

Stellungnahme zur öffentlichen Konsultation zum Forschungsdatengesetz

der Sektion Ethical, Legal & Social Aspects (ELSA)
des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e.V.

Vorbemerkung

In der Nationalen Forschungsdateninfrastruktur (NFDI) e.V. mit derzeit 247 Mitgliedsorganisationen werden wertvolle Datenbestände von Wissenschaft und Forschung für das gesamte deutsche Wissenschaftssystem systematisch erschlossen, vernetzt und nachhaltig sowie qualitativ nutzbar gemacht. NFDI besteht aus 26 Konsortien und einem Verbund von Konsortien, die sich dem Forschungsdatenmanagement in den Natur-, Lebens-, Ingenieurs-, Kultur- und Sozialwissenschaften widmen. Innerhalb des NFDI-Vereins haben sich rechtlich unselbstständige Abteilungen gebildet, sogenannte Sektionen. In diesen Sektionen werden Querschnittsthemen über die Grenzen der Fachkonsortien hinweg transdisziplinär vorangetrieben. Eine Aufgabe dieser Sektionen ist die Begleitung von Gesetzgebungsprozessen, und so auch die Erstellung von Stellungnahmen aus der Perspektive wissenschaftlicher Praxis. Gemeinsam gestalten der Verein, die Konsortien und Sektionen die Zukunft des Forschungsdatenmanagements in Deutschland und knüpfen an (inter-)nationale Entwicklungen und bestehende Initiativen an.

Einleitung

Das im Koalitionsvertrag der Parteien SPD, Bündnis 90/Die Grünen und FDP vorgesehene Forschungsdatengesetz (FDG) soll den Zugang zu Forschungsdaten für öffentliche und private Forschung verbessern und umfassend vereinfachen.¹ In Forschungsdaten liegt großes Potenzial für Gesellschaft, Wirtschaft und Wissenschaft sowie für Umwelt und Klima, das oft zu wenig oder gar nicht genutzt wird. Das ungenutzte Potenzial sollte durch dafür geeignete Forschungsklauseln – beispielsweise im Rahmen des FDG – erschlossen werden. Geeintes Ziel der Klauseln ist, Raum für die effektive und innovative Nutzung der wertvollen Daten zu schaffen.

¹ SPD, Bündnis 90/Die Grünen und FDP. (2021). Koalitionsvertrag 2021-2025 zwischen SPD, Bündnis 90/Die Grünen und FDP: Mehr Fortschritt wagen. Bündnis für Freiheit, Gerechtigkeit und Nachhaltigkeit. Absatz: Forschungsdaten, S. 21.

Dieses Ziel verfolgt auch der Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. zusammen mit 26 Konsortien und einem Verbund aus Konsortien der verschiedenen Wissenschaftsdisziplinen, die sich dem FAIRen Umgang mit Daten widmen, um die Verfügbarkeit, den Zugang, die Interoperabilität und Wiederverfügbarkeit von Daten zu verbessern. Der NFDI-Verein begrüßt daher das anstehende Gesetzesvorhaben und möchte an dieser Stelle die Aspekte, die bei der Gestaltung eines Forschungsdatengesetzes (FDG) berücksichtigt werden müssen, hervorheben. Dabei liegt im Fokus der Zugang zu Forschungsdaten aus verschiedenen Quellen sowie das Sichern der Nachhaltigkeit der Daten.

1 Bedarfe

a. Daten für die Forschung

Daten sind längst nicht nur Kernbestandteil der wissenschaftlichen Forschung, sondern auch unserer alltäglichen Welt. Ihre Bedeutung für die Gesellschaft und Wissenschaft nimmt kontinuierlich zu und enthüllt ein bemerkenswertes Potenzial für die gemeinwohlorientierte Forschung. Dennoch werden (Forschungs-)Daten in Deutschland heute noch immer zu wenig genutzt. Heterogene oder gar fehlende Zugangsregelungen tragen zu dieser Tendenz bei.

Gleichzeitig setzen Innovation und Förderung des wissenschaftlichen Fortschritts einen diskriminierungsfreien Datenzugang voraus. Bessere Datenzugangsregelungen begünstigen nicht nur die Nachprüfbarkeit von Forschungsergebnissen und einen „Faktencheck“, sondern fördern auch interdisziplinäre Forschung und Kooperationen. Zudem sind sie Voraussetzung für eine nachhaltige Nutzung für Wirtschaft, Gesellschaft, Verwaltung und Politik, deren Akteure wiederum evidenzbasierte Entscheidungen treffen können. Es besteht daher ein Bedarf an einem einheitlichen gesetzlichen Rahmen, der die Datenzugangsmöglichkeiten für die Wissenschaft und Forschung klar regelt.

Aus Sicht des NFDI-Vereins sollen **bessere Zugangs- und Verknüpfungsmöglichkeiten** für alle Daten ermöglicht werden, die dazu beitragen, gegenwärtige und zukünftige gesellschaftliche Herausforderungen zu verstehen und Lösungen zu entwickeln. Dabei sollten generell einzelne Forschungsfelder nicht ausgeklammert werden. Schließlich arbeiten alle Disziplinen auf ihre Weise mit (heterogenen) Forschungsdaten. Entsprechende Zugangshürden liegen aber nicht allein in der jeweiligen Disziplin, sondern in der Beschaffenheit der jeweiligen nachzunutzenden Daten, die über Disziplinengrenzen hinweg gemäß FAIR-Grundsätzen nachgenutzt werden können und sollten.

Relevant sind somit im Rahmen der NFDI die Datenzugänge für alle Konsortien mit ihren jeweiligen disziplinspezifischen Bedarfen, ihren speziellen Datenbezügen und Entstehungskontexten aus ihren Forschungscommunities. Zu berücksichtigen sind darüber hinaus aber auch die Bedarfe abgelehnter NFDI-Konsortien, die sich durch besondere Anforderungen an das Forschungsdatenmanagement auszeichnen. Aus der Verknüpfung dieser verschiedenen Belange ergibt sich ein Geflecht aus inter- und transdisziplinären Bezügen; es gibt kaum noch Forschungsbereiche, in denen der Forschungsgegenstand nicht in Form von Forschungsdaten Betrachtung findet.

Forschungsdaten sind also Daten, die im Forschungsprozess entstehen oder auf die die Wissenschaft für Forschungszwecke zugreift, um sie als methodisch erforderliche Grundlage für den konkreten Forschungsprozess zu verwenden.²

Durch die Interdisziplinarität der Forschungsbereiche wird die Bedeutung eines **diskriminierungsfreien Datenzugangs** begründet. Dabei sind allen wissenschaftlich Forschenden die gleichen Zugangsrechte und -pflichten zu gewähren. Forschende und Forschungseinrichtungen müssen Vorgaben der Datensicherheit, Geheimhaltungspflichten und den Ausschluss kommerzieller Nutzung einhalten (siehe NFDI Positionspapier zum direkten Recht eines Datenzugangs für die öffentliche Forschung³).

Besonders wertvoll für die Forschung sind Daten, die einzigartig sind, nicht ohne Weiteres neu generiert werden können oder einen hohen wissenschaftlichen, historischen und gesellschaftlichen Wert haben (z.B. Zeitzeugnisse, Oral History etc.), unabhängig von Herkunft, Quelle und der Heterogenität der Daten. Darunter fallen nicht nur neue Daten, sondern auch existente Datensätze inkl. Metadatenschemata, die eine Mindestvoraussetzung für die effektive Nachnutzung sind. Schon **bestehende Datensätze, „Rohdaten“ und publizierte Daten** aus den unterschiedlichsten Bereichen von Wirtschaft und Gesellschaft sollten für die Forschung bereitgestellt werden. Zudem sind gesellschaftlich relevante Daten, die für die Daseinsvorsorge unentbehrlich sind – z.B. im Bereich Energie, Umwelt, Telekommunikation, IoT – von besonderem Interesse. Aber auch Daten aus dem Social-Media-Bereich, Unternehmensarchive und sonstige Archive nicht-öffentlicher Einrichtungen sind forschungsrelevant.

Im Fokus steht außerdem das **nachhaltige Aufbereiten und Zurverfügungstellen von Forschungsdatensätzen**: Die Zugangs-, Nutzungs- und Verknüpfungsmöglichkeiten für viele Forschungsdaten sind langfristig nicht lohnenswert, weil sie sich auf zahlreiche Forschungsdatenzentren, Behörden, Bibliotheken, Hochschulen und Archive verteilen und nach einer aktiven Nutzungsphase (z.B. 10 Jahre) das Interesse zurückgeht. Der Aufwand zur Pflege und zum Erhalt der Datensätze steigt jedoch.

² Das Verständnis von Forschungsdaten ist an die einschlägige Definition von Forschungsdaten angelehnt, die durch den Rat für Informationsinfrastrukturen (RfII) verabschiedet wurde. Siehe hierzu Rat für Informationsinfrastrukturen, Schlüsselbegriffe, <https://rfii.de/de/themen/>.

³ Verabschiedet durch die Konsortialversammlung, siehe <https://www.nfdi.de/positionspapier-recht-auf-datenzugang-fuer-oeffentliche-forschung-schafft-gesellschaftlichen-mehrwert/>.

Ein Beispiel dafür sind Datensätze der öffentlichen Verwaltung, die vorwiegend ereignis- oder themenbedingt hohen Relevanzschwankungen unterliegen, aber dennoch aus öffentlichem Interesse gepflegt werden müssen. Forschungsdaten von bleibendem Wert sollten daher in einen einheitlichen rechtlich-organisatorischen Rahmen gebracht werden. Ein seit fast 40 Jahren gegebener Rahmen sind z.B. die Archivgesetze des Bundes und der Länder, die entsprechend anzuwenden sind und regelmäßig modernisiert werden.

Dabei muss die Qualität von Forschungsdaten oder einer den Datensatz ausweisenden Stelle für jeden Forschenden eindeutig einschätzbar sein. Grundsätzlich können nur Forschungsdaten für eine Forschung herangezogen werden, die gut kuratiert sind und die über zertifizierte Systeme mit eindeutigen, langzeitverfügbaren PIDs (Persistent Identifier) versehen sind.

Häufig steht dem Einrichten solcher Identifikationsmerkmale die rechtliche Inhaberschaft bzw. Provenienz entgegen. Daraus können sich erhebliche Probleme ergeben, insbesondere wenn neue Ergebnis-Datensätze durch Datenverarbeitung aus Forschungsdaten hervorgehen. In diesen Fällen ist Software oft ein essentieller Bestandteil der Datenverarbeitung. Um die Reproduzierbarkeit der Daten zu gewährleisten, sind demnach nicht nur die Daten selbst verfügbar zu machen, sondern auch die Details der benutzten Datenverarbeitung und damit auch die benutzte Software. Insofern muss Forschungssoftware nicht nur selbst als Datenart, sondern auch als essentieller Bestandteil anderer Arten von Daten gesehen werden.

Für die Forschung relevant können alle Arten von Daten aus Wirtschaft, Wissenschaft und Gesellschaft sein oder werden. Jedoch sind nicht alle Daten gleich – es gibt unterschiedliche Anforderungen im Hinblick auf Dauer der Archivierung, Qualität und Umfang der Dokumente und die erwartete Nachnutzung. **Forschungsdaten sind per se heterogen bzw. divers.** Aus ressourcentechnischen Gründen müssen Forschungsorganisationen/ Datenmittleragenturen sowie Fachinformationsdienste jene Mindestanforderungen entwickeln, die über Aspekte der grundsätzlichen Archivierbarkeit/Archivierung entscheidet. Dann muss das Nachnutzungspotential ermittelt werden, welches über die erforderlichen Ressourcen und Aufwände für das Archivierungslevel bestimmt.

b. Daten aus der Forschung

Für **öffentliche Stellen, Unternehmen und Zivilgesellschaft** sind Zugangs- und Nutzungsmöglichkeiten zu öffentlich finanzierten Daten sowie Daten aus internationalen Großforschungsanlagen von besonderem Interesse.

Jedoch ist bei allen Daten, die mit öffentlichen Mitteln beforscht und durch öffentlich finanzierte Infrastrukturen erhoben wurden, zu beachten:

- Forschungsdaten müssen so lange im Bereich der Vertraulichkeit bleiben, bis die Forschenden ihren eigenen wissenschaftsinternen Prozess der Erkenntnisgewinnung abgeschlossen haben. Eine generelle Unterstützung der vorhandenen Maßnahmen bspw. im Rahmen des DFG-Kodex „Leitlinien zur guten wissenschaftlichen Praxis“ sollte das Gesetz unbedingt berücksichtigen.
- Der Fokus des Gesetzes sollte nicht auf der Bereitstellung von Forschungsdaten im Kontext der Auftragsforschung liegen. Dies würde Geschäftsmodelle tangieren und kann Industriepartner ggf. abschrecken, entsprechende Projekte einzugehen. Grundsätzlich öffentlich geförderte Projektaktivitäten (bspw. durch Ministerien wie dem BMBF oder durch die DFG) sollten berücksichtigen, dass die dort erhobenen Daten möglichst offen zur Verfügung gestellt werden.

Zudem soll bei öffentlich finanzierten Projekten zentral verzeichnet werden, durch wen Forschungsdaten erhoben und wo diese nach Abschluss des Projektes archiviert wurden.

Bei Wissenschaftskommunikationsprojekten besteht insbesondere Bedarf an Forschungsdatensätzen, bei denen die Nachnutzung rechtlich klar geregelt ist, z.B. mit einer Nutzungslizenz (vgl. FAIR-Prinzipien R1.1).⁴ Zudem werden langfristige Forschungsdaten sowohl von der Forschung als auch von Wirtschaft und Zivilgesellschaft nachgefragt. Die Zwecke sind äußerst vielfältig und reichen von globalen Herausforderungen (z.B. Klimaforschung, Pandemieprävention), über Gewinnerzielung und Technologieentwicklung (z.B. effiziente Datenanalysemöglichkeiten in Echtzeit, Training von Big Data) bis zum rein privaten Interesse (z.B. Ahnenforschung).

Genau wie die Daten *für* die Forschung, so können auch Daten *aus* der Forschung zur Steigerung des Gemeinwohls beitragen. Sie sind die Grundlage neuer Erkenntnisse, ihre Nachnutzbarkeit ermöglicht ein effizienteres/transparenteres Ressourcenmanagement. Letztendlich sind Daten für die Entscheidungsfindung (auf allen Ebenen und in allen Bereichen) unerlässlich. Wichtig ist jedoch das **Beibehalten des Forschungszwecks**. Beispielsweise könnten öffentliche Stellen von dem Zugang zu Forschungsdaten profitieren, indem die Daten zur Verbesserung des Gesundheits- und Pflegesektors in Deutschland, z.B. Verbesserung von Diagnose, Therapie und Prävention, eingesetzt werden. Darüber hinaus führt eine bessere Datenverfügbarkeit zur Verminderung unbewusster Voreingenommenheit (unconscious bias, z.B. gender bias). Abseits davon unterstützt eine breite Zugänglichkeit der Forschungsdaten Wissenschaftskommunikationsprojekte, bei denen Jugendliche mit echten Forschungsdaten arbeiten und so einen authentischen Einblick in moderne Forschung erhalten.⁵ Weitere gesellschaftliche Aspekte für eine Zugänglichkeit der Forschungsdaten selbst sind:

⁴ Siehe <https://www.go-fair.org/fair-principles/r1-1-metadata-released-clear-accessible-data-usage-license/>

⁵ Siehe <https://www.teilchenwelt.de/angebote/masterclasses/>.

- der ökologische Fußabdruck, um Biodiversität- und Klimaindikatoren für die Unternehmen und Zivilgesellschaft zu erfassen;
- die Begründung von politischen Entscheidungen, z.B. Klimapolitik;
- die Steigerung des Innovationspotenzials;
- die Unterstützung der demokratischen Willensbildungsprozesse;
- die Transparenz gegenüber Geldgebern und der Gesellschaft insgesamt;
- institutionelle Selbstreflexion (aus historischer Sicht).

Auch die Forschung verfügt über heterogene Datensätze und -quellen:

- Daten aus Forschungseinrichtungen;
- sämtliche Daten, die in Bibliotheken, Archiven, Museen und Datenzentren vorliegen bzw. erzeugt werden;
- Forschungsdaten, die durch kostenintensive Verfahren (z.B. Forschungsexpeditionen in schwer zugänglichen Gebieten oder mit/von ressourcenintensiven Technologien/Prozessen) erhoben oder gewonnen wurden;
- Forschungsdaten, die in Hochschulen sowie in staatlicher oder privater Forschungstätigkeit entstehen.

Für die Nutzung in der Wissenschaftskommunikation sind auch ältere Datensätze interessant, d.h. die neuesten Forschungsdaten müssen zu diesem Zweck nicht sofort öffentlich zugänglich werden. Zur didaktischen Reduktion können sich ältere Datensätze teilweise sogar besser eignen.

c. Was würde die Bereitschaft zum Datenteilen auf Seiten von Unternehmen bzw. der Wirtschaft, Wissenschaft, Zivilgesellschaft und öffentlichen Stellen erhöhen? Welche Anreize könnten gesetzt werden?

Damit die **Bereitschaft zum Datenteilen** allgemein erhöht werden kann, soll genügend Wissen darüber vermittelt werden, wie technische, soziale und juristische Herausforderungen zu überwinden sind. Anreize sollen geschaffen werden, s.u., und der Mehrwert sollte noch abgebildet werden. Dadurch können Hemmnisse und Ängste von allen Seiten abgebaut werden, damit Daten nicht nur geteilt, sondern **gemeinsam genutzt werden können**. Das „Five Safes“ Modell von UK Data Service z.B. bietet mehr Optionen, als nur Daten herauszugeben.⁶

⁶ UK Data Service, What is the Five Safes framework? (2022), <https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/>.

Unternehmen bzw. der Wirtschaft

Die Bereitschaft zum **Datenteilen auf Seite der Wirtschaft** kann bspw. durch folgende Maßnahmen erhöht werden:

- klare rechtliche Rahmenbedingungen, etwa zum Haftungsausschluss;
- Finanzielle Wertschöpfungsmodelle für Nachnutzung unternehmerischer Daten;
- Standards für die Qualität von Daten: viele Unternehmen werden die Nachnutzung von frei verfügbaren Daten von eigenen Kriterien an die Qualität abhängig machen;
- Datenteilen als Voraussetzung für Fördermittel.

Wissenschaft

Auch in der Wissenschaft soll das Bewusstsein über Datennutzung gefördert werden. Wichtig dafür sind:

- die frühe Sensibilisierung der Forschenden im Rahmen ihrer akademischen Laufbahn (z.B. Datenmanagement als Pflichtteil von Bachelor-/Master-Studiengängen) sowie deutschlandweite Einführung von Datenmanagement-Plänen in Lehrplänen⁷;
- angemessene Ressourcenausstattung;
- gleichwertige Anerkennung von zitierbaren FAIRen Datenpublikationen (z.B. in den References der Journale) und zugehöriger Software-Tools;
- Schaffung unbefristeter Mittelbaustellen für die Forschungsunterstützung sowie Binden gewonnener und angereicherter Expertise von/in Personal zur Qualitätssicherung;
- gesicherte Qualitätsmanagementverfahren und -zertifizierungen, die Vertrauen in bereitgestellte Daten geben, um wirklich den Nutzungswert der Daten einschätzen zu können und damit die Nachnutzung zu erhöhen.

Zudem sollten Datenpublikationen (z.B. im beruflichen Lebenslauf) und auch entwickelte Softwareprodukte einen ebenso hohen Stellenwert wie ein veröffentlichter wissenschaftlicher Artikel einnehmen.

Zivilgesellschaft

- Wertschätzung, Nennung des Beitrages zum Erkenntnisgewinn für Wissenschaft und Gesellschaft (vgl. Motivation für Citizen Science)
- größere Sinnstiftung für Citizen-Science-Projekte
- erhöhte Transparenz und Unterstützung

⁷ Berücksichtigung des zeitlichen und finanziellen Aufwands für die notwendige Aufbereitung von Forschungsdaten vor Veröffentlichung in der Planung und Förderung von Forschungsprojekten.

Öffentlichen Stellen

Für die Erhöhung des Datenteilens allgemein sind Incentivierung und Gratifikationsmechanismen nötig. Viele der derzeitigen Open-Science-Bemühungen fordern Wissenschaftlerinnen und Wissenschaftler auf, "alle Ihre Daten zu teilen". Dies ist auch deshalb bisher nicht zum Mainstream geworden, da Wissensproduzierende hierfür nicht angemessen belohnt werden. Bei Forschungsdatenpublikationen fehlt es an der gleichzeitig mitlaufenden Entwicklung von Reputationsmechanismen für die Forschenden. Vor allem aus wirtschaftlichen Gründen (begrenzte Forschungsprojektmittel) wird oft gezögert, unnötige Aufwände einzubringen. Das Gesetz sollte einen Beitrag dazu leisten, dass Forschungsdatenpublikationen durch Forschendenteams und deren Sorgfalt und wissenschaftliche Genauigkeit incentiviert werden. Dies könnte beispielsweise gelingen durch:

- Verbesserung der Wissenschaftskommunikation;
- Aufklärung, Erhöhung von Servicementalität;
- Forschungsergebnisse für die eigene Einrichtung gewinnbringend nutzen (insbesondere Museen und kuratierend tätige Einrichtungen);
- Vereinheitlichung der Zugangswege und einheitliche Interpretationen der Gesetzgebung zu Zugangswegen über die Disziplinen hinweg im Rahmen des Archivrechts.

Allgemeine Anreize

- Reputationsgewinn, wenn ein Datensatz Grundlage in weiteren Forschungsprojekten /-ergebnissen ist;
- Bereitstellung von implementierbaren IT-Standards (z.B. TC215 ISO);
- für die Wissenschaft: höhere Zitation, größere Reichweite der Erkenntnisse, Erweiterung der eigenen Datensätze, Validierung der Forschung
- Beförderungen sollten Open Science berücksichtigen;
- Research Assessment sollte Daten berücksichtigen (siehe z.B. CoARA⁸);
- Anerkennung/Bewertung der Veröffentlichung von Forschungsdaten nach FAIR-Prinzipien durch Universitäten etc.;
- Training, Ausbildung von (Big Data) Fachpersonal sowie zusätzliche Stellen und Förderung an den Experimenten und Forschungseinrichtungen für modernes Big Data-Handling, sowie Verpflichtung zur Daten-Publikation bei Verwendung öffentlicher Gelder;

⁸ Siehe <https://coara.eu/>.

- Anerkennung der Datenpublikation als (Teil) der Forschungsleistung – nicht nur Publikation, auch Bereitstellung mit Zugangshürden (im Sinne von FAIR zugänglich, aber nicht offen);
- Forschungspreise für hochqualitative Datensätze.

Öffentliche Förderung könnte die Veröffentlichung von Daten zur Grundlagenforschung samt deskriptiver Metadaten zum Zeitpunkt der Datenerhebung zum Standard erheben, wobei Sperrfristen und Nicht-Veröffentlichung einer Begründung im Antrag bedürfen. Darüber hinaus sollten Datenpublikationen, Veröffentlichung “negativer” Ergebnisse und Studien zur Reproduzierbarkeit gesondert von Förderern gefordert und explizit für einen positive Förderentscheidung einbezogen werden, um die wissenschaftliche Fehlerkultur zu bestärken.

d. Wie schätzen Sie die Bedeutung von Forschungsk Kooperationen (z.B. mit Partnern aus der Wissenschaft und der Wirtschaft) in Bezug auf das sektorübergreifende Teilen von Daten ein?

Nur mit einem Datenaustausch können auch sinnvolle Kooperationen zwischen Grundlagenforschung und Wirtschaft, z.B. in Bezug auf allgemein zugängliche Software, geschlossen werden. Es besteht daher ein hoher Bedarf an Leuchtturmprojekten, die eine offene, nachnutzbare Datenkultur im Sinne der FAIR Data Principles pflegen – also maschinenlesbare, verknüpfte Daten mit hohem Informationsgehalt in kuratierten Repositorien.

Eine Kooperation macht Forschungsdaten, die z.B. in der archivischen Welt sehr zersplittert vorliegen (2488 Archive in Deutschland laut Archivportal-D⁹), erst voll nutzbar. Kooperationen mit Wirtschaft und Privatpersonen sind für Daten über Naturphänomene (z.B. Gewässerkunde) oder nicht mehr lebende Personen (z.B. Kunstgeschichte, Ortsgeschichte) besonders sinnvoll und hilfreich. Für personenbezogene Daten kommt mangels entsprechender Einwilligungen möglicherweise eine Teilbereitstellung (durch Anonymisierung, Pseudonymisierung) an Wirtschaft und Privatpersonen infrage. Für die Kooperation mit der Wissenschaft über personenbezogene Daten sind, im Archivrecht und der DSGVO, Privilegien vorgesehen.

⁹ Siehe <https://www.archivportal-d.de/struktur>.

Insbesondere größere und renommiertere Forschungsk Kooperationen können dafür sorgen, dass die notwendigen Kompetenzen und die teilweise umfangreichen FDM Maßnahmen für die Verwendung von Forschungsdaten in ihren Projekten bereitgestellt und auch durchgängig eingehalten, standardisiert und „unter Policy“ gesetzt werden. Größere Forschungsverbände können ihrerseits auch zur Weiterentwicklung von FDM-Technologien und -Maßnahmen beitragen. Kleine Forschungsgruppen haben bspw. keine Supportstrukturen und neue Aufgabenrollen zum FDM (z.B. „Project Chief Data Curator“, „Data Librarian Team“) im Projektkontext verfügbar.

e. Welche Rolle und welche Aufgaben sollten Datenmittlerstrukturen beim Datenteilen einnehmen?

Daten müssen als Teil einer kritischen, möglicherweise dezentralen Infrastruktur einer Gesellschaft betrachtet werden und in diesem Sinn müssen diese auch geschützt bzw. die daraus gewonnenen Erkenntnisse bedacht werden, möglicherweise in einem föderierten System im Zusammenhang mit Datenmittlerstrukturen und Portalen, über welche die Daten entdeckt und beantragt werden können.¹⁰

Datenmittlerstrukturen bestehen aus den technischen (Rechenzentren) und den organisatorischen Infrastrukturen (z.B. Gedächtniseinrichtungen, fachspez. Datenzentren). Die Strukturen sind in die aktive Nutzungsphase der Forschungsdaten und die Archivierungsphase (derzeit verteilt auf Archive, Behörden, Bibliotheken und Forschungseinrichtungen) im Anschluss daran zu differenzieren. Eine strikte Unterteilung in datenaltruistische Organisationen und Datenvermittlungsdienste (wie im EU Data Governance Act) führt im Forschungsdatenbereich vermutlich nicht weiter, da sich in der Wissenschaft das Bereithalten der Infrastruktur kaum vom Wissen über die Forschungsdaten trennen lässt.

Im Kern sind Datenmittler und Datentreuhänder essentiell für den **Aufbau von Vertrauen in die Rechtmäßigkeit der Datennutzung** sowie für die **Vermittlung von kritischen Daten**.

Bereits die Expertengruppe der EU-Kommission beschrieb, dass Vertrauen zwischen dem privaten und öffentlichen Sektor der Schlüssel für einen wirklichen B2G (Business-to-Government) ist.¹¹ Ähnliches würde auch für B2A (Business-to-Academia) gelten. Folglich muss Vertrauen durch rechtlich und technisch sichere Datenmittlern etabliert werden. Denn diese können nicht nur technische Strukturen zur Verfügung stellen, sondern auch eigenständige Akteure mit eigener, auch ethischer Verantwortung für Datenbestände sein.

¹⁰ Siehe z.B. https://forschen-fuer-gesundheit.de/menu_find.php.

¹¹ Siehe <https://digital-strategy.ec.europa.eu/en/news/commission-appoints-expert-group-business-government-data-sharing>.

Übergreifende Infrastruktur- und Katalogdienste wie z.B. die NFDI, die Deutsche Digitale Bibliothek und ihre Teilportale (z.B. Archivportal-D, Zeitungsportal) sollten aber als Datenmittler eine herausgehobene und dauerhaft finanzierte Stellung erhalten. Langfristig sollten Datenmittler Daten inhaltlich und nicht nur technisch kuratieren können. Folglich soll die angebotene Unterstützung über die technische Infrastruktur und technische Fragen hinausgehen und sich auch auf soziale und rechtliche Fragen beziehen. Somit würden sich **Datenmittler als universelles Serviceangebot** etablieren, das u.a. folgende Aspekte beinhaltet:

- Standardisierung:
 - Nutzung von DOIs oder anderer geeigneter PIDs für publizierte Datensätze mindestens nachdrücklich empfehlen
 - Entwickeln von leicht verständlichen Standards für alle vier Buchstaben von FAIR
 - Entwicklung von Regeln zur Erzeugung von Rechtssicherheit
- Prüfung der datenschutzrechtlichen Anforderungen:
 - adäquate (an die informierte Einwilligung angepasste) Archivierung und (Teil-)Veröffentlichung der Daten
 - regelmäßige Updates zu datenschutzrechtlichen Themen
 - Archivierung und Veröffentlichung von qualitativen Daten nach Prüfung bezüglich Anonymisierung/Pseudonymisierung
- Klärung von Nachnutzungsbedingungen und Lizenzen;
- Lizenzierung von Metadaten zu Datensätzen in CC0 und Zurverfügungstellen dieser über geeignete Schnittstellen
- Übernahme einer Treuhänderfunktion, sodass auch sensible Daten langfristig im Original aufbewahrt werden können
- Verlinkung von Daten über die eigenen Infrastrukturen hinaus

2 Hindernisse

Der Transfer von Daten aus und in die Forschung sieht sich in der NFDI mit mehreren Hindernissen konfrontiert. Die unter 3. aufgezeigten Lösungen referenzieren sodann auf die ausgeführten Hindernisse.

Allem voran hemmen **rechtliche Hindernisse** und Unsicherheiten die Forschenden, einen offenen Umgang mit Forschungsdaten zu pflegen. Grundlage dafür ist eine einheitliche Definition des Forschungsdatenbegriffs. Damit verbunden ist auch das Bedürfnis, verbundene Begriffe wie „forschungsrelevante Daten“ sowie Datenkategorien zu definieren. Da die Relevanz von Daten für die jeweilige Forschungsdisziplin bzw. das jeweilige Konsortium unterschiedlich ausfällt, bedarf es hier objektiver und nachvollziehbarer Kriterien.

Die Abstraktionsfähigkeit sollte dabei auch Rücksicht auf die in der jeweiligen Disziplin möglichen Maßnahmen nehmen; je Disziplin besteht ein anderes Verständnis von Forschungsrelevanz und nachhaltiger Datennutzung. Diese Klarheit ist letztlich auch für die Durchsetzbarkeit der im FDG festgehaltenen Aufgaben, Pflichten und Rechtsrahmen notwendig. Dies schließt abstrakte Kriterien zur Bestimmung der Datenqualität ein, sodass ein je Disziplin auszufüllender Spielraum besteht. Auf diese Weise könnten konkrete, zeitgemäße Bedarfe der Forschung stetig einbezogen werden. Ein mangelhafter Anwendungsbereich und brüchige Definition fördern die bestehende Unsicherheit; entsprechend klar und eindeutig müssen die Regelungen des Gesetzes insgesamt sein. Eine Durchsetzbarkeit ist sowohl gegenüber der Wissenschaft als auch der Wirtschaft unter Respekt der Forschungsfreiheit (Art. 5 Abs. 3 GG) gewünscht.

Abseits unmittelbarer rechtlicher Hürden, die mit einem FDG unmittelbar gelöst werden könnten, stellt die Vielzahl an konkurrierenden und bereichsspezifischen Regelungen für Forschungsdaten ein Hemmnis dar. Archivrecht, Sozialrecht, Informationsfreiheitsgesetze, Datenschutzrecht, Urheberrecht usw. überlagern und überformen einander, weshalb sowohl Forschenden als auch Rechtswissenschaftler:innen ein Verständnis nur gemeinsam und mit großem Aufwand bewerkstelligen können. Die Verständnishürde sollte deutlich herabgesetzt werden.

In **technischer Hinsicht** ergeben sich Hindernisse besonders aus dem Verständnis der Aufbereitung von Forschungsdaten als inter- und multidisziplinäre Aufgabe. Technische Grundvoraussetzung dafür ist eine Interoperabilität von Datenformaten und Forschungsergebnissen – eine Interoperabilität muss technisch, semantisch, syntaktisch und letztlich auch organisatorisch verstanden werden. Hinzu tritt für Semantik und Syntax eine Möglichkeit, ein Verständnis über die eigene Disziplin hinaus zu fördern. Auch fehlt es an einem Bewusstsein, dass es für eine nachhaltige Datenaufbereitung einer entsprechend nachvollziehbaren Dokumentation der Forschungstätigkeit mangelt; hier bedarf es der Sensibilisierung. Es fehlen also entsprechende bzw. ausreichend dokumentierte Standards, die eine einheitliche Ausgestaltung begünstigen. Eine gesetzliche Lösung sollte dabei aber bereichsspezifische Eigenheiten wie z.B. bestehende Regelungen der Verwaltungsdigitalisierung berücksichtigen. Zugleich darf die Einheitlichkeit nicht dazu führen, dass per Gesetz ein “single point of failure” entsteht.

Technische und rechtliche Hürden münden letztlich in **organisatorischen Hemmnissen**. Fehlende Standards und Schnittstellen, aber auch rechtliche Unsicherheiten zeigen sich in fehlenden dezidierten Datenrepositorien für Forschungsdaten. Es fehlen vertrauenswürdige Datenmittler bzw. Datentreuhänder als Transferdienstleistende. Zugleich braucht es dezentrale Repositorien, die eine Orientierung bieten und bei der Überwindung der aufgezeigten Hindernisse eine Unterstützung bieten.

3 Lösungsvorschläge

Aus Sicht der NFDI empfiehlt sich eine überwiegend **legislative Lösung** zur Bewältigung der o.g. Hindernisse, um Vertrauen und Sicherheit bei den Forschenden und Forschungseinrichtungen zu generieren. Durch das FDG auf Bundesebene könnte ein einheitlicher Rahmen geschaffen werden, der den Ländern eine maßgebliche Orientierungswirkung ist. Unstimmigkeiten in Auslegung und Verständnis bestehender Landesregelungen könnten ggf. geklärt werden.

Konkrete legislative Änderungen und Ergänzungen ergeben sich aus Sicht der NFDI wie folgt:

- Definition des Forschungsdatenbegriffs: Übernahme der umfassenden Definition in § 3 Abs. 10 Datennutzungsgesetz sowie Art. 2 Abs. 9 der Richtlinie (EU) 2019/1024 (sog. PSI- bzw. Open-Data-Richtlinie). Inhaltlich sollte die finale Definition allerdings auch (Zwischen-)Produkte und nicht-digitale Fragmente wissenschaftlicher Forschung einbeziehen. Ebenso sind die Begriffe „Forschungsdaten“, „Forschungsdatenmanagement“ und „Forschungsdateninformationsmanagement“ in Anlehnung an den Begriff trennscharf voneinander zu definieren:
 - (genuine) Forschungsdaten: Daten, die im wissenschaftlichen Prozess oder Forschungsprojekt erhoben werden. Sie weisen unterschiedliche „Reifegrade“ auf, z.B. unverarbeitete Primärdaten, prozessierte Sekundärdaten und geprüfte, publikationsfähige bzw. lizenzierbare Daten.
 - Forschungsdatenmanagement (FDM): Alle lebenszyklus-abhängigen Aufgaben im Forschungsprozess. Hierzu gehören bspw. Kuratierungs- und Datenmanagementprozesse sowie die Anwendung der dazu notwendigen Informationstechnologien.
 - Forschungsdateninformationsmanagement: Vorgänge in einer Forschungsorganisation, Daten bzw. Informationen über ihre Forschungstätigkeiten zu erheben.
- Klarstellung der Rolle der Forschungsdatenzentren: Mit Perspektive der NFDI und der Langzeitarchivierung durch spezielle Forschungsdatenzentren muss geklärt werden, welche Position Forschungsdatenzentren einnehmen und wie diese sich von Archiven i.S.d. Archivrechts unterscheidet. Sofern sie als Nachweissysteme für Forschungsleistungen verstanden werden, bedarf es dafür umfangreicher Rahmenregelungen.

- Rechtliche Rahmenbedingung zur Anerkennung einer „Datenautorenschaft“: Die FAIR-Prinzipien im Blick sollte damit die Zitierbarkeit rein digitaler Forschungsergebnisse erleichtert und gesichert werden. Ausdrücklich nicht gemeint, da rechtlich nicht umsetzbar, ist die Autorenschaft an reinen Messdaten wie beispielsweise Altaraußmaßen oder Abständen von Molekülbestandteilen; hierzu sei auf die Differenzierung zwischen Meinung und Tatsachen in der Rechtsprechung des BVerfG verwiesen.¹²
- Konkretisierung von möglichen Qualitätskriterien für Forschungsdaten: Als Voraussetzung geeignete bundeseinheitlicher Vorgaben qualitativer Grundsätze (z.B. FAIR) und ethischer Maßstäbe (vgl. § 6 HessUnivG a.F. – heute § 1 Abs. 3 HessHG¹³). Diese sollten jedoch ausreichend Spielraum für die heterogenen Wissenschaftsdisziplinen lassen, die Kriterien modular anzuwenden oder gar eigene in einem vorgegebenen Prozess zu etablieren und anerkennen zu lassen.
- Konkretisierung der Dauerhaftigkeit einer Datenaufbewahrung: Umfang und Detailgrad der aufzubewahrenden Forschungsdaten anhand objektiver Kriterien; ggf. in Anlehnung an Kriterien des Archivrechts zur Bestimmung von Archivgut.
- Durchsetzbare Sanktionen bei widerrechtlichem Zugang Dritter zu unveröffentlichten Forschungsdaten, beispielsweise vor Ablauf vereinbarter oder gesetzlicher Fristen.
- Gesetzlich verankerte Verschwiegenheitspflicht von Forschungsdatenzentren bei Forderungen zur Offenlegung/Veröffentlichung von sensiblem Archivgut.

Auch wenn der Fokus bei der Stellungnahme zum FDG auf legislativen Lösungsansätzen liegt, gibt es auch Lösungen, die außerhalb eines Forschungsdatengesetzes der rein rechtlichen Sphäre liegen. **Die nicht-legislativen Lösungsansätze tangieren und ergänzen die genannten legislativen Anpassungsvorschläge:**

Den Anwendungsbereich des FDG betreffend bleibt zu bedenken, dass die Langzeitarchivierung und Zugänglichkeit/Nachnutzbarkeit von Forschungsdatensätzen stets von den zugreifenden Systemen und Anwendungen abhängt. Dies macht erforderlich, dass durch Förderlinien und Gesetze nicht nur die Neuentwicklung von Anwendungen und Practices, sondern auch bestehende Best Practices verstetigt und gepflegt werden. Entsprechend sollte das FDG, wenn nicht schon legislativ, die Forschungsdaten umgebende Landschaft von Soft- und Hardware, Ontologien sowie Grundlagen zur (Nach-)Nutzung von Forschungsdaten, in die Förderung von Forschungsdaten-Vorhaben einbeziehen. Es ist damit zu rechnen, dass mit der steigenden Archivierung von Forschungsdaten auch der Bedarf an geeigneten und verzerrungsfreien Abruf- oder Zugriffsmethoden steigt.

¹² BVerfGE 90, 241 (247) sowie E 33, 1 (14).

¹³ Hierzu ausführlich *Vettermann/Petri*, „Should I CARE about FAIR? – Ein rechtlicher Blick auf die Prinzipien des Forschungsdatenmanagements“, RuZ 1/2023 – im Erscheinen.

Die Sensibilisierung sowie Begleitung der Forscher:innen durch Datenexpert:innen z.B. in Datenkompetenzzentren oder durch Helpdesks erscheinen als wichtige flankierende Maßnahmen zur Stärkung von Datenkompetenz. Auch darüber hinaus sollte die berufliche Auseinandersetzung mit Forschungsdaten als Schlüsselkompetenz begriffen werden, indem Voraussetzungen für (zukünftige) forschungsdatenbezogene Berufsbilder geschaffen werden. Die gewonnenen Erfahrungen im rechtlichen wie infrastrukturellen Umgang mit Forschungsdaten entlang des Lebenszyklus sind wertvoll und notwendig für eine Effektuierung des FDG.

4 Sonstiges

Zunehmende Relevanz bekommt die Nachhaltigkeit von Forschungsdaten aufgrund der Begrenztheit der Ressourcen. Ausschlaggebend für eine nachhaltige und offene Datenökonomie ist die Förderung geeigneter, möglichst kostenfreier Repositorien für Forschungsdaten. Damit einhergehen muss aber ein bewusstes Aufbereiten und Erfassen der Forschungsdaten im Sinne einer bestmöglichen Nachnutzbarkeit – hier sei auf die o.g. Sensibilisierungsmaßnahmen verwiesen. Zudem ist der Aufbau von nachhaltigen Informationsinfrastrukturen essentiell für konsistente Datenzugänge und vertrauensvollen Datenaustausch.

Ergänzend zu den vorhandenen Archivgesetzen sollte das FDG darüber hinaus Kooperationsmodelle zwischen den jeweiligen Forschungsdaten- und Rechenzentren sowie den Gedächtniseinrichtungen ermöglichen. Auf diese Weise ließen sich technische Kompetenzen bei der Datenspeicherung und -vorhaltung mit denen der Archive bezüglich der Festlegung des bleibenden Werts von Forschungsdaten, ihrer Erschließung und den Anforderungen an die digitale Archivierung zusammenführen. Für die Bestimmung des bleibenden Werts kann auch die Einbeziehung der datenproduzierenden Stelle sinnvoll oder gar notwendig sein.

Impressum

Nationale Forschungsdateninfrastruktur (NFDI) e.V.

Albert-Nestler-Str. 13

76131 Karlsruhe

Tel: + 49 721 988 994 0

Fax: + 49 721 988 994 29

E-Mail: info@nfdi.de

<https://www.nfdi.de/>

Ansprechpersonen

Vorstand des NFDI-Vereins:

Prof. Dr. York Sure-Vetter (Direktor)

Eva Lübke (Kaufmännische Leiterin)

Sprecher:in der ELSA-Sektion:

Prof. Dr. Franziska Boehm

Prof. Dr. Ulrich Sax

Das Papier spiegelt die Gesamtheit der Positionen der NFDI-Konsortien wider. Koordination und Federführung durch die Sektion Ethical, Legal und Social Activities (ELSA), namentlich Vasilka Stoilova und Dr. Oliver Vettermann.