



Nationale Forschungsdateninfrastruktur (NFDI)

Nathalie Hartl¹ · Elena Wössner · York Sure-Vetter

Angenommen: 30. Juli 2021 / Online publiziert: 26. August 2021
© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2021

Zusammenfassung

In der Nationalen Forschungsdateninfrastruktur (NFDI) werden wertvolle Forschungsdaten für das gesamte deutsche Wissenschaftssystem systematisch erschlossen, vernetzt und nachhaltig nutzbar gemacht. Bislang sind diese meist dezentral, projektbezogen oder nur zeitlich begrenzt verfügbar. Mit der NFDI soll ein digitaler Wissensspeicher unter Berücksichtigung der FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable) geschaffen werden. Bereits vorhandene Daten können zur Bearbeitung weiterer Forschungsfragen genutzt werden und neue Erkenntnisse und Innovationen ermöglichen. Bis zu 30 NFDI-Konsortien, Zusammenschlüsse verschiedener Einrichtungen innerhalb eines Forschungsfeldes, arbeiten zusammen interdisziplinär an der Zielumsetzung. Um die Aktivitäten zum Aufbau einer Nationalen Forschungsdateninfrastruktur zu koordinieren, wurde der gemeinnützige Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. mit Sitz in Karlsruhe im Oktober 2020 gegründet. Gemeinsam gestalten Verein und NFDI-Konsortien die Zukunft des Forschungsdatenmanagements in Deutschland. Darüber hinaus soll NFDI auch am Aufbau internationaler Initiativen, beispielsweise der European Open Science Cloud (EOSC), mitwirken.

Täglich wächst die Menge an wertvollen Forschungsdaten. Doch große Teile davon sind aktuell nur für eingeschränkte Personenkreise sowie zeitlich begrenzt verfügbar. Forschungsdaten werden häufig auf einzelnen Servern oder Rechnern in wenig standardisierten Formaten gespeichert. Eine sinnvolle Archivierung mit Möglichkeit zur Nachnutzung ist in die wenigsten Forschungsprozesse standardmäßig integriert, sodass Zugänge zu Wissen verlorengehen. Um dies zu ändern, ist die Nationale Forschungsdateninfrastruktur (NFDI) von Bund und Ländern ins Leben gerufen worden.

Ziel von NFDI ist es, ein systematisches sowie nachhaltiges Forschungsdatenmanagement aufzubauen, zu koordinieren und dabei bereits bestehende Strukturen besser zu vernetzen. Relevante Daten sollen nicht länger ungenutzt auf den Servern einzelner Einrichtungen ohne zentralen Zugang schlummern, sondern nach den FAIR-Prinzipien [1] (Findable, Accessible, Interoperable und Reusable – also auffindbar, zugänglich, interoperabel und wiederverwendbar) zur Verfügung gestellt werden.

Ein verbesserter Zugang zu digitalen Datensätzen soll den Wissenschaftsstandort Deutschland stärken, den Austausch verschiedener Forschungseinheiten sowie -disziplinen fördern und zudem das Forschungsdatenmanagement auf internationaler Ebene vereinfachen [2]. Um Potenziale fächerübergreifend ausschöpfen zu können und gleichzeitig Bedarfe unterschiedlicher Communitys zu berücksichtigen, werden bis zu 30 NFDI-Konsortien durch Bund und Länder gefördert, die jeweils verschiedene Forschungsbereiche vertreten. Um die Aktivitäten zum Aufbau einer Nationalen Forschungsdateninfrastruktur fachlich zu steuern und zu koordinieren wurde der gemeinnützige Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. mit Sitz in Karlsruhe gegründet. Gemeinsam gestalten die NFDI-Konsortien und der NFDI-Verein die Zukunft des Forschungsdatenmanagements in Deutschland.

Von der Notwendigkeit eines besseren Forschungsdatenmanagements zur Vereinsgründung

Die Empfehlung, eine koordinierte Forschungsdateninfrastruktur in Deutschland aufzubauen, kam vom Rat für Informationsinfrastrukturen (RfII). Mit einer Nationalen Forschungsdateninfrastruktur in Form eines bundesweiten, verteilten, wachsenden Netzwerks soll ein „neues Rückgrat für

✉ Nathalie Hartl
info@nfdi.de

¹ Nationale Forschungsdateninfrastruktur (NFDI) e.V.,
Karlsruhe, Deutschland

das Forschungsdatenmanagement“ in Deutschland geschaffen werden [3]. Dahinter steht die Vision, Forschung disziplinübergreifend effektiver zu gestalten. Denn immer mehr wissenschaftliche Entdeckungen werden heute mit bereits vorhandenen Forschungsdaten gemacht. Diese können z. B. unter neuen Fragestellungen betrachtet werden oder mit anderen Datensätzen verknüpft werden. Ein gut zugänglicher digitaler Wissensspeicher erlaubt neuen Erkenntnisgewinn und befördert Innovationen, die der Gesellschaft nutzen. Bund und Länder haben sich dem Vorschlag des RfII angeschlossen und im November 2018 eine Vereinbarung zum Aufbau und zur Förderung einer NFDI getroffen [4].

Fachgetriebene Vernetzung als Schlüssel zum Erfolg

In der Bund-Länder-Vereinbarung wurde festgelegt, insgesamt bis zu 30 Konsortien zu fördern, die in 3 Ausschreibungsrunden in einem wissenschaftsgeleiteten Verfahren ausgewählt werden. Die Deutsche Forschungsgemeinschaft (DFG) koordiniert das Verfahren zur Begutachtung der Bewerbungen und ein Expertengremium prüft und bewertet die Anträge [5]. Die Gemeinsame Wissenschaftskonferenz (GWK) entscheidet auf dieser Basis letztlich, welche Konsortien gefördert werden. Bei der Auswahl der Konsortien wird darauf Wert gelegt, die Wissenschaftslandschaft möglichst breit abzubilden. Das spiegelt sich auch bei der Betrachtung der bereits geförderten Konsortien aus den ersten beiden Förderrunden wider. Die 9 Konsortien DataPLANT, GHGA, KonsortSWD, NFDI4BioDiversity, NFDI4Cat, NFDI4Chem, NFDI4Culture, NFDI4Health und NFDI4Ing sind im Oktober 2020 gestartet. Im Juli 2021 haben weitere 10 Konsortien, BERD@NFDI, DAPHNE4NFDI, FAIRmat, MaRDI, NFDI4DataScience, NFDI4Earth, NFDI4Microbiota, NFDI-MatWerk, PUNCH4NFDI und Text+ ihre Förderzusagen erhalten [6]. Damit arbeiten verschiedenste Bereiche – von Kultur-, über Sozial-, Geistes- und Ingenieurwissenschaften bis hin zu Lebens- und Naturwissenschaften – an gemeinsamen Lösungen für individuelle Probleme.

Der konsortialübergreifende Austausch ist dabei Grundlage für den Erfolg von NFDI. So ergeben sich auch Vernetzungen, die auf den ersten Blick ungewöhnlich erscheinen können, beispielsweise zwischen Chemie und Kulturwissenschaft: Die Untersuchung und Restauration von Kunstwerken kann mittels chemischer Analysen von verwendeten Materialien und den dazu vorliegenden Daten optimiert werden.

Querschnittsthemen, die mehrere Konsortien betreffen, sollen im Verein in sogenannten Sektionen bearbeitet werden. Beispiele für Querschnittsthemen sind die gemeinsame Modellierung von Metadaten-Schemata und Ontologien,

gemeinsame Infrastrukturdienste und Dienstleistungen zum Thema Forschungsdatenmanagement, rechtliche und ethische Fragen rund um das Teilen von Forschungsdaten oder auch die Etablierung der Zitierung von Forschungsdaten in wissenschaftlichen Veröffentlichungen. Durch das gemeinsame Bearbeiten der Querschnittsthemen können sich fächerübergreifende Standards entwickeln. Ziel ist, dass diese Standards durch die Einbindung der unterschiedlichen Communitys eine breite Akzeptanz erfahren.

Auch innerhalb der einzelnen Konsortien findet eine fachliche Vernetzung statt, da hinter jedem Konsortium mehrere Partner, wie beispielsweise Universitäten und Forschungszentren, stehen. Im Konsortium NFDI4Health engagieren sich z. B. 18 (ko-)antragstellende Einrichtungen und 35 weitere Partner. Diese arbeiten unter anderem daran, die Forschung im Rahmen der COVID-19-Pandemie effektiver zu machen [7]. Etabliert werden soll ein Verfahren, mit dem Gesundheitsdaten aus verschiedenen Quellen und Studien miteinander verknüpft werden können (Record Linkage), um ein möglichst umfassendes Bild der Erkrankten zu erhalten. Neue Erkenntnisse zu Infektionswegen und zum Verlauf von COVID-19 können so schneller gewonnen werden. Bei dem Austausch ist es wichtig, die besondere Sensibilität der Daten sowie damit verbundene rechtliche Fragestellungen zu berücksichtigen.

Aber nicht nur im Fall der Coronapandemie kann eine optimierte, gemeinsame Dateninfrastruktur dabei helfen, Sachverhalte besser zu verstehen und daraus Handlungen oder Innovationen abzuleiten. Auch das Thema Klimawandel und die damit verbundenen Auswirkungen lassen sich besser erfassen, wenn viele Daten berücksichtigt werden können. Das Konsortium NFDI4Biodiversity widmet sich beispielsweise dem klimaverursachten Biodiversitätsverlust und möglichen Lösungen. KonsortSWD forscht unter anderem zu den Auswirkungen des Klimawandels auf internationale und nationale Migration, indem historische Klimadaten mit Migrationsdaten verknüpft werden.

Insgesamt zeigen die verschiedenen Beispiele aus den Reihen der Konsortien, welche Potenziale eine nachhaltige Strategie für Forschungsdatenmanagement entfalten kann.

Der NFDI-Verein: Viele Organe – ein Ziel

Das *NFDI-Direktorat* bildet den Vereinsvorstand, koordiniert die Aktivitäten des Vereins, fördert den Austausch und repräsentiert den Verein in der Öffentlichkeit. Die Geschäftsstelle des Direktorats mit mehreren Mitarbeiter:innen ist in Karlsruhe angesiedelt. Daneben gibt es das *Kuratorium* als administrativ-strategisches Kontrollgremium, den *Wissenschaftlichen Senat* als inhaltlich-strategisches Gremium, die *Konsortialversammlung*, welche konsortialübergreifend die inhaltlich-technischen Grund-

sätze bestimmt, sowie die *Mitgliederversammlung*, in der alle Vereinsmitglieder zusammenkommen [8]. Weitere Informationen zu den Vereinsorganen enthält die Satzung.¹

Auf den ersten Blick mag die Struktur von NFDI mit den verschiedenen Vereinsorganen komplex anmuten. Erst die Einbindung einer großen Anzahl an Communitys und Institutionen ermöglicht es jedoch, neue und allgemein akzeptierte Standards zum Forschungsdatenmanagement nach den FAIR-Prinzipien zu formulieren, die künftig nahtlos in Forschungsprozesse integriert werden. Das bedeutet im Detail:

- *Findable*: Es existieren mehr und mehr Datensätze, die jedoch oft im „Big-Data-Dschungel“ untergehen. NFDI möchte eine Art Kompass schaffen, mit dem sich Forschende durch das Datendickicht navigieren können. Metadaten und Persistent Identifier sind dabei die Koordinaten auf der Daten-Dschungel-Landkarte, die dabei helfen, Gesuchtes leichter auffindbar zu machen.
- *Accessible*: Zu viele Datensätze fristen ihr Dasein in einsamen Kellerregalen und auf vergessenen, überholten Speichermedien. Um ihr Potenzial zu nutzen, müssen sie bewahrt und systematisch für Mensch und Maschine zugänglich gemacht werden. Durch digitale Langzeitarchivierung und standardisierte Kommunikationsprotokolle soll die Suche im staubigen Kellerregal durch wenige Klicks am Laptop abgelöst werden. Zugänglich bedeutet hier nicht, dass jede „Kellertür“ immer und für alle offensteht. Ziel ist ein dauerhafter und offener Zugriff auf Metadaten, die vor allem auch Informationen über die Voraussetzungen für den Datenabruf enthalten. Auch bei wem der „Kellerschlüssel“ abgeholt werden kann, soll darin gespeichert sein.
- *Interoperable*: Englisch oder Deutsch, Python oder Java – verschiedene Forschungscommunities nutzen verschiedene Sprachen mit unterschiedlichen Syntaxen, Semantiken und Vokabeln. So auch bei Forschungsdaten. Unterschiedliche Formate, Terminologien und Standards machen einen Austausch und das Zusammenführen von Daten ineffizient. Die Entwicklung gemeinsamer Standards mit kompatiblen Formaten sowie einer interoperablen Darstellung von Metadaten und gemeinsam definierten Terminologien und Ontologien führt zu einer für Menschen verständlichen und für Maschinen verarbeitbaren Sprache. Oder zumindest zu einer geeigneten Übersetzungsmöglichkeit.
- *Reusable*: Früher mussten Forscher:innen für neue Erkenntnisse oft um die Welt segeln, heute werden immer mehr Entdeckungen vom Laptop aus gemacht – und zwar mit bereits vorhanden Daten. Finden Wissenschaftler:innen interessante Datensätze, ist deren Wiederver-

wendung und Weiterverarbeitung häufig weder effizient noch effektiv, da wichtige Informationen zu Kontext, Provenienz (Datenherkunft), Nutzungslizenzen und Zitation nicht unmittelbar vorhanden und schwer bis gar nicht auffindbar sind. Gemeinsam entwickelte Leitfäden zur Ausgestaltung der Metadaten und empfehlenswerten Datenformaten tragen wohl nicht mehr zur Entdeckung eines neuen Kontinents bei, aber vielleicht zur Entdeckung eines neuen Planeten. Oder zum Erhalt unserer Erde und ihrer Vielfalt.

Wandel der Wissenschaftskultur

Wesentliche Ziele von NFDI sind erreicht, wenn sich Forschende bewusst und offen dem Thema Forschungsdatenmanagement zuwenden und Datenkompetenz entwickeln. NFDI kann somit ein Katalysator für den kulturellen Wandel in allen Wissenschaftsdisziplinen sein, der die Entwicklungen hin zum kompetenten Umgang mit Forschungsdaten, nachhaltigen Forschungsdatenmanagement und selbstverständlichen Teilen von Forschungsdaten – unter Beachtung rechtlicher Rahmenbedingungen – beschleunigt.

Eine starke Motivation zum Teilen von Daten kann die Etablierung eines Reputationssystems sein, das Personen ähnlich wie beim Publizieren von Forschungsergebnissen für ihre Leistungen honoriert. Im Rahmen der geplanten Sektion Training & Education sollen mit bedarfsorientierten Aus- und Weiterbildungsangeboten die Kompetenzen von (Nachwuchs-)Wissenschaftler:innen in den Bereichen Data Literacy und FAIR-Konzepten gestärkt sowie das Bewusstsein für Forschungsdatenmanagement erhöht werden.

Internationale Anbindung

Auch auf internationaler Ebene verfolgen Initiativen das Ziel, den Zugang zu wissenschaftlichen Daten in Zukunft zu vereinfachen. Die Europäische Kommission hat dazu die European Open Science Cloud (EOSC) ins Leben gerufen. Angestrebt wird, unter Beteiligung verschiedener Nationen, eine Cloud-Infrastruktur zu schaffen [10]. NFDI wird Deutschland als mandatiertes Mitglied in der EOSC Association repräsentieren und sicherstellen, dass die nationalen Entwicklungen nahtlos an die internationalen anknüpfen.

Eine weitere Dateninfrastruktur-Initiative auf europäischer Ebene ist Gaia-X [11], an der Wirtschaft und Wissenschaft gemeinsam mitwirken. Im Mai 2021 haben NFDI und Gaia-X das Verbundprojekt FAIR Data Spaces gestartet, welches vom Bundesministerium für Bildung und Forschung (BMBF) gefördert wird. Gemeinsam soll ein cloudbasierter Datenraum unter Berücksichtigung der FAIR-Prinzipien kreiert werden. Bislang nicht genutzte

¹ Mehr über den Aufbau erfahren Sie in unserer Satzung: [9].

Potenziale sollen hierbei durch den Wissenstransfer zwischen Wirtschaft und Wissenschaft ausgeschöpft werden [12].

Darüber hinaus sind die einzelnen Konsortien ebenfalls mit internationalen Organisationen vernetzt, beispielsweise das Konsortium GHGA (German Human Genome-Phenome Archive) mit dem europäischen Genomarchiv (EGA) und der Global Alliance for Genomics and Health (GA4GH). Die gemeinsame Bereitstellung von Genomdaten kann unter anderem dabei helfen, dank einer breiteren Datenbasis seltene genetische Erkrankungen besser zu verstehen.

NFDI möchte sowohl deutschlandweit als auch international ein übergreifendes Forschungsdatenmanagement etablieren und Partner zusammenbringen, die gemeinsame Werte und ein hoher Qualitätsanspruch an Daten einen.

Mitglied von NFDI werden

Aktuell hat der NFDI-Verein 142 Mitglieder (Stand 20.07.2021), darunter Ministerien, Universitäten, Hochschulen, Institute und andere Organisationen aus dem Wissenschaftsbereich. Mehr Informationen sowie das Antragsformular zur Mitgliedschaft sind auf der Internetseite des Vereins www.nfdi.de zu finden.

NFDI besser kennenlernen

Der Verein ist sowohl auf den sozialen Netzwerken Twitter @NFDI_de als auch LinkedIn (National Research Data Infrastructure Germany (NFDI)) vertreten. Hier werden aktuelle Neuigkeiten sowie Jobangebote geteilt.

In der Vortragsreihe InfraTalks geben Wissenschaftler:innen aus den Konsortien immer am ersten Montag des Monats Einblicke in ihre Forschung und aktuelle Problemfelder. Übertragen werden die Vorträge auf dem YouTube-Kanal NFDI Direktorat. Hier sind die InfraTalks auch noch nach den Live-Streams abrufbar.² Weitere Informationen finden sich auf der Vereinshomepage.

² Link YouTube-Kanal NFDI Direktorat: [13].

Literatur

1. Wilkinson M et al (2016) The FAIR guiding principles for scientific data management and stewardship. <https://www.nature.com/articles/sdata201618>. Zugegriffen: 16. Juli 2021 (Sci Data 3, Artikelnr. 160018)
2. Bundesministerium für Forschung und Bildung (2021) Nationale Forschungsdateninfrastruktur. <https://www.bmbf.de/de/nationale-forschungsdateninfrastruktur-8299.html>. Zugegriffen: 16. Juli 2021
3. Rat für Informationsinfrastrukturen (2016) Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. <https://rfii.de/download/rfii-empfehlungen-2016/>. Zugegriffen: 16. Juli 2021
4. GWK (2018) Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI). <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf>. Zugegriffen: 16. Juli 2021
5. Deutsche Forschungsgemeinschaft (2021) Informationen und Erklärvideos zu NFDI und dem Begutachtungsprozess bei der Auswahl der Konsortien. <https://www.dfg.de/foerderung/programme/nfdi/>. Zugegriffen: 16. Juli 2021
6. Nationale Forschungsdateninfrastruktur (2021) Übersicht zu den geförderten Konsortien. <https://www.nfdi.de/konsortien-2>. Zugegriffen: 16. Juli 2021
7. NFDI4Health (2021) Task force COVID-19. <https://www.nfdi4health.de/de/task-force-covid-19/>. Zugegriffen: 16. Juli 2021
8. Kraft S, Schmalen A, Seitz-Moskaliuk H, Sure-Vetter Y, Knebes J, Lübke E, Wössner E (2021) Nationale Forschungsdateninfrastruktur (NFDI) e. V.: Aufbau und Ziele. Bausteine Forschungsdatenmanagement, Nr. 2/2021. <https://bausteine-fdm.de/article/view/8332>. Zugegriffen: 17. August 2021
9. Nationale Forschungsdateninfrastruktur (2021) Informationen zum NFDI-Verein. <https://www.nfdi.de/verein>. Zugegriffen: 16. Juli 2021
10. European Commission (2021) European open science cloud. <https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud>. Zugegriffen: 16. Juli 2021
11. Gaia-X (2021) Homepage. <https://www.gaia-x.eu/>. Zugegriffen: 16. Juli 2021
12. Bundesministerium für Bildung und Forschung (2021) Karliczek: Mit FAIR-Data Spaces schaffen wir einen gemeinsamen Datenraum für Wissenschaft und Wirtschaft. <https://www.bmbf.de/de/karliczek-mit-fair-data-spaces-schaffen-wir-einen-gemeinsamen-datenraum-fuer-wissenschaft-14723.html>. Zugegriffen: 16. Juli 2021
13. NFDI Direktorat (2021) YouTube-Kanal. <https://www.youtube.com/channel/UCTz321rUFOvrKOGkFfhyhLQ>. Zugegriffen: 16. Juli 2021