

Datendienste nachhaltig gestalten

Ein Diskussionsimpuls zur Weiterentwicklung von Forschungsdateninfrastrukturen

Rat für Informationsinfrastrukturen, Göttingen, Mai 2020

Der Rat für Informationsinfrastrukturen (RfII) hat in den Jahren 2017 und 2018 Diskussionsimpulse zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) publiziert.¹ Mit der Bewilligung der ersten Konsortien werden ab 2020 Weichen für Datendienste gestellt, die dann in der Verantwortung der NFDI sowie von Bund und Ländern stehen werden. Aber auch außerhalb beziehungsweise parallel zur NFDI werden in Deutschland und in Europa Informationsinfrastrukturen weiterentwickelt und aufgebaut, unter anderem die *European Open Science Cloud* (EOSC) mit ihren transnationalen Diensten und Angeboten.

Die Gemeinsame Wissenschaftskonferenz (GWK) hat den RfII in der Rolle eines Beraters für Politik und Wissenschaft eingesetzt. Vor diesem Hintergrund wird er die Entwicklung im Bereich der Forschungsdateninfrastrukturen auch weiterhin durch Diskussionsimpulse kompakt und zeitnah unterstützen. Dabei ist ihm eine gut ausbalancierte Entwicklung wichtig, die sowohl die Überschneidung von Bedarfen von Wissenschaft, öffentlicher Verwaltung und Wirtschaft berücksichtigt als auch die internationale Zusammenarbeit.

Nachhaltiges Nutzen für die Wissenschaft organisieren

Forschungsdatendienste in der öffentlichen Wissenschaft müssen darauf angelegt sein, eine wissenschaftsweite Nutzung in möglichst nachhaltiger Form zu gewährleisten. „Nutzung“ von Daten geht dabei zunächst nicht mit einer „Verwertung“ im kommerziellen Sinne einher.² Diese Zielstellung drückt sich zum Beispiel im breit akzeptierten Leitmotiv der Offenheit und Zugänglichkeit von Wissenschaft aus (mit einem breiten Spektrum an Programmatiken wie z. B. Open Access, Open Science oder FAIR). In Deutschland soll die NFDI vor eben diesem Hintergrund durch Communities beziehungsweise Fachgemeinschaften ausgestaltet werden, um sich dann, angetrieben durch die fach-

liche Nachfrage, als wissenschaftseigenes Netzwerk von Forschungsdatendiensten zu entwickeln. Darüber hinaus entstehen mit öffentlicher Förderung weitere Datensammlungen und -dienste für die Wissenschaft, die es anschlussfähig zu gestalten gilt.

Auf dem Weg zu guten Lösungen ist eine Vielzahl von Gestaltungsalternativen denkbar. Dies betrifft neben der Qualitätssicherung auch die „Betriebsmodelle“ und hierauf aufbauende „Geschäftsmodelle“ (in einem nicht profitorientierten, aber auf Nachhaltigkeit und Ressourcenschonung angelegten Wortsinn), die Verteilung von Aufgaben und Verantwortlichkeiten sowie die Bereitstellungs- und Nutzungsformen (mit ggf. Regeln für eine verantwortungsvolle Bereitstellung und Nutzung, Aneignungsverboten etc.). Der RfII formuliert vor diesem Hintergrund einige aus wissenschaftlicher Perspektive gebotene Anforderungen an Daten und Datendienste.

Gestaltungsmerkmale für Daten(sammlungen) und Datendienste

Zielgruppenorientierung. Die Angebote müssen zunächst die Bedarfe ihrer (wissenschaftlichen) Zielgruppen reflektieren. Aushandlungsprozesse hierzu inklusiv und gerecht zu führen, gehört zu den grundlegenden Anforderungen zum Beispiel an die Konsortien der NFDI, aber auch an transnationale Strukturen wie die EOSC oder andere Forschungsdateninfrastrukturen.³ Während des Aufbaus sollte dabei die aktuell existierende wissenschaftliche Nachfrage, und nicht eine lediglich antizipierte (mögliche oder denkbare) künftige Nachfrage im Vordergrund stehen.

Standortneutrale Zugangsregime. Große nationale Forschungsdateninfrastrukturen wie die NFDI sind als überregionale, transdisziplinäre Netzwerke angelegt, mit dem Ziel eines gleichermaßen guten Zugangs für Wissenschaftlerinnen und Wissenschaftler unabhängig vom jeweiligen Standort. Sofern der Zugang für Nutzende an Voraussetzungen geknüpft ist (etwa an

¹ Vgl. RfII (2018) – Diskussionspapier: In der Breite und forschungsnah; RfII (2018) – Positionspapier Zusammenarbeit als Chance; RfII (2017) – Schritt für Schritt oder: Was bringt wer mit?

² Nutzungen durch weitere Akteure im außerwissenschaftlichen Kontext sind nicht grundsätzlich ausgeschlossen, stehen aber bei der innerwissenschaftlichen Organisation von Daten meist nicht im Vordergrund.

³ Vgl. zur Nutzereinbindung ausführlich: RfII (2018) – Diskussionspapier: In der Breite und forschungsnah, mit Blick auf europäische Gestaltungsansätze siehe RfII (2018) – Stellungnahme zur EOSC.

die Einwilligung in Datenschutzbestimmungen oder spezifische methodische Kompetenzen), sollten interne Regelungen sicherstellen, dass es keine regionalen oder organisationsbezogenen Bevorzugen innerhalb und zwischen den Akteuren gibt, die das Netzwerk tragen. Wo Forschungsdateninfrastrukturen zunächst für den nationalen Bedarf konzipiert werden, müssen Anforderungen an die Anschlussfähigkeit zu supra- und internationalen Wissenschaftsstrukturen frühzeitig mitgedacht und mitgestaltet werden. Insbesondere die Kooperationsfähigkeit im europäischen Forschungsraum gilt es sicherzustellen.

Wissenschaftsweite Teilhabe. Daten und Datendienste sollten auch in fachlicher Hinsicht wissenschaftsweit und transdisziplinär nutzbar sein. Die in fachübergreifende Infrastrukturen wie zum Beispiel die NFDI ‚mitgebrachten‘ Metadaten müssen dort, wo dies möglich und sinnvoll ist, in gemeinschaftlich vereinbarte und idealerweise international anerkannte Ontologien überführt werden und sollten für den Einsatz maschineller Verfahren geeignet sein. Neben einer solchen Fortentwicklung der Aus- und Kennzeichnung von Forschungsdaten sollten auch Pfade für die Entwicklung transformierter Datensätze oder „Datenprodukte“ mitgedacht werden, die Nutzungsmöglichkeiten für andere Disziplinen und Domänen schaffen.⁴

Wissenschaftsadäquate Nutzungsbedingungen. Datendienste für die Forschung zeichnen sich durch die Qualitätsmerkmale der Nachvollziehbarkeit, der Beständigkeit und der Unabhängigkeit (von wissenschaftsfremden Interessen) aus. Nutzende sollten daher anbieterseitig nicht nur bestmögliche Informationen zur Provenienz und wissenschaftlichen Qualität von Daten erhalten, sondern es müssen bei Bedarf auch Langzeitverfügbarkeit von Inhalten, Zugangs- und Nutzungsrechten gewährleistet sein.⁵

Nutzung außerhalb der Wissenschaft. Wissenschaftseigene Angebote sollten grundsätzlich auch für Nachfragen aus Wirtschaft und Gesellschaft zugänglich sein. Gleichwohl können die Nutzungsinteressen und die Bedingungen der Arbeit mit den Daten dann andere sein als beim Teilen von Daten innerhalb der Wissenschaft. Aspekte der Datennutzung an der

Schnittstelle zu Wirtschaft und Gesellschaft sollten deshalb beim Aufbau neuer Dienste konzeptionell und schon frühzeitig berücksichtigt werden. Bereits praktizierte und erprobte Modelle sowie Erfahrungswerte aus der Breite der beteiligten wissenschaftlichen Communities und Fachgemeinschaften können hierfür eine Grundlage bieten.

Nutzung und qualitative Erweiterung von Daten im Forschungsprozess. Datendienste sollen als eine gemeinsame Ressource den Fortschritt von Forschung steigern können. Deshalb ist es erstrebenswert, dass Forschungsergebnisse, die beispielsweise mit Daten und Datenprodukten aus wissenschaftlichen Beständen erzielt wurden – etwa erweiterte oder transformierte Daten – auch wieder in den Bestand zurückgeführt werden und so zur Erweiterung und Qualität des Datenpools beitragen. Um dieses Ziel zu erreichen, sollten die Betreiber der wissenschaftlichen Datendienste geeignete Vereinbarungen mit den Datennachfragern und -nutzern treffen.⁶

Wissenschaftliche Reputation. Dateninfrastrukturen und ihre Dienste bringen die Forschung in dem Maße entscheidend voran, wo (und wie gut) sie die Qualität von Forschungsdaten, Methodik und Forschungsergebnissen sichern. Neben den auf diese Weise zu erzielenden Reputationsgewinnen sollten auch die Vorbereitung von Daten zur wissenschaftlichen Nutzung und die Entwicklung sowie der Betrieb von Datendiensten in geeigneter Weise reputationsförderlich sein.⁷ Die im Aufbau befindliche NFDI kann hierzu einen wesentlichen Beitrag leisten und auch entsprechende Impulse in den europäischen Forschungsraum geben.

Kompetenzaufbau. Nach wie vor beobachtet der Rfll starke Engpässe und fehlende Qualifizierungen für Fachkräfte an der Schnittstelle Wissenschaft/ Informationsinfrastruktur. Forschungsinfrastrukturen kommt hier eine besondere Rolle zu.⁸ Neben der Vermittlung von fachspezifischen, technischen und Informatikbezogenen Qualifikationen werden im Bereich der Koordinierung und Steuerung von Infrastrukturverbänden und Datendiensten auch wirtschaftliche und juristische Kompetenzen erworben, zum Beispiel bei der Ausarbeitung von Betriebs- und

4 Eine domänenübergreifende Erschließung von Daten kann mit Aufwand verbunden sein, vgl. Rfll (2019) – Herausforderung Datenqualität, S. 75f., und dort zitierte Literatur. Dies gehört zu den Gründen, warum die Idee einer „offenen Bereitstellung“ im konkreten Fall fachliche Anstrengungen erfordern kann, um die erwünschte Nachnutzung zu gewährleisten (vgl. Rfll (2019) – Stellungnahme aktuelle Entwicklungen Open Data).

5 Forschungsdateninfrastrukturen bzw. ihre Betreiber sollten insbesondere Aussagen zur Betriebskontinuität der Dienste treffen. Nach Recherchen des Rfll befinden sich Nutzungsbedingungen und Dienstverträge im Bereich wissenschaftlicher Daten bislang oftmals nicht auf dem Stand von Regelungen, die z. B. für Zugänge im Bereich der digitalen wissenschaftlichen Literatur üblich sind.

6 Siehe hierzu auch: Rfll (2018) – Diskussionspapier: In der Breite und forschungsnah, S. 5.

7 Der Rfll hat sich zur Notwendigkeit einer Steigerung der Reputation von wissenschaftlicher Datenarbeit und der Erstellung wissenschaftlicher Datenprodukte in seinem Positionspapier „Herausforderung Datenqualität“ geäußert, Rfll (2019) – Herausforderung Datenqualität.

8 Rfll (2019) – Digitale Kompetenzen.

Geschäftsmodellen oder der Vertragsgestaltung mit kommerziellen Dienstleistern beziehungsweise Anbietern. Insbesondere die Konsortien der NFDI sollten sich in dieser Hinsicht auch als Treiber für die Heranbildung von Fachkräften begreifen.

Technisch-organisatorische Anschlussfähigkeit. Zum Umgang mit und zur Nutzung von Referenzmodellen sollte unter den Betreibern wissenschaftlicher Datendienste ein breiter Diskurs begonnen werden.⁹ Dies gilt in besonderem Maße für die NFDI-Konsortien, die auch existierende respektive parallel entstehende wissenschaftliche Informations- beziehungsweise IT-Infrastrukturen sowie hierauf zugeschnittene Betriebsmodelle im Blick behalten müssen: etwa auf lokaler Ebene der Universitäten und Forschungseinrichtungen, aber auch im Bereich des Hoch- und Höchstleistungsrechnens und der EOSC. Ein Zusammenspiel mit Bestrebungen in den Sektoren Wirtschaft und Verwaltung, Datenbestände zu vernetzen, wäre darüber hinaus förderlich.

Reform- und Kompromissbereitschaft. Zur Bildung fach- und domänenübergreifender Forschungsdateninfrastrukturen muss es zu Abstimmungen und gemeinsamen Entscheidungen zwischen den Akteuren kommen, die technische Systeme, verwendete Standards, gemeinsam zu nutzende Dienste, die Übernahme von Kosten oder den Austausch von Know-how und gegebenenfalls Ressourcen betreffen. Dies kann in Teilbereichen bedeuten, dass man etablierte Verfahren aufgeben oder anpassen muss. Hier sind zugunsten übergeordneter Ziele und der Finanzierbarkeit des Gesamtsystems Kompromissfähigkeit und eine Kommunikationskultur geboten, die Partikularinteressen überwinden hilft.

Prinzip der besten Eignung. Die oftmals gemeinschaftliche Organisation von Diensten bringt Arbeitsbeziehungsweise Aufgabenteilungen mit sich, die sowohl zwischen den innerwissenschaftlichen Verbundpartnern auszuhandeln sind, als auch mit externen, wissenschaftlichen sowie kommerziellen Akteuren und ihren zu integrierenden Angeboten. Handlungsleitendes Prinzip sollte dabei – nebst der Wahrung der Souveränität der Wissenschaft – die Übertragung einer Aufgabe an den jeweils fachlich-technisch am besten geeigneten Akteur sein. Grundlage ist eine klare Beschreibung der Aufgaben und ein Selbstverständnis darüber, welche Verantwortlichkeiten in der Hand der

Wissenschaft verbleiben und welche Aufgaben in Zusammenarbeit mit externen Dienstleistern oder Partnern gut, international anschlussfähig und vor allem wissenschaftskonform gelöst werden können.

„Bezahlen mit Daten“. Beim Aufbau eines Dienstleistungsportfolios werden Entscheidungen zur Einbindung von Drittanbietern getroffen werden müssen. Hierbei ist Sorge zu tragen, dass Geschäftsmodelle Dritter, die Daten und Datenpuren aus wissenschaftlichen Arbeitsprozessen beziehungsweise Zwischenergebnisse aus der Forschung aufzeichnen und kommerziell vermarkten, nicht zulasten der Rechte der einzelnen Wissenschaftlerinnen und Wissenschaftler, der Forschung insgesamt oder der Ziele von wissenschaftlichen Einrichtungen und Infrastrukturverbänden gehen.¹⁰ Im Einzelfall stellen sich Grenzfragen, die wissenschaftsintern noch nicht hinreichend ausgeleuchtet oder gar ausverhandelt sind. Hierauf abzielende Verständigungsprozesse sollten im Gesamtsystem von den Verantwortlichen möglichst transparent und auch kritisch geführt werden.

Datensouveränität der Forschenden. Wo öffentlich geförderte wissenschaftliche Akteure beziehungsweise von ihnen gebildete Verbundstrukturen oder Einrichtungen der Infrastruktur Verträge mit (kommerziellen) Intermediären schließen, ist die Wahrung der Datensouveränität der Forschenden sowie der Wissenschaft und ihrer Institutionen insgesamt eine wichtige Anforderung an die Vertragsgestaltung, auch im Detail. Entsprechende Verträge mit wirtschaftlichen Partnern sollten öffentlich zugänglich sein und in einer Weise ausgestaltet werden, die einer für die Wissenschaft nachteiligen Monopolbildung entgegenwirkt.¹¹ |

Rat für Informationsinfrastrukturen (RfII)

Geschäftsstelle
Papendiek 16
37073 Göttingen
0551-392 70 50
info@rfii.de www.rfii.de

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung – keine Bearbeitung 4.0 Lizenz (CC BY-ND). Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

⁹ Zu nennen wäre etwa die ausdifferenzierte Referenzarchitektur für Datenplattformen der International Data Spaces Association (Whitepaper 2019) oder das OAIS Referenzmodell für die Langzeitarchivierung. Auch Reifegradmodelle für Forschungsinfrastrukturen (z. B. RISE, RISE-DE) sowie Standardklauseln für wissenschaftsadäquate Nutzungsbedingungen erscheinen sinnvoll.

¹⁰ Zu „Daten als Gegenleistung“ vgl. auch Datenethikkommission (2019) – Gutachten der Datenethikkommission.

¹¹ So wie es auch die Eckpunkte einer Datenstrategie der Bundesregierung vorsehen. Siehe unter <https://www.bundesregierung.de/re-source/blob/975226/1693626/60b196d5861f71cdefb9e254f5382a62/2019-11-18-pdf-datenstrategie-data.pdf?download=1> (zuletzt geprüft am 04.05.2020).