

Angela Borgwardt

Mehr Daten nutzbar machen! Anforderungen an ein Forschungsdatengesetz

Eine Stunde für die Wissenschaft
Paper No. 11

AUF EINEN BLICK

Im Zuge der digitalen Transformation hat die Bedeutung von Daten für die Wissenschaft immer mehr zugenommen. Datenintensive Forschung leistet einen zentralen Beitrag zur Lösung gesellschaftlicher Herausforderungen und unterstützt gesellschaftliche Weiterentwicklung und Innovation. Gegenwärtig bestehen in Deutschland jedoch Defizite beim Datenzugang und bei der Nutzung von Daten durch die Wissenschaft. Die großen Potenziale von Forschungsdaten bleiben aufgrund technischer und rechtlicher Hürden häufig noch ungenutzt. Um die Möglichkeiten von Forschungsdaten für die Wissenschaft im Gemeinwohlinteresse besser zu erschließen, plant die Bundesregierung ein Forschungsdatengesetz. In vorliegendem Papier werden Empfehlungen gegeben, welche Aspekte bei der Ausgestaltung des Gesetzes berücksichtigt werden sollten.

WACHSENDE BEDEUTUNG VON FORSCHUNGSDATEN

Die Bedeutung von Daten für die Forschung wird zunehmend wichtiger. Forschung hat zwar schon immer mit Daten gearbeitet, doch sind die Möglichkeiten datenintensiver Forschung durch die Digitalisierung stark angestiegen: Es können immer mehr Daten erhoben, gesammelt, verarbeitet und bereitgestellt werden. Zudem findet Forschung immer häufiger in komplexen, interdisziplinären Projekten und großen Forschungsverbänden statt, in denen sehr viele Daten produziert werden, was ein umfassendes Datenmanagement erforderlich macht. Aber nicht nur die Menge der digitalen Daten wächst exponentiell an, sondern es sind auch immer mehr digitale Werkzeuge verfügbar, um eine große Menge an Daten zu archi-

vieren, auszuwerten und nutzbar zu machen. Datengetriebene Forschung spielt inzwischen eine zentrale Rolle bei der Bearbeitung wissenschaftlicher Fragen und leistet so auch einen wesentlichen Beitrag zur Lösung gesellschaftlicher Herausforderungen, zum Beispiel in den Bereichen Klima, Mobilität oder Gesundheit. Damit unterstützt sie entscheidend gesellschaftliche Weiterentwicklung und ermöglicht Innovation.

Datenintensive Forschung verändert auch die Wissenschaft selbst. Die intensive Nutzung digitaler Daten eröffnet der Forschung eine Vielzahl neuer Möglichkeiten und Methoden der Datengewinnung, des Datenaustauschs und der Datenanalyse. Um die mit Forschungsdaten verbundenen Potenziale bestmöglich nutzen zu können, braucht es nicht nur geeignete rechtliche und infrastrukturelle Rahmenbedingungen, sondern nach Ansicht des Wissenschaftsrats auch einen Kulturwandel, der aktiv gestaltet werden muss und verschiedene Prinzipien umfasst, etwa Teilen und Kooperieren, Themen- und Methodenvielfalt und die Sicherung wissenschaftlicher Standards (WR 2020).



Wissenschaftsrat: Zum Wandel in den Wissenschaften durch datenintensive Forschung

WAS SIND FORSCHUNGSDATEN?

Forschungsdaten sind „Aufzeichnungen in digitaler Form, bei denen es sich nicht um wissenschaftliche Veröffentlichungen handelt und die im Laufe von wissenschaftlichen Forschungstätigkeiten erfasst oder erzeugt und als Nachweise im Rahmen des Forschungsprozesses verwendet werden oder die in der Forschungsgemeinschaft allgemein für die Validierung von Forschungsfeststellungen und -ergebnissen als notwendig erachtet werden“. (Datennutzungsgesetz § 3, Abs. 10)

„Zu Forschungsdaten zählen u. a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden.“ (DFG 2015)

Qualitätsgesicherte Forschungsdaten dokumentieren die Ergebnisse des Forschungsprozesses, bilden einen Grundpfeiler wissenschaftlicher Erkenntnis und können Grundlage weiterer Forschung sein. „Die nachhaltige Sicherung und Bereitstellung von Forschungsdaten dient daher nicht nur der Prüfung früherer Ergebnisse, sondern in hohem Maße auch der Erzielung künftiger Ergebnisse. Sie bildet eine strategische Aufgabe, zu der Wissenschaft, Politik und andere Teile der Gesellschaft gemeinsam beitragen müssen.“ (Allianz der Wissenschaftsorganisationen 2010)

HINDERNISSE UND HANDLUNGSBEDARF

Gegenwärtig bleiben die großen Potenziale von Forschungsdaten in Deutschland aufgrund technischer und rechtlicher Hürden häufig noch ungenutzt. Folgende Defizite sind festzustellen:

- eine schlechte Auffindbarkeit von Daten;
- ein unklarer Rechtsstatus von Daten, die von Forscher_innen und anderen Akteur_innen generiert werden;
- fehlende Rechtssicherheit bei der Verknüpfung und Anonymisierung von Daten;
- eine problematische Abgrenzung personenbezogener von nichtpersonenbezogenen Daten;
- die ungeklärte Frage, ab wann das Urheberrecht greift;
- fehlende standardisierte Datenzugänge (vgl. RfII 2023).

Die große Bedeutung von Forschungsdaten wurde insbesondere in der Covid-19-Pandemie deutlich (vgl. WR 2020: 59ff.). In Deutschland zeigten sich dabei positive Aspekte im Bereich Grundlagenforschung (z. B. bei der Impfstoffentwicklung), aber auch Defizite in der Datenlandschaft, die dazu führten, dass wichtige wissenschaftliche Erkenntnisse nicht gewonnen werden konnten. So waren viele erforderliche Daten entweder nicht vorhanden oder nicht zugänglich, um im medizinischen Bereich notwendige datenbasierte wissenschaftliche Studien durchzuführen und wichtige politische Entscheidungen durch wissenschaftliche Evidenz zu stützen (vgl. Beermann et al. 2022).

EMPFEHLUNGEN ZUR AUSGESTALTUNG DES GESETZES

Vor diesem Hintergrund diskutierten in der hochschulpolitischen Reihe der Friedrich-Ebert-Stiftung „Eine Stunde für die Wissenschaft“ am 26.6.2023 Expert_innen aus Wissenschaft und Politik darüber, wie das FDG ausgestaltet werden sollte. Ruppert Stüwe, Mitglied des Bundestags und des Ausschusses für Bildung, Forschung und Technikfolgenabschätzung, und Dr. Stefan Lange, Leiter der Geschäftsstelle des Rates für Informationsinfrastrukturen (RfII), verdeutlichten in ihren Impulsen die aus ihrer Sicht wichtigsten Punkte.

DER RAT FÜR INFORMATIONSDATASTRUKTUREN (RFII)

Ein zentraler Akteur im Bereich Forschungsdaten ist der Rat für Informationsinfrastrukturen (RfII), ein 2014 von der Gemeinsamen Wissenschaftskonferenz (GWK) eingesetztes Expert_innengremium, das Akteur_innen aus Politik und Wissenschaft in Fragen der digitalen Informationsinfrastrukturen berät. Inzwischen erstreckt sich die Beratung auf alle Fragen der digitalen Transformation der Wissenschaft, insbesondere auf Forschungsdateninfrastrukturen (RfII 2021a).

Die 24 Mitglieder des Rates vertreten die wissenschaftlichen Nutzer_innen und Betreiber_innen von Informationsinfrastrukturen: So sind neben Forscher_innen auch Infrastruktureinrichtungen, Bund und Länder sowie das öffentliche Leben (zurzeit neben einer Landesdatenschutzbeauftragten auch Repräsentant_innen aus Unternehmen, darunter ein Verlag) mit Vertreter_innen im Rat beteiligt. Der Rat entwickelt politikfähige Empfehlungen auf Systemebene, will die Selbstorganisation der Wissenschaft stimulieren und Synergiepotenziale im System identifizieren.

Der RfII beteiligt sich auch an verschiedenen Gesetzgebungsverfahren im Bereich Forschungsdaten auf europäischer und nationaler Ebene und veröffentlicht Positionspapiere wie auch Stellungnahmen zu aktuellen Themen (z. B. RfII 2016, 2021, 2022). In einem 2023 verabschiedeten Diskussionsimpuls werden Empfehlungen zum Forschungsdatengesetz und zur Einrichtung eines Dateninstituts gegeben.

Datenzugang und -nutzung für die öffentliche Forschung verbessern

Bei der Ausgestaltung des Gesetzes muss geklärt werden, was unter Forschungsdaten zu verstehen ist, wer Anspruch auf Datenzugang haben sollte und welcher gesellschaftliche Nutzen mit dem Gesetz verbunden ist.

RfII-Geschäftsstellenleiter Stefan Lange betonte die Wichtigkeit, „Forschungsdaten“ im Gesetz genau zu defi-

nieren. Das FDG sollte nicht nur Forschungsdaten im engeren Sinne in den Blick nehmen, wie es im Datennutzungsgesetz vorgesehen ist, sondern auch

- Sekundärdaten, die den Entstehungsprozess der Primärdaten dokumentieren und kontextualisieren sowie
- Metadaten, die auf Prozesse der Qualitätssicherung (Kuratierung) im Forschungsprozess und in wissenschaftlichen Datenbanken oder Repositorien zurückgehen und die Verwendbarkeit der Daten sicherstellen, beispielsweise deren Find- und Verknüpfbarkeit erleichtern (RfII 2023: 3).

Das FDG soll laut RfII dafür sorgen, dass Daten, die unter Einsatz öffentlicher Gelder erhoben wurden, der Wissenschaft auch möglichst umfassend zur Verfügung stehen, damit das Wissenschaftssystem seine zentralen Aufgaben für die Gesellschaft wahrnehmen kann (RfII 2023: 1). Privilegierte Datenzugänge für die öffentlich geförderte Forschung seien insbesondere für eine evidenzbasierte Politikberatung unverzichtbar, da diese nur dann in hinreichend übergreifender Form, qualitätsgesichert sowie unter Berücksichtigung einer gesamtgesellschaftlichen Perspektive erfolgen könne (RfII 2023: 1).

Lange betonte, dass öffentlich finanzierte Forschungsdaten ein wissenschaftliches Gemeinschaftsgut sind: „Sie gehören niemandem persönlich, auch nicht den Wissenschaftler_innen selbst, die sie erheben.“ Im öffentlichen Interesse sei es unabdingbar, einen Forschungszugang zu öffentlich erhobenen und gehaltenen Daten auch außerhalb der Wissenschaft (z. B. in Behörden, Sozialversicherungsträgern) zu schaffen und diese verknüpfbar zu machen, indem bisherige gesetzliche Restriktionen – im Einklang mit einem forschungsfreundlich verstandenen Datenschutzrecht – beseitigt werden.

Entsprechend solle das geplante FDG auch nicht alle möglichen Zugänge zu Forschungsdaten regulieren, sondern dort für Klarheit sorgen, wo durch bisherige, nicht miteinander harmonisierte Regelungen Unsicherheiten entstanden sind und den Datenzugang für die Forschung behindern, erläuterte Lange. Das Gesetz solle zum Beispiel ermöglichen, dass der Forschung künftig Gesundheits- und Mobilitätsdaten oder Registerdaten zur Verfügung gestellt werden können. Auch die Verknüpfbarkeit von Daten und Rahmenbedingungen für deren Bereitstellung sollten im Gesetz geregelt werden. Der RfII unterstützt somit die Forderung der Deutschen Forschungsgemeinschaft (DFG), das Gesetz vor allem als ein „Datenzugangsgesetz für die Forschung“ auszugestalten, um Wissenschaftler_innen einen möglichst ungehinderten Zugang zu Daten zu bieten.

Ruppert Stüwe, der im Forschungsausschuss im Bundestag als Berichterstatter für die Digitalisierung in der Forschung bzw. für Forschungsdateninfrastrukturen mit der Frage des FDG befasst ist, stimmte Lange zu: „Überall da, wo öffentliches Geld in die Erstellung und Bereitstellung der Daten geflossen ist, sollten die Daten auch für die öffentliche Wissenschaft zugänglich sein. Ebenso sollten die Ergebnisse der Wissenschaft auf dieser Basis auch wieder allen zugänglich sein und dem Gemeinwohl zugute kom-

men.“ Entscheidend für die Nutzung der Daten sollte der Forschungszweck sein, betonte Stüwe. Dabei solle der Zweck der unabhängigen Forschung vom Zweck einer Produktentwicklung oder Ähnlichem abgegrenzt werden. So müsse beispielsweise auch künftig verhindert werden, dass Unternehmen unter dem Stichwort „Forschung“ eine Fishing-Expedition durch öffentlich generierte Datensätze durchführen, um im privaten oder rein ökonomischen Interesse ein Produkt (weiter) zu entwickeln.

Auch müsse der Umgang mit Daten geregelt werden, bei denen keine ökonomischen, sondern andere Interessen im Vordergrund stehen, etwa von Strafverfolgungsbehörden. So sollte es beispielsweise künftig möglich sein, dass Forscher_innen wissenschaftlich mit Daten arbeiten, die bei qualitativen Interviews im Rahmen von psychologischen Studien entstanden sind, ohne dass der Datenschutz und die Rechte der Datenspende_innen verletzt werden.

Der Gegenstand und Anspruch des geplanten Gesetzes muss nach Stüwe klar unterschieden werden vom Thema Open Access und Open Data, wo es um den Zugang zu wissenschaftlich erhobenen Daten geht, die ohne Zugangshürden veröffentlicht werden und allen zur Verfügung stehen. Dagegen solle im FDG der Umgang mit öffentlich erhobenen Forschungsdaten so geregelt werden, dass Wissenschaftler_innen spezifische Daten für Forschungsvorhaben anfragen können und Zugang zu Daten erhalten, wenn ein klares, nachvollziehbares Forschungsinteresse besteht. Das Forschungsdatengesetz diene dem verbesserten Zugang der Forschung zu Daten, insbesondere wenn diese in außerwissenschaftlichen Kontexten vorliegen.

Zugriff auf privatwirtschaftliche Daten differenziert regeln

Forschungsdaten entstehen sowohl in der öffentlich finanzierten Forschung als auch in der privaten Industrieforschung. Wenn der Datenzugang der öffentlichen Forschung insgesamt verbessert werden soll, stellt sich die Frage, inwieweit auch deren Zugang zu privatwirtschaftlichen Daten optimiert werden kann.

Unter bestimmten Voraussetzungen sollte die Wissenschaft die Möglichkeit haben, auf privatwirtschaftlich gehaltene Daten zuzugreifen, meinte Lange. Die öffentlich getragene Wissenschaft würde in ihren Leistungsmöglichkeiten im Bereich datenintensiver Forschung zurückfallen, wenn sie sich auf die Arbeit mit selbst erhobenen, öffentlichen Forschungsdaten beschränken müsste. Dieser Zugang sollte nicht unlimitiert sein, doch sollten die Möglichkeiten erweitert werden: Der 2022 verabschiedete europäische Data Governance Act ermöglicht Zugang zu solchen Daten bislang nur bei außergewöhnlichen Anlässen (wie etwa einer Krisensituation). Dies sei, so Lange, zum Beispiel für die Zwecke einer krisenvorbeugenden wissenschaftsgestützten Politikberatung eine starke Einschränkung.

Ein wichtiger Punkt ist dabei, dass der öffentlich finanzierten Wissenschaft für den Datenzugang keine zu hohen Kosten entstehen dürfen. Sie sollte privatwirtschaftlichen

Datengebenden höchstens eine angemessene Aufwandsentschädigung für die Datenbereitstellung erstatten müssen (RfII 2023: 7). Auch müsse darauf geachtet werden, derzeit funktionierende freiwillige Arrangements zum Datenaustausch zwischen Wissenschaft und Wirtschaft nicht durch zusätzliche legislative Hürden im FDG zu erschweren oder zu verschlechtern. So dürften beispielsweise berechnete Geschäftsgeheimnisse und auf Daten basierende Geschäftsmodelle im Unternehmensbereich nicht verletzt oder konterkariert werden (RfII 2023: 7).

Stüwe stellte klar, dass bei privatwirtschaftlichen Daten in Zukunft deutlicher geregelt werden sollte, dass Daten und Forschungsergebnisse, die in Kooperationen zwischen Wissenschaftler_innen und Unternehmen erzielt worden sind, der Öffentlichkeit zur Verfügung gestellt werden müssen. Der Bereich der „Geschäftsgeheimnisse“ sei möglichst eng zu definieren, wenn private Unternehmen mit öffentlichen Daten oder mit Unterstützung der öffentlichen Infrastruktur forschen.

Datenschutz sicherstellen

Bei der gesetzlichen Regelung von Forschungsdaten in einem Bundesgesetz wie dem FDG müssen bestehende Datenschutzregelungen auf unterschiedlichen politischen Ebenen beachtet werden. Als EU-Mitgliedstaat ist Deutschland beispielsweise dazu verpflichtet, bei der Nutzung von Daten die europäische Datenschutzgrundverordnung (DSGVO) auf nationaler Ebene umzusetzen.

DATENSCHUTZ IN DER EU

Der Datenschutz ist als Grundrecht im Recht der Europäischen Union verankert (Vertrag über die Arbeitsweise der Europäischen Union, EU-Charta der Grundrechte, Art. 8). Im Jahr 2018 ist zudem die Datenschutz-Grundverordnung der EU (DSGVO) in Kraft getreten (EU 2016), die im Wesentlichen zwei Ziele verfolgt:

- Förderung des grenzüberschreitenden Datenverkehrs innerhalb der EU im digitalen Zeitalter durch einen einheitlichen und unmittelbar geltenden Rechtsrahmen, der von allen EU-Mitgliedstaaten umgesetzt werden soll (vor allem gleiche Wettbewerbsbedingungen für Unternehmen, bessere Möglichkeiten bei der Bekämpfung von Kriminalität und Terrorismus).
- Stärkung des EU-Grundrechts auf Schutz der personenbezogenen Daten (z. B. Name und Privatanschrift, Standortdaten auf einem Mobiltelefon), indem den Betroffenen mehr Kontrolle und Transparenz bei der Datenverarbeitung gegeben wird (höhere Anforderungen an eine rechtswirksame Einwilligung der betroffenen Personen und mehr Rechte in Bezug auf Information und Auskunft).

Aufgrund des föderalistisch geprägten Systems der Bundesrepublik ist Datenschutz aber weitgehend Sache der Länder, die unterschiedliche Datenschutzregelungen in ih-

ren Landesgesetzen festgehalten haben. Für die Umsetzung und Wahrung des Datenschutzes sind wiederum Landesdatenschutzbeauftragte zuständig, die bei der Auslegung von Gesetzen – auch bei der DSGVO – oft unterschiedliche Ansichten vertreten.¹

Für Stüwe hat es oberste Priorität, den erleichterten Zugang zu Daten und die Verknüpfung von Daten datenschutzkonform sicherzustellen. Er hält es jedoch für wenig zielführend, dass gegenwärtig Bundesländer, Datenschutzbeauftragte und einzelne Einrichtungen unterschiedlich mit Daten und der Frage des Datenschutzes umgehen. Hier müsse mit dem FDG mehr Rechtssicherheit und Verlässlichkeit erreicht werden. Deshalb solle im Gesetz auch festgelegt werden, dass eine bestimmte Auslegung des Datenschutzes über Länder- und Institutionengrenzen hinweg gilt.

FDG mit anderen Gesetzen harmonisieren

Ein nationales FDG zur Nutzung, Verwendung und Verwertung von Daten muss hohen Anforderungen genügen, weil es mit bestehenden Gesetzen und parallelen Gesetzesvorhaben auf Landes-, Bundes- und EU-Ebene zu harmonisieren ist.

Nach Auffassung des RfII kann ein FDG nur dann seine Wirkung entfalten, wenn alle gesetzlichen Rahmenbedingungen (mit) adressiert werden, die bei diesem Thema zum Tragen kommen. So sind wichtige Bereiche der gewünschten erleichterten Datenzugänge oder Datenverknüpfungsmöglichkeiten bereits auf unterschiedlichen politischen Ebenen (z. B. Datenschutzrecht in Europa, Bund, Ländern) und in weiteren Gesetzen (z. B. Bundesstatistikgesetz, Steuergesetze, Urheberrechtsgesetz) geregelt oder im Begriff, geregelt zu werden (z. B. Gesundheitsdatennutzungsgesetz)² (RfII 2023: 8). Auch die derzeitige Ausgestaltung des Data Act in der EU (vgl. Europäische Kommission 2023)³ bestimmt letztlich den Handlungsspielraum auf nationaler Ebene bei der Ausgestaltung des FDG.

Geeignete Infrastruktur schaffen

Neben gesetzlichen Regelungen braucht es auch eine geeignete materielle Infrastruktur, um den Datenzugang und die Datennutzung für die Forschung zu verbessern.

¹ Der unterschiedliche Umgang der Länder mit Datenschutz zeigte sich z.B. in der Covid-19-Krise bei unterschiedlichen Vorschriften der Länder zur Datenerfassung von Restaurantgästen oder Patient_innen (vgl. Schemmel 2022).

² Das Gesundheitsdatennutzungsgesetz soll dazu dienen, Daten zum Wohl der Patient_innen besser nutzen zu können: „Kern des Gesetzes ist die erleichterte Nutzbarkeit von Gesundheitsdaten für gemeinwohlorientierte Zwecke. Dazu wird unter anderem eine dezentrale Gesundheitsdateninfrastruktur mit einer zentralen Datenzugangs- und Koordinierungsstelle für die Nutzung von Gesundheitsdaten aufgebaut“ (BMG 2023). Das Bundeskabinett hat am 30.8.2023 das Vorhaben zum Gesundheitsdatennutzungsgesetz beschlossen, das am 1.1.2024 in Kraft treten soll.

³ Im Juni 2023 haben sich die Mitgliedstaaten der EU mit dem Europäischen Parlament auf einen Entwurf des Data Act geeinigt. Damit soll ein für die gesamte EU geltender rechtlicher Rahmen für das Nutzen und Teilen von Daten geschaffen werden, um den Datenaustausch zu verbessern oder überhaupt erst zu ermöglichen.

Stüwe betonte zum einen die Wichtigkeit, neue Institutionen zu schaffen, die Daten verwalten und zugänglich machen, sie aufbereiten und auffindbar machen sowie Anträge prüfen usw. – und somit auch eine Transformation von Daten, die für die Forschung genutzt werden sollen, leisten und begleiten können. Zum anderen bedürfe es einer Hardware-Infrastruktur, die Daten über längere Zeiträume sicher und verfügbar hält. Nach Stüwe spricht dabei viel für eine dezentrale, fachspezifische Struktur, um der Unterschiedlichkeit von Forschungsdaten – von Daten im Gesundheitsbereich bis hin zu qualitativen Interviews – gerecht werden zu können.

Der RfII empfiehlt, die beiden Vorhaben – Einführung eines Forschungsdatengesetzes und Einrichtung eines Dateninstituts – als ineinandergreifende Zahnräder in einem innovationsfähigen umfassenden Datenökosystem zu begreifen und mit weiteren Bausteinen im Bereich der nationalen, europäischen und internationalen Dateninfrastrukturen zu verzahnen (RFII 2023: 1). Ein Dateninstitut sollte im Interesse der Forschung verschiedene Aufgaben übernehmen, vor allem die mangelnde Auffindbarkeit von Daten begrenzen, Rechtssicherheit bei der Verknüpfung und Anonymisierung von Daten herstellen, das Abgrenzungsproblem von personenbezogenen und nichtpersonenbezogenen Daten klären und standardisierte Datenzugänge unter anderem zu privatwirtschaftlichen Daten festlegen (RFII 2023: 6). Bei der Realisierung der parallelen Vorhaben muss nach Lange unbedingt verhindert werden, dass zusätzliche Hürden und Doppelstrukturen aufgebaut werden und unscharfe Verantwortlichkeiten bzw. überlappende Regulierungen zu Widersprüchen und neuen Rechtsunsicherheiten führen.

Da von einer sehr unterschiedlichen Qualität der Forschungsdaten ausgegangen werden muss – je nachdem, welche Institutionen oder Wirtschaftsunternehmen Daten für die Wissenschaft bereitstellen –, spricht sich der RFII dafür aus, qualitätsgesicherte Datentreuhändermodelle zu etablieren. Datentreuhänder_innen könnten die Daten auf dritter, neutraler Ebene verwalten und sie so aufbereiten (z. B. durch Anonymisierung oder Pseudonymisierung)⁴, dass sie wissenschaftlich genutzt werden können, aber über konkrete Geschäftsmodelle oder einzelne Datenspende_innen keine Auskunft mehr geben. So kann die notwendige Qualität der Daten sichergestellt und zugleich der erforderliche Datenschutz gewährleistet werden. Lange verdeutlichte, dass mit einem Datentreuhändersystem allerdings relativ hohe Kosten verbunden sind und es viel Zeit und Ressourcen braucht, bis genügend qualifiziertes Personal so-

⁴ Das Bundesdatenschutzgesetz (BDSG) und die europäische Datenschutzgrundverordnung (DSGVO) fordern die Anonymisierung und Pseudonymisierung von Daten, um den Datenschutz von Personen zu gewährleisten. Bei der „Anonymisierung“ werden personenbezogene Daten so verändert, dass Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur sehr schwer einer natürlichen Person zugeordnet werden können; bei der „Pseudonymisierung“ können personenbezogene Daten ohne zusätzliche Informationen nicht mehr einer spezifischen Person zugeordnet werden – die zusätzlichen Informationen müssen gesondert aufbewahrt werden und unterliegen bestimmten Maßnahmen des Datenschutzes. Beide Ansätze sind bei der Umsetzung mit erheblichen Herausforderungen verbunden (vgl. Bertram 2020).

wohl für die Datentreuhandstellen als auch aufseiten der Datengebenden und Nutzenden – Betriebe, Verwaltungen, Hochschulen und Forschungseinrichtungen – ausgebildet bzw. verfügbar sind (siehe auch RfII 2019).

Nationale Forschungsdateninfrastruktur ausbauen

Die bisher wichtigste Struktur bei der Sicherung der Qualität, Validität und Seriosität von Daten ist die Nationale Forschungsdateninfrastruktur (NFDI): Hier werden im Rahmen der Wissenschaft – bezogen auf Disziplinen und Forschungsfelder – Kompetenzen gebündelt, um Daten jenseits von Silo-Lösungen qualitätsgesichert zu sammeln, zu archivieren und innerwissenschaftlich zur Verfügung stellen. Die NFDI sei eine Erfolgsgeschichte, meinte Lange. Deshalb sei es sehr wichtig, dass die NFDI nicht nur dauerhaft erhalten bleibt, sondern auch als wichtige Infrastruktur eines europäischen Forschungsdatenraums weiter in die internationale Dimension hinein ausgebaut wird.

NATIONALE FORSCHUNGSDATENINFRASTRUKTUR (NFDI)

Die Nationale Forschungsdateninfrastruktur wurde auf Initiative des Rates für Informationsinfrastrukturen von der Gemeinsamen Wissenschaftskonferenz (GWK) eingerichtet und wird von Bund und Ländern bis 2029 zunächst projektförmig finanziert. Sie wird seit 2019 aufgebaut und soll:

- Datenbestände von Wissenschaft und Forschung für das gesamte Wissenschaftssystem systematisch erschließen, vernetzen und nachhaltig sowie qualitativ nutzbar machen;
- als dauerhafter digitaler Wissensspeicher dienen;
- relevante Daten nach den FAIR-Prinzipien zur Verfügung stellen (FAIR = Findable, Accessible, Interoperable und Reusable: auffindbar, zugänglich, interoperabel und wiederverwendbar) (NFDI o. J.).

Die NFDI besteht aus sogenannten Konsortien, in denen das Forschungsdatenmanagement organisiert ist: „Konsortien sind auf langfristige Zusammenarbeit angelegte Zusammenschlüsse von Nutzenden und Anbietern von Forschungsdaten und umfassen Hochschulen, außeruniversitäre Forschungseinrichtungen, Ressortforschungseinrichtungen, Akademien und andere öffentlich geförderte Informationsinfrastruktureinrichtungen. Sie sind in der Regel nach Fachgruppen beziehungsweise Methoden organisiert und haben zum Ziel, den Zugang zu und die Nutzung der für sie relevanten Forschungsdaten zu gewährleisten und nachhaltig zu gestalten“ (DFG 2021).

Die NFDI-Konsortien wurden durch ein von der Deutschen Forschungsgemeinschaft (DFG) geleitetes Verfahren in drei Förderrunden ausgewählt (vgl. DFG 2022).

Zugriff für außerwissenschaftliche Interessierte regeln

Im Umgang mit Forschungsdaten stellt sich die Frage, ob auch außerwissenschaftlich Interessierte Zugang erhalten sollten.

Aus Sicht von Stüwe ist der entscheidende Punkt, dass Interessierte bei der Datenanfrage immer die konkreten Zwecke der Datenverwendung benennen müssten. Ausschlaggebend sollte immer der Zweck der Forschung sein, sei es in öffentlichen Institutionen oder privatwirtschaftlichen Unternehmen. Dabei sollte eine klare Grenze gezogen werden zwischen einem Forschungsinteresse, das dem Gemeinwohl dient (z. B. klinische Studien, die oft in Universitätskliniken in Kooperation von Unternehmen und öffentlicher Hand durchgeführt werden), und Forschung, die einem Privatinteresse folgt. Privatwirtschaftliche Akteur_innen sollten zwar auch Zugang zu Forschungsdaten erhalten können, allerdings nur unter bestimmten Bedingungen. Es müsse beispielsweise reguliert werden, dass die Ergebnisse der darauf aufbauenden Forschung nicht privatisiert werden, sondern die Nutzung der Daten an die Bedingung geknüpft wird, die Ergebnisse dann auch der Öffentlichkeit zugänglich zu machen.

Auch laut RfII sollte es möglich sein, Daten mit außerwissenschaftlichen Interessierten zu teilen. Hingegen sei es nicht sinnvoll, externen Dritten einen bedingungslosen, pauschal geregelten Zugang zu Forschungsdaten zu ermöglichen. In jedem Fall sei sicherzustellen, dass eine gute und qualitätswahrende außerwissenschaftliche Nutzung gewährleistet ist. Deshalb sei es wichtig, dass externe Dritte konkrete Fragestellungen für ihren Datenbedarf angeben müssen bzw. der Zweck für die Nutzung der Forschungsdaten spezifiziert wird. Daraus könne dann auch abgeleitet werden, welchen Grad der Verarbeitung (z. B. Anonymisierung) Forschungsdaten haben müssen, um für außerwissenschaftliche Zwecke brauchbar zu sein, mit welchem Aufwand bzw. mit welchen Kosten dies verbunden wäre und ob die angegebenen Zwecke selbst mit wissenschaftlichen Wertvorstellungen und guter wissenschaftlicher Praxis vereinbar sind (RfII 2023: 11).

Der Schutz der Autonomie wissenschaftlicher Forschung ist für den RfII ein zentraler Punkt: Für externe Dritte sollte es nicht möglich sein, im laufenden Forschungsprozess auf Zwischenprodukte oder Forschungsspuren zuzugreifen, etwa durch Tracking von Forscher_innenhandeln. Auch sollte es keine undifferenzierte Publizitätspflicht für Forschungsdaten geben: Erst wenn Wissenschaftler_innen Daten selbstständig publizieren und qualitätsgesichert veröffentlichen, sollten diese Daten auch für externe Nutzer_innen zugänglich sein. Das gelte beispielsweise für Forschungsdaten, die in Repositorien, Forschungsdatenzentren oder Infrastrukturen wie der NFDI – zum Teil bereits aufbereitet als „Forschungsdatenprodukte“ (z. B. „Public Use Files“) – hinterlegt und ohne Weiteres zugänglich sind. Forschungsdaten sollten nach Lange auch kuratiert sein, das heißt, die Metadaten müssen Auskunft über den Entstehungskontext und die Autor_innenschaft geben, damit Daten nicht einfach zweckentfremdet oder als Eigenes ausgegeben werden können.

Sicherheitsarchitektur etablieren

Beim Zugang zu Daten spielt eine wichtige Rolle, dass in vielen Disziplinen und Forschungsfeldern Daten in internationaler Kooperation erzeugt oder angefordert werden. Zugriffe auf Daten für Forschung erfolgen oftmals global – auch aus Staaten, in denen Forschungsfreiheit und die Autonomie der Wissenschaft nicht unbedingt gesichert ist. Somit besteht auch die Gefahr, dass andere Länder Zugang zu öffentlichen Daten erhalten und diese dann für Zwecke nutzen, die Nachteile für die deutsche Gesellschaft und Wirtschaft mit sich bringen – sei es, weil die Daten sicherheitsrelevant sind, Betriebs- und Geschäftsgeheimnisse offenbaren oder personenbezogene Informationen nicht ausreichend geschützt sind.

Nach Auffassung von Stüwe stellt ein unerwünschter Datenabfluss ein reales Risiko dar. Deutschland müsse seine Dateninfrastruktur besser schützen als bisher, was für den Bereich Wissenschaft und Forschung ganz besonders gelte. Notwendig sei eine ernsthafte Diskussion über das Thema Sicherheit und insbesondere Datensicherheit in der Wissenschaft, was mit geeigneten Maßnahmen und ausreichenden Ressourcen unterlegt werden müsse.

Der RfII fordert, dass nicht alle Daten grundsätzlich für alle offen zugänglich sein sollten. Prinzip sollte vielmehr sein: „So offen wie möglich, so geschlossen wie nötig“, verdeutlichte Lange. Man werde innerwissenschaftlich, aber auch in der Beziehung zu Dritten regulativ zu Lösungen kommen müssen, die einer „abgestuften Sicherheitsarchitektur“ folgen. In dieser sollte dann festgelegt werden, welche Daten global von wem abgerufen werden können, welche nur unter bestimmten Auflagen zugänglich sind und welche unter Sicherheitsgesichtspunkten gar nicht verfügbar sein sollen.

LITERATURVERZEICHNIS

Allianz der deutschen Wissenschaftsorganisationen 2010: Grundsätze zum Umgang mit Forschungsdaten, o. O., https://gfzpublic.gfz-potsdam.de/pubman/item/item_2949914_3/component/file_2949913/Grundsätze_Forschungsdaten_2010.pdf (7.9.2023).

Beermann, Sandra; Dörr, Marcus; Grill, Eva; Karch, André; Lange, Berit; Zeeb, Hajo 2022: Coronapandemie: Die Rolle epidemiologischer Forschung in Gesundheitskrisen, in: Deutsches Ärzteblatt 2022, 119 (17): A753/B-621, <https://www.aerzteblatt.de/archiv/224889/Corona-pandemie-Die-Rolle-epidemiologischer-Forschung-in-Gesundheitskrisen> (12.9.2023).

Bertram, Catharina 2020: Anonymisierung und Pseudonymisierung, 9.12.2020, <https://www.johner-institut.de/blog/gesundheitswesen/anonymisierung-und-pseudonymisierung/> (31.8.2023).

BMBF – Bundesministerium für Bildung und Forschung 2023: Forschungsdatengesetz, Stellungnahmen aus der öffentlichen Konsultation, <https://www.bmbf.de/bmbf/del/service/gesetze/gesetze.html> (12.9.2023).

BMI – Bundesministerium des Innern und für Heimat o. J.: Datenschutz-Grundverordnung, <https://www.bmi.bund.de/SharedDocs/faqs/DE/themen/it-digitalpolitik/datenschutz/datenschutzgrundvo-liste.html> (25.9.2023).

BMG – Bundesministerium für Gesundheit 2023: Gesundheitsdatennutzungsgesetz, <https://www.bundesgesundheitsministerium.de/service/gesetze-und-verordnungen/detail/gesundheitsdatennutzungsgesetz.html> (12.9.2023).

DFG – Deutsche Forschungsgemeinschaft 2015: Leitlinien zum Umgang mit Forschungsdaten, Bonn, www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf (8.4.2020).

DFG – Deutsche Forschungsgemeinschaft 2021: Nationale Forschungsdateninfrastruktur: Förderung von Konsortien, Ausschreibung 2021, 12.5.2021, https://www.dfg.de/foerderung/info_wissenschaft/2021/info_wissenschaft_21_37/index.html (25.9.2023).

DFG – Deutsche Forschungsgemeinschaft 2022: Nationale Forschungsdateninfrastruktur, Statistische Übersicht zu den Förderentscheidungen in drei Ausschreibungsrunden, Bonn, https://www.dfg.de/download/pdf/foerderung/programme/nfdi/nfdi_statistische_uebersicht.pdf (20.9.2023).

DFG – Deutsche Forschungsgemeinschaft 2023: Die Wissenschaft braucht ein Forschungsdatengesetz, Bonn, https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/stellungnahme_forschungsdatengesetz.pdf (12.9.2023).

Europäische Kommission 2023: Gestaltung der digitalen Zukunft Europas: Datengesetz, <https://digital-strategy.ec.europa.eu/de/policies/data-act> (15.9.2023).

Europäischer Rat; Rat der Europäischen Union o. J.: Datenschutz in der EU, <https://www.consilium.europa.eu/de/policies/data-protection/> (20.9.2023).

EU – Europäische Union 2016: Datenschutz in der EU, <https://www.consilium.europa.eu/de/policies/data-protection/> (20.9.2023).

Rfll – Rat für Informationsinfrastrukturen 2016: Leistung aus Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen, <https://rfii.de/?p=1998> (10.8.2023).

Rfll – Rat für Informationsinfrastrukturen 2019: Digitale Kompetenzen – dringend gesucht! Empfehlungen zu Berufs- und Ausbildungsperspektiven für den Arbeitsmarkt Wissenschaft, Göttingen, <https://rfii.de/download/digitale-kompetenzen-dringend-gesucht?p=1998> (18.10.2023).

Rfll – Rat für Informationsinfrastrukturen 2021a: Website des Rfll, <https://rfii.de> (1.11.2023).

Rfll – Rat für Informationsinfrastrukturen 2021b: Nutzung und Verwertung von Daten im wissenschaftlichen Raum: Empfehlungen zur Ausgestaltung von Datendiensten an der Schnittstelle zwischen Wissenschaft und Wirtschaft, Göttingen, <https://rfii.de/download/nutzung-und-verwertung-von-daten-im-wissenschaftlichen-raum-september-2021/> (10.8.2023).

Rfll – Rat für Informationsinfrastrukturen 2022: Datenpolitik, Open Science und Dateninfrastrukturen: Neue Initiativen im europäischen Raum, Göttingen, <https://rfii.de/download/datenpolitik-open-science-und-dateninfrastrukturen-aktuelle-entwicklungen-im-europaeischen-raum-juli-2022/> (12.9.2023).

Rfll – Rat für Informationsinfrastrukturen 2023: Diskussionsimpuls: Anforderungen an die Ausgestaltung eines Forschungsdatengesetzes und eines Dateninstituts, Göttingen, <https://rfii.de/download/rfii-diskussionsimpuls-forschungsdatengesetz-2023/> (1.9.2023).

NFDI – Nationale Forschungsdateninfrastruktur o. J.: Website, <https://www.nfdi.de/> (19.9.2023).

Schemmel, Frank 2022: Ist eine Datenschutzbehörde für jedes Bundesland wirklich notwendig? DataGuard, <https://www.dataguard.de/blog/datenschutzbehoerden> (10.9.2023).

SPD; Bündnis 90/Die Grünen; FDP 2021: Koalitionsvertrag 2021–2025: Mehr Fortschritt wagen: Bündnis für Freiheit, Gerechtigkeit und Nachhaltigkeit, <https://www.bundesregierung.de/resource/blob/974430/1990812/1f422c60505b6a88f8f3b3b5b8720bd4/2021-12-10-koav2021-data.pdf?download=1> (10.8.2023).

WR – Wissenschaftsrat 2020: Zum Wandel in den Wissenschaften durch datenintensive Forschung, https://www.wissenschaftsrat.de/download/2020/8667-20.pdf?__blob=publicationFile&v=4 (10.9.2023).

AUTORIN

Dr. Angela Borgwardt ist Politikwissenschaftlerin und arbeitet als freie Publizistin und Moderatorin zu wissenschafts- und gesellschaftspolitischen Themen.

IMPRESSUM

ISBN: 978-3-98628-344-5

November 2023

© **Friedrich-Ebert-Stiftung**

Herausgeberin: Abteilung Analyse, Planung und Beratung
Godesberger Allee 149, 53175 Bonn

www.fes.de/apb

Für diese Publikation ist in der FES verantwortlich:
Florian Dähne

Bestellungen/Kontakt: apb-publikationen@fes.de

Satz: minus design, Berlin

Bild Seite 1: © 2023, Johannes Beck/minus design

Die in dieser Publikation zum Ausdruck gebrachten Ansichten sind nicht notwendigerweise die der Friedrich-Ebert-Stiftung. Eine gewerbliche Nutzung der von der FES herausgegebenen Medien ist ohne schriftliche Zustimmung durch die FES nicht gestattet. Publikationen der Friedrich-Ebert-Stiftung dürfen nicht für Wahlkampfzwecke verwendet werden.