen überwiegen die Datenschutzinteressen der betroffenen Personen erheblich. Bei Big-Data-Forschung, die ausschließlich auf eine

[33] Vgl. Kühling 2020: 186.

Verbesserung von Prävention, Diagnose und Therapie in der medizinischen Praxis abzielt, ist davon regelmäßig auszugehen.[34]

Zu den Vorschriften im nationalen bereichsspezifischen Recht, die die Datenverarbeitung zu Forschungszwecken einem großzügigeren datenschutzrechtlichen Rahmen unterstellen wollen, zählt auch § 75 Sozialgesetzbuch X (SGB X). Die Vorschrift erlaubt eine Übermittlung von Sozialdaten, soweit dies für ein bestimmtes Vorhaben der wissenschaftlichen Forschung im Sozialleistungsbereich erforderlich ist und die Interessenabwägung zugunsten der Forschung ausfällt. Den besonderen Bedürfnissen der Forschung wird in dieser Norm zudem dadurch Rechnung getragen, dass nach § 75 Abs. 2 SGB X auch Forschungsfragen, die sich erst später auftun, ebenfalls auf derselben Datengrundlage bearbeitet werden dürfen, sofern ein »inhaltlicher Zusammenhang« zwischen alter und neuer Forschungsfrage zu bejahen ist. Auch die Forschungsklauseln im Landeskrankenhausrecht sind Beispiele dafür, wie das Datenschutzrecht eine Verarbeitung von personenbezogenen Daten für wissenschaftliche Forschungsvorhaben unter bestimmten Voraussetzungen privilegiert (s. z. B. § 7 BremKHDSG oder § 25 LKG-Bln).

4.2 *Einwilligung*

Zentraler Legitimationstatbestand für eine Verarbeitung personenbezogener Daten – egal ob ›normaler‹ Daten oder besonders schutzwürdiger Gesundheitsdaten – ist darüber hinaus stets die Einwilligung der betroffenen Person selbst. Auch mit einer Datenverarbeitung à la Big Data kann sich der Einzelne – ganz im Sinne informationeller Selbstbestimmung – einverstanden erklären und damit eine Legitimationsgrundlage für Big-Data-Anwendungen schaffen.[35]

4.2.1 Wirksamkeitsvoraussetzungen

Gerade bei Big-Data-Anwendungen ist dann aber ein besonderes Augenmerk darauf zu legen, dass alle Voraussetzungen für eine *wirksame* Einwilligung erfüllt sind. Wirklich »selbstbestimmt« (und damit wirksam) ist eine Einwilligung in die Verarbeitung personenbezogener Daten nur

[34] Vgl. Buchner 2018: 140.

[35] Vgl. zu den ethischen Aspekten der Einwilligung Abschnitt 3 (»Autonomie und informationelle Selbstbestimmung in der Praxis«) des Teils 3 (Ethische Aspekte) des vorliegenden Sachstandsberichts.

dann, wenn die betroffene Person diese Einwilligung bewusst und bestimmt sowie freiwillig und informiert erteilt hat. Schon die Gewährleistung der Freiwilligkeit einer Einwilligung kann in Big-Data-Szenarien eine besondere Herausforderung sein, etwa wenn es sich um die Verarbeitung von Sozialdaten im System der Gesetzlichen Krankenversicherung handelt oder wenn im Zuge medizinischer Behandlungen eine Einwilligung bei besonders vulnerablen Personengruppen (Patienten, Studienteilnehmer) eingeholt wird. Ebenso ist auch auf die Informiertheit und Zweckbestimmtheit einer Einwilligung gerade bei Big-Data-Anwendungen besonders zu achten, weil sich diese Anforderungen mit der Zweckfreiheit und Ergebnisoffenheit eines Datenverarbeitungsprozesses im Stile von Big Data grundsätzlich nur schwer vereinbaren lassen.

4.2.2 Insbesondere: Bestimmtheit der Einwilligung

Nach Art. 6 Abs. 1 lit. a DS-GVO muss jede Einwilligung in eine Datenverarbeitung *»für einen oder mehrere festgelegte Zwecke«* erteilt werden; aus Art. 5 Abs. 1 lit. b DS-GVO folgt zudem, dass diese Zweckfestlegung *»eindeutig«* sein muss. Egal wie zweckoffen und ergebnisfrei daher eine Big-Data-Anwendung auch angelegt sein mag: Sollen die damit einhergehenden Datenverarbeitungsprozesse durch eine Einwilligung legitimiert werden, kann sich diese nicht auf allgemeine Formulierungen und Zweckbestimmungen beschränken, da diese dem Bestimmtheitsgebot nicht hinreichend Rechnung tragen. Für Betroffene muss auch bzw. gerade bei Big-Data-Anwendungen klar sein, welche Daten in welchem Umfang und zu welchem Zweck verarbeitet werden und an welche dritten Stellen diese Daten gegebenenfalls übermittelt werden.

Andererseits ist es jedoch mit dem Bestimmtheitserfordernis durchaus auch vereinbar, dass sich der Einzelne im Kontext von Big Data mittels Einwilligung auch mit einer sehr weitreichenden Nutzung seiner Daten einverstanden erklärt – vorausgesetzt, aus der Einwilligungserklärung sind die erfassten Daten klar ersichtlich.[36] Zu berücksichtigen ist auch, dass die Zweckbestimmung einer Einwilligung einer Datenverarbeitung zu anderen Zwecken dann nicht entgegensteht, wenn diese zu wissenschaftlichen Forschungszwecken erfolgt (s. dazu näher Abschnitt 4.3.2 »Forschungsprivilegierung«).

[36] Vgl. Tinnefeld et al. 2020: 429 f.

4.2.3 *Broad Consent*

Eine Lockerung erfährt das Erfordernis einer zweckbestimmten Einwilligung für den Fall einer Datenverarbeitung zu wissenschaftlichen Forschungszwecken zudem durch die Rechtsfigur des *broad consent*. Die Einwilligung in eine Datenverarbeitung zu Forschungszwecken kann danach auch ›breiter‹, d. h. unbestimmter ausfallen und sich allgemein auf bestimmte Bereiche wissenschaftlicher Forschung erstrecken.[37] Die DS-GVO nimmt diesen Ansatz in ihrem Erwägungsgrund 33 auf. Ausdrücklich ist dort die Rede davon, dass sich der Zweck einer Datenverarbeitung für Zwecke der wissenschaftlichen Forschung zum Zeitpunkt der Datenerhebung oftmals noch nicht vollständig angeben lasse. Daher sollte es möglich sein, dass Betroffene ihre Einwilligung auch »für bestimmte Bereiche wissenschaftlicher Forschung« geben, wenn dies »unter Einhaltung der anerkannten ethischen Standards der wissenschaftlichen Forschung geschieht«. Einschränkend heißt es des Weiteren dann noch, dass dem Einzelnen die Möglichkeit eröffnet werden muss, seine Einwilligung »nur für bestimme Forschungsbereiche oder Teile von Forschungsprojekten in dem vom verfolgten Zweck zugelassenen Maße zu erteilen«. Für die medizinische Forschung ermöglicht die Rechtsfigur des *broad consent* jedenfalls eine deutliche Flexibilisierung des Umfangs mit personenbezogenen Daten.[38]

4.2.4 Datenspende

Im Sinne einer inhaltlichen Weiterentwicklung und Überarbeitung des geltenden Rechts hat der Deutsche Ethikrat u. a. die Idee einer »Datenspende« als noch flexibleres Einwilligungsmodell in die Diskussion eingebracht, um Datenschutz und Big Data miteinander in Einklang zu bringen. Das traditionelle Einwilligungsmodell soll auf diese Weise nicht nur prozedural erweitert, sondern auch bereichsbezogen geöffnet werden. Insbesondere soll es mittels einer Datenspende ermöglicht werden, umfassend einer Datennutzung ohne enge Zweckbindung zugunsten der klinischen und medizinbezogenen Grundlagenforschung zuzustimmen.[39]

Die Datenspende geht also nochmals über die Rechtsfigur des *broad consent* hinaus, indem auf den Bestimmtheitsgrundsatz bei der Einwil-

---

[37] Vgl. Herbst 2016: 373. Vgl. zu ethischen Aspekten des *broad consent* Abschnitt 3.1 (»Autonomiezentrierte Ansätze zur Weiterentwicklung des Konzeptes der informierten Einwilligung«) des Teils 3 (Ethische Aspekte) des vorliegenden Sachstandsberichts.

[38] Vgl. Kühling 2020: 188.

[39] Vgl. Deutscher Ethikrat 2018: 266 f.

ligung gänzlich verzichtet wird. Damit stellt sich allerdings die Frage, ob ein solcher einwilligungsbasierter Lösungsansatz überhaupt so ausgestaltet werden kann, dass damit nicht eine endgültige Aufgabe informationeller Selbstbestimmung einhergeht. Immerhin dürfen auch nach Art. 8 Abs. 2 EU-Grundrechtecharta Daten nur »nach Treu und Glauben für festgelegte Zwecke« verarbeitet werden. Der Deutsche Ethikrat bleibt insoweit die Antwort schuldig, wie sich eine Datenspende mit diesen grundrechtlichen Wertungen vereinbaren lässt.

Darüber hinaus stellt sich auch die Frage, ob eine Datenspende nicht so ausgestaltet werden muss, dass der betroffenen Person stets die Möglichkeit erhalten bleibt, einmal getroffene Entscheidungen auch wieder rückgängig zu machen. Grundsätzlich gilt, dass die Widerrufsmöglichkeit als Gegenpart zur Einwilligung ebenso wie die Einwilligung selbst »grundrechtliche Realisierung informationeller Selbstbestimmung«[40] ist und daher nach geltenden verfassungsrechtlichen Maßstäben auch kein Verzicht erklärt werden kann. Die freie Widerrufbarkeit der Einwilligung soll nicht zuletzt auch den Einzelnen vor sich selbst schützen. Auch wenn der Einzelne auf sein Grundrecht auf informationelle Selbstbestimmung verzichtet, muss ihm die Befugnis verbleiben, diesen Grundrechtsverzicht wieder rückgängig zu machen. Die einzelne Person soll gerade nicht zu einem bloßen ›Objekt‹ der Datenverarbeitung werden, weil es ihr verwehrt ist, ihre informationelle Selbstaufgabe wieder zu revidieren und in ihre Rolle als selbstbestimmtes Subjekt zurückzukehren.[41] Wenn damit aber verfassungsrechtlich vorgegeben ist, dass jede Form einer Einwilligung in die Datenverarbeitung stets wieder rückgängig gemacht werden kann, läuft damit die eigentliche Idee einer Daten*spende* ins Leere.

4.3 *Zweckbindung – Zweckvereinbarkeit*

Ebenso wie für das Verbotsprinzip gilt auch für den Zweckbindungsgrundsatz, dass dieser nicht zu einem unauflösbaren Konflikt zwischen Datenschutzrecht und Big Data führt. Der Zweckbindungsgrundsatz geht nicht so weit, dass eine Datenverarbeitung zu einem anderen als dem ursprünglich verfolgten Zweck überhaupt nicht möglich ist. Bereits verfügbare Daten dürfen nach Art. 6 Abs. 4 DS-GVO durchaus auch zu einem anderen als dem ursprünglich verfolgten Erhebungszweck genutzt werden, vorausgesetzt die weitere Nutzung kann ebenfalls auf einen ent-

[40] Sydow / Ingold 2018: Art. 7 Rn. 46.
[41] Vgl. Buchner 2006: 232 ff.

sprechenden Erlaubnistatbestand gestützt werden und der Zweck der weiteren Nutzung ist nicht unvereinbar mit dem ursprünglichen Erhebungszweck. Mit guten Gründen kann man daher auch einwenden, dass es sich bei dem bis dato üblicherweise so bezeichneten Zweckbindungsgrundsatz jedenfalls seit Geltung der DS-GVO der Sache nach eher um einen Grundsatz der »Zweckvereinbarkeit« handelt.[42]

4.3.1 Kriterien einer Zweckvereinbarkeit

Für die Beurteilung, ob der Zweck einer Weiterverarbeitung von Daten mit dem ursprünglich verfolgten Erhebungszweck vereinbar ist oder nicht, kommt es auf die in Art. 6 Abs. 4 DS-GVO normierten Beurteilungskriterien an, die zwar nicht abschließend sind, wohl aber das Grundgerüst für die Beurteilung einer Zweckvereinbarkeit bilden. In Art. 6 Abs. 4 lit. a – lit. e finden sich fünf Kriterien, die im Fall von Big-Data-Anwendungen oftmals den Ausschlag eher gegen als für eine Vereinbarkeit geben werden.[43] Dies gilt schon für die ersten beiden Kriterien, die »Verbindung« zwischen dem ursprünglichen und dem neuen Zweck (lit. a) sowie den »Zusammenhang« der Datenerhebung (lit. b): Je weiter der Zweck der ursprünglichen Verarbeitung und der der Weiterverarbeitung auseinanderliegen und je unvorhersehbarer für die betroffene Person eine weitere Verarbeitung ist, desto mehr spricht dafür, dass die weitere Verarbeitung der Daten mit dem ursprünglichen Zweck unvereinbar ist. Ein typisches Wesenselement von Big Data ist es gerade, auch bis dato nicht vorhersehbare Korrelationen aufzudecken.

Auch das Kriterium der Art der personenbezogenen Daten (lit. c) erschwert für Big-Data-Anwendungen im medizinischen Kontext die Annahme einer Zweckvereinbarkeit, da strenge Maßstäbe gerade dann anzulegen sind, wenn es sich um besonders schutzwürdige Daten wie Gesundheitsdaten handelt. Und schließlich können auch die »möglichen Folgen« einer beabsichtigten Weiterverarbeitung für Betroffene (lit. d) im Fall von Big-Data-Anwendungen gegen eine Vereinbarkeit sprechen, wenn diese darauf abzielen, einzelne Personen in bestimmte Kategorien einzuordnen, um daran dann möglicherweise auch noch für die Betroffenen nachteilige Entscheidungen anzuknüpfen.[44]

Insgesamt sind also die Kriterien, an denen sich auch Big-Data-Anwendungen hinsichtlich der Zweckbindung bzw. Zweckvereinbarkeit

[42] Vgl. Kühling / Buchner / Herbst 2020: Art. 5 Rn. 24 ff.
[43] Vgl. Buchner 2018: 136.
[44] Vgl. ibid.

einer Datenverarbeitung messen lassen müssen, relativ streng. Dies mag für den Ethikrat dann auch (ein) Beweggrund gewesen sein, in Abkehr vom Zweckbindungsgrundsatz eher auf andere »technisch-organisatorische sowie materiell- und verfahrensrechtliche Sicherungen« zu setzen, die den »Mangel an Konkretheit von gesundheitsrelevanten Big-Data-Anwendungen« kompensieren sollen.[45] Viel gewonnen ist mit diesem Vorschlag allerdings nicht, da nicht so recht ersichtlich ist, wie sich der Ethikrat konkret diese »Kompensation« vorstellt – immerhin zählt der Zweckbindungsgrundsatz zu den ganz zentralen datenschutzrechtlichen Sicherungsinstrumentarien, die einen effektiven Schutz informationeller Selbstbestimmung gewährleisten sollen.[46]

4.3.2 Forschungsprivilegierung

Unproblematisch bzw. irrelevant ist der Zweckbindungsgrundsatz demgegenüber von vornherein dann, wenn personenbezogene Daten zu wissenschaftlichen Forschungszwecken verarbeitet werden. Insoweit setzt sich das datenschutzrechtliche Muster fort, dass die Datenverarbeitung zu Forschungszwecken eine Privilegierung erfährt, weil die datenschutzrechtlichen Anforderungen deutlich niedriger angesetzt werden. 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 Voraussetzungen für die Annahme einer Zweckvereinbarkeit, wie sie in Art. 6 Abs. 4 DS-GVO normiert sind und die sonst bei Big Data oftmals nur schwer zu erfüllen sind, spielen damit bei einer Datenverarbeitung zu wissenschaftlichen Forschungszwecken von vornherein keine Rolle.

4.4 *Transparenz und Richtigkeit*

Gemäß Art. 5 Abs. 1 lit. a DS-GVO müssen personenbezogene Daten »auf rechtmäßige Weise, nach Treu und Glauben *und in einer für die betroffene Person nachvollziehbaren Weise*« verarbeitet werden. Letzteres Gebot der Transparenz schließt nicht nur eine heimliche Datenverarbeitung aus, sondern verlangt vor allem auch eine umfassende Information der betroffenen Person über die Verarbeitung der sie betreffenden Daten. Gemäß

[45] Deutscher Ethikrat 2018: 26.
[46] Vgl. Kühling 2020, 185 f.

Art. 5 Abs. 1 lit. d DS-GVO müssen personenbezogene Daten zudem »sachlich richtig und erforderlichenfalls auf dem neuesten Stand sein«. Der Verantwortliche muss nach Art. 5 Abs. 1 lit. d DS-GVO »alle angemessenen Maßnahmen« ergreifen, damit unrichtige Daten unverzüglich gelöscht oder berichtigt werden.

4.4.1 Gesundheitsalgorithmen als Herausforderung für das Recht

Beide Grundsätze, sowohl die Transparenz der Datenverarbeitung als auch die Richtigkeit der Datenbestände, stellen unter Big Data eine besondere Herausforderung dar. Je umfangreicher Datenbestände sind, desto intransparenter und fehleranfälliger ist auch die Verarbeitung dieser Daten. Für den Einzelnen ist es im Fall von Big-Data-Anwendungen regelmäßig schwer bis unmöglich, die Datenverarbeitungsprozesse als solche noch nachzuvollziehen und/oder die Richtigkeit einer Datenverarbeitung zu kontrollieren. »Transparent« sind für diesen nur noch die konkreten Ergebnisse einer Big-Data-Anwendung – die im medizinischen Kontext dann aber auch so aussehen können, dass infolge einer Big-Data-Anwendung etwa ein potenzieller Versicherungsnehmer oder eine potenzielle Versicherungsnehmerin als ›zu krank‹ oder ein Patient bzw. eine Patientin im Krankenhaus als vergleichsweise weniger oder überhaupt nicht mehr ›behandlungswürdig‹ eingestuft wird. Schon heute werden sogenannte Gesundheitsalgorithmen eingesetzt, die über die Behandlung von Patientinnen und Patienten im Gesundheitswesen »entscheiden«.[47] Für die betroffenen Personen sind diese Entscheidungen regelmäßig intransparent und auch nicht daraufhin kontrollierbar, ob die der Big-Data-Anwendung zugrundliegenden Daten »sachlich richtig« und »auf dem neuesten Stand« im Sinne des Art. 5 Abs. 1 lit. d DS-GVO sind.

4.4.2 Rechtliche Lösungsansätze

Die Lösung für solcherlei Probleme sucht das Datenschutzrecht seit jeher und unter Geltung der DS-GVO noch einmal verstärkt in der Information der betroffenen Personen. So hat der europäische Gesetzgeber in den Art. 13 und 14 DS-GVO umfangreiche Informationspflichten verankert, die die Transparenz einer Datenverarbeitung so weit wie möglich vorantreiben sollen. Der Erfolg dieser Herangehensweise in der Praxis ist aller-

[47] Vgl. Ärzteblatt 2019.

dings zweifelhaft. Zum einen bleibt die tatsächliche Umsetzung all der Transparenzpflichten hinter den rechtlichen Vorgaben zurück.[48] Zum anderen ist aber auch eine rechtstreue Umsetzung sämtlicher Informationspflichten nicht zwangsläufig transparenzfördernd, wenn die Vielzahl an mitunter viel zu detaillierten und im Einzelnen kaum mehr nachvollziehbaren ›Informationen‹ im Ergebnis zu einer Informationsüberflutung bei den betroffenen Personen führt (Information overload).

Entscheidend dafür, ob gerade in Zeiten von Big Data eine echte Transparenz für Betroffene überhaupt realistisch ist, wird darüber hinaus sein, ob und inwieweit die datenschutzrechtlichen Informations- und Auskunftspflichten auch die Nachvollziehbarkeit von *algorithmenbasierten Entscheidungen* herstellen können. Der europäische Gesetzgeber hat diese Herausforderung durchaus gesehen, indem er die Informations- und Auskunftspflichten in den Art. 13 Abs. 2 lit. f, Art. 14 Abs. 2 lit. g sowie Art. 15 Abs. 1 lit. h DS-GVO ausdrücklich auch auf die »involvierte Logik« einer automatisierten Entscheidung erstreckt hat. Wie transparent Big Data in der Zukunft sein wird, hängt maßgeblich davon ab, wie diese rechtliche Vorgabe in der Praxis ausgelegt wird. Unter dem alten Recht war die Rechtsprechung hierzulande noch wenig transparenzfreundlich und stattdessen eher auf den Schutz von sogenannten Geschäftsgeheimnissen bedacht.[49] So zählte der Bundesgerichtshof in seiner Entscheidung zur Offenlegung der Scoreformel von Auskunfteien (Schufa-Score) die meisten Rechengrößen zu den als Geschäftsgeheimnis geschützten Inhalten, über die der einzelne Betroffene nicht in Kenntnis gesetzt werden muss.

Unter Geltung der DS-GVO werden sich solcherlei Einschränkungen der Informationspflichten allerdings nicht mehr aufrechterhalten lassen.[50] Kernanliegen der europäischen Reform des Datenschutzrechts war es gerade, die Rechte des Einzelnen auf Information und Auskunft nachhaltig zu stärken.[51] Art. 13 Abs. 2 lit. f, Art. 14 Abs. 2 lit. g und Art. 15 Abs. 1 lit. h DS-GVO fordern ausdrücklich »aussagekräftige Informationen« über die involvierte Logik einer Entscheidung. Damit der Einzelne aber Entscheidungen, die ihn betreffen, zumindest in Grundzügen nachvollziehen kann, bedarf er hierfür auch Informationen, wie algorithmenbasierte Entscheidungen zustande kommen, insbesondere welche Daten

[48] Vgl. Wiebe / Helmschrot 2019 (Beispiel Online-Dienste).
[49] Vgl. Bundesgerichtshof 2014: 343.
[50] Vgl. Kühling / Buchner / ders. 2020: Art. 22 DS-GVO, Rn. 35 f.
[51] Vgl. Albrecht / Jotzo 2017: 38 f., 51.

mit welcher Gewichtung in eine Berechnung einfließen oder welche Vergleichsgruppen gebildet werden.[52]

4.5 *Fazit: Die Vereinbarkeit von Datenschutzrecht und Big Data*

Grundsätzlich kann bereits mit dem geltenden Recht den Herausforderungen von Big Data in vielerlei Hinsicht Rechnung getragen werden. Zum einen kann (bzw. muss) das Recht so ausgelegt werden, dass es die besonderen Risiken von Big Data hinreichend berücksichtigt – so etwa mit Blick auf die Transparenz und Richtigkeit algorithmenbasierter Entscheidungen. Zum anderen ist das geltende Recht aber auch flexibel genug, um eine ›Big-Data-freundliche‹ Auslegung zu ermöglichen – dies gilt vor allem für all die Big-Data-Anwendungen im Forschungskontext. Zu kurz greift daher auch die Kritik des Deutschen Ethikrats an der Unvereinbarkeit »des überkommenen Datenschutzrechts mit den Besonderheiten von Big-Data-Anwendungen«[53]. Es bedarf keineswegs einer Abkehr von geltenden datenschutzrechtlichen Grundsätzen, sondern lediglich einer Auslegung, die sowohl den Chancen als auch den Risiken von Big Data adäquat Rechnung trägt.

5. Ausblick: Forschungsprivilegierung

Gerade weil das geltende Recht für den Bereich der Forschungsdatenverarbeitung den größten Spielraum für eine Vereinbarkeit von Datenschutz und Big Data bereithält, ist damit zu rechnen, dass sich künftig mehr und mehr Datenverarbeiter bei Big Data auf irgendwelche »Forschungszwecke« berufen werden, um sich auf diese Weise von den lästigen Fesseln des Datenschutzrechts zu befreien. Zwar mag auch die Big-Data-›Forschung‹ großer IT-Unternehmen wie Google, Amazon oder Facebook mitunter auf einen wissenschaftlichen Erkenntnisgewinn abzielen. Damit einher geht aber regelmäßig auch eine kommerzielle Nutzung personenbezogener Daten. Letzteres muss eine datenschutzrechtliche Privilegierung als ›Forschung‹ jedenfalls dann ausschließen, wenn nicht das Ziel einer transparenten Wissensgenerierung für die Allgemeinheit im Vor-

[52] Vgl. Kühling / Buchner / ders. 2020: Art. 22 DS-GVO, Rn. 35a; Simitis / Hornung / Spiecker gen. Döhmann / Dix 2019: Art. 15, Rn. 25.
[53] Deutscher Ethikrat 2018: 129, 156.

dergrund steht, sondern stattdessen die kommerzielle Verwertung eines neuen Erkenntnisgewinns.

In Betracht kommt die Privilegierung einer Forschungsdatenverarbeitung im unternehmerischen Kontext stets nur dann, wenn auch diese die Kernmerkmale wissenschaftlicher Forschung erfüllt: die Transparenz des Forschungsprozesses und der Ergebnisse, die Unabhängigkeit und Selbstständigkeit der Forschenden sowie das Ziel eines Erkenntnisgewinns im Allgemeininteresse frei von sachfremden Erwägungen.[54] Für das Recht wird eine ganz zentrale Herausforderung künftig gerade auch darin liegen, sicherzustellen, dass wissbegierige Unternehmen nicht beliebig unter dem Deckmantel der wissenschaftlichen Forschung bei jedem Einsatz von Big Data die Privilegien einer Forschungsdatenverarbeitung für sich in Anspruch nehmen.

Literaturverzeichnis

Ärzteblatt (2019): Gesundheits-Algorithmus unter Rassismusverdacht. URL https://www.aerzteblatt.de/blog/106972/Gesundheits-Algorithmus-unter-Rassismusverdacht [17. März 2020].

Albrecht, J. P. / Jotzo, F. (2017): Das neue Datenschutzrecht der EU. Baden-Baden: Nomos.

Boehme-Neßler, V. (2016): Das Ende der Anonymität. In: Datenschutz und Datensicherheit 40, 419–423.

Böge, F. (2020): Wenn der Smartphone-Nutzer Rot sieht. In: Frankfurter Allgemeine Zeitung (5. März 2020), 7.

Buchner, B. (2018): Big Data und Datenschutz im Gesundheitswesen. In: Zeitschrift für medizinische Ethik 2, 131–146.

Ders. (2006): Informationelle Selbstbestimmung im Privatrecht. Tübingen: Mohr Siebeck.

Bundesgerichtshof (2014): Urteil vom 28. Januar 2014. VI ZR 156/13. In: Datenschutz und Datensicherheit 38 (5), 341–343.

Bundesverfassungsgericht (1983): Urteil vom 15. Dezember 1983. 1 BvR 209/83 u. a. In: Verein der Richter des Bundesverfassungsgerichts (Hg.): Entscheidungen des Bundesverfassungsgerichts. Tübingen: J. C. B. Mohr, 1–71.

Deutscher Ethikrat (Hg.) (2018): Big Data und Gesundheit – Datensouveränitaät als informationelle Freiheitsgestaltung. Stellungnahme. Berlin.

Hahn, E. (2019): Das »Recht auf Nichtwissen« des Patienten bei algorithmengesteuerter Auswertung von Big Data. In: Medizinrecht 37, 197–202.

Hamacher, K. / Katzenbeisser, S. / Kussel, T. / Stammler, S. (2020): Genomische Daten und der Datenschutz. In: Datenschutz und Datensicherheit 2, 87–93.

[54] Vgl. Kühling / Buchner / ders. u. Tinnefeld 2020: Art. 89 DS-GVO, Rn. 13.

Heller, P. (2020): Big Data soll Ausbreitung stoppen. In: Frankfurter Allgemeine Zeitung (5. März 2020), 72.

Herbst, T. (2016): Rechtliche und ethische Probleme des Umgangs mit Proben und Daten bei großen Biobanken. In: Datenschutz und Datensicherheit 40, 371–375.

Karg, M. (2015): Anonymität, Pseudonyme und Personenbezug revisited? In: Datenschutz und Datensicherheit 39 (8), 538–543.

Katzenmeier, C. (2019): Big Data, E-Health, M-Health, KI und Robotik in der Medizin. In: Medizinrecht 37, 259–271.

Krawczak, M. / Weichert, T. (2017): Vorschlag einer modernen Dateninfrastruktur für die medizinische Forschung in Deutschland (Version 1.9). URL https://www.uni-kiel.de/medinfo/documents/TWMK%20Vorschlag%20DInfMedForsch%20v1.9%20170927.pdf [17. März 2020].

Kühling, J. (2020): Gesundheitsdatenschutzrecht im Zeitalter von »Big Data«. In: Datenschutz und Datensicherheit 44, 182–188.

Kühling, J. / Buchner, B. (2020 i. E.): Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG. München: C. H. Beck [zitiert als Kühling / Buchner / Bearbeiter 2020].

Ladeur, K.-H. (2016): Wissenserzeugung im Sozialrecht und der Aufstieg von »Big Data«. In: Buchner, B. / Ladeur, K.-H. (Hg.): Wissensgenerierung und -verarbeitung im Gesundheits- und Sozialrecht. Tübingen: J. C. B. Mohr, 89–105.

Mayer-Schönberger, V. / Cukier, K. (2013): Big Data: Die Revolution, die unser Leben verändern wird. München: Redline Verlag.

Roßnagel, A. / Scholz, P. (2000): Datenschutz durch Anonymität und Pseudonymität – Rechtsfolgen der Verwendung anonymer und pseudonymer Daten. In: Multimedia und Recht 12, 721–731.

Sarunski, M. (2016): Big Data – Ende der Anonymität? In: Datenschutz und Datensicherheit 40, 424–427.

Schneider, J. (2017): Datenschutz: Nach der EU-Datenschutz-Grundverordnung. München: C. H. Beck.

Simitis, S. / Hornung, G./ Spiecker gen. Döhmann, I. (2019): Datenschutzrecht. Baden-Baden: Nomos [zitiert als Simitis / Hornung / Spiecker gen. Döhmann / Bearbeiter 2019].

Sydow, G. (2018): Europäische Datenschutzgrundverordnung. Baden-Baden: Nomos [zitiert als Sydow / Bearbeiter 2018].

Tinnefeld, M.-T. / Buchner, B. / Petri, T. / Hof, H. (2020): Einführung in das Datenschutzrecht. Berlin: De Gruyter.

Watteler, O. / Kinder- Kurlanda, K. E. (2015): Anonymisierung und sicherer Umgang mit Forschungsdaten in der empirischen Sozialforschung. In: Datenschutz und Datensicherheit 39, 515–519.

Weichert, T. (2020): Die Forschungsprivilegierung in der DS-GVO. In: Zeitschrift für Datenschutz 1, 18–24.

Ders. (2014): Big Data, Gesundheit und der Datenschutz. In: Datenschutz und Datensicherheit 38, 831–838.

Wiebe, A. / Helmschrot, C. (2019): Untersuchung der Umsetzung der DS-GVO durch Online-Dienste. URL https://www.bmjv.de/SharedDocs/Downloads/DE/News/Artikel/112919_DSGVO_Studie.pdf?__blob=publicationFile&v=2 [17. März 2020].