

# Bausteine Forschungsdatenmanagement

## Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern

### NFDI4Health – Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten<sup>i</sup>

Juliane Fluck	Birte Lindstädt	Wolfgang Ahrens	Oya Beyan
Benedikt Buchner	Johannes Darms	Ralf Depping	Jens Dierkes
Holger Fröhlich	Juliane Gehrke	Martin Golebiewski	Linus Grabenhenrich
Horst K. Hahn	Toralf Kirsten	Sebastian Klammt	Harald Kusch
Matthias Löbe	Markus Löffler	Frank Meineke	Wolfgang Müller
Hubertus Neuhausen	Ute Nöthlings	Tobias Pischon	Fabian Prasser
Ulrich Sax	Carsten Oliver Schmidt	Matthias Schulze	Sebastian C. Semler
Sylvia Thun	Dagmar Waltemath	Lothar H. Wieler	Hajo Zeeb

Iris Pigeot

**2021**  
**Zitiervorschlag**

Fluck, Juliane et al. 2021. NFDI4Health – Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2021: S. 72-85. DOI: [10.17192/bfdm.2021.2.8331](https://doi.org/10.17192/bfdm.2021.2.8331).

Dieser Beitrag steht unter einer  
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

<sup>i</sup>Juliane Fluck (ORCID: [0000-0003-1379-7023](https://orcid.org/0000-0003-1379-7023)), Birte Lindstädt (ORCID: [0000-0002-8251-1597](https://orcid.org/0000-0002-8251-1597)),

# 1 Zusammenfassung

Epidemiologische und klinische Studien sind standardisiert und gut dokumentiert, jedoch erfüllen Studienprotokolle, eingesetzte Erhebungsinstrumente und erhobene Daten die Anforderungen der FAIR-Prinzipien nicht in ausreichendem Maße. NFDI4Health wird daher eine Struktur schaffen, die eine zentrale Suche nach existierenden, dezentral verwalteten Datenkörpern und zugehörigen Dokumenten sowie einen FAIRen Zugang zu diesen erleichtert. Dazu werden die Auffindbarkeit und der Zugang zu strukturierten Gesundheitsdaten aus Registern, administrativen Gesundheitsdatenbanken, klinischen und epidemiologischen sowie Public Health-Studien verbessert und die Qualität und Harmonisierung der zugrundeliegenden Daten optimiert.

Eine weitere Herausforderung entsteht durch die Verwendung personenbezogener Gesundheitsdaten. Diese sind hoch sensibel, so dass ihre Nutzung restriktive Datenschutzbestimmungen und informierte Einwilligungserklärungen der StudienteilnehmerInnen erfordert, was jedoch ihre Wiederverwendbarkeit einschränkt. NFDI4Health zielt daher darauf ab, den Austausch und die Verknüpfung von personenbezogenen Gesundheitsdaten sowie verteilte Datenanalysen unter Einhaltung datenschutzrechtlicher und ethischer Bestimmungen zu erleichtern. Um dies möglichst effizient zu erreichen, wird NFDI4Health die Entwicklung neuer, maschinenprozessierbarer Zustimmungsmöglichkeiten sowie innovativer Datenzugriffsservices auf Grundlage der FAIR-Prinzipien vorantreiben und die Interoperabilität von IT-Lösungen für Metadatenrepositorien stärken. Komplementiert wird dies durch die Entwicklung entsprechender Angebote für Training und Ausbildung, um der Herausforderung der Umsetzung der Lösungen in den Universitäten und Forschungseinrichtungen zu begegnen. Schließlich wird durch die gemeinsame Arbeit in der NFDI4Health die Kooperation zwischen klinischer und epidemiologischer/Public Health-Forschung gestärkt.

---

Wolfgang Ahrens (ORCID: [0000-0003-3777-570X](https://orcid.org/0000-0003-3777-570X)), Oya Beyan (ORCID: [0000-0001-7611-3501](https://orcid.org/0000-0001-7611-3501)), Benedikt Buchner (ORCID: [0000-0003-1438-5127](https://orcid.org/0000-0003-1438-5127)), Johannes Darms (Deutsche Zentralbibliothek für Medizin (ZB MED)), Ralf Depping (ORCID: [0000-0003-2537-7229](https://orcid.org/0000-0003-2537-7229)), Jens Dierkes (ORCID: [0000-0002-0121-9261](https://orcid.org/0000-0002-0121-9261)), Holger Fröhlich (ORCID: [0000-0002-5328-1243](https://orcid.org/0000-0002-5328-1243)), Juliane Gehrke (TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung, Berlin), Martin Golebiewski (ORCID: [0000-0002-8683-7084](https://orcid.org/0000-0002-8683-7084)), Linus Grabenhenrich (ORCID: [0000-0002-9300-6625](https://orcid.org/0000-0002-9300-6625)), Horst K. Hahn (ORCID: [0000-0001-7512-5762](https://orcid.org/0000-0001-7512-5762)), Toralf Kirsten (ORCID: [0000-0001-7117-4268](https://orcid.org/0000-0001-7117-4268)), Sebastian Klammt (ORCID: [0000-0001-7852-4769](https://orcid.org/0000-0001-7852-4769)), Harald Kusch (ORCID: [0000-0002-9895-2469](https://orcid.org/0000-0002-9895-2469)), Matthias Löbe (ORCID: [0000-0002-2344-0426](https://orcid.org/0000-0002-2344-0426)), Markus Löffler (ORCID: [0000-0003-2513-1317](https://orcid.org/0000-0003-2513-1317)), Frank Meineke (ORCID: [0000-0002-9256-7543](https://orcid.org/0000-0002-9256-7543)), Wolfgang Müller (ORCID: [0000-0002-4980-3512](https://orcid.org/0000-0002-4980-3512)), Hubertus Neuhausen (ORCID: [0000-0001-5223-9177](https://orcid.org/0000-0001-5223-9177)), Ute Nöthlings (ORCID: [0000-0002-5789-2252](https://orcid.org/0000-0002-5789-2252)), Tobias Pischon (ORCID: [0000-0003-1568-767X](https://orcid.org/0000-0003-1568-767X)), Fabian Prasser (ORCID: [0000-0003-3172-3095](https://orcid.org/0000-0003-3172-3095)), Ulrich Sax (ORCID: [0000-0002-8188-3495](https://orcid.org/0000-0002-8188-3495)), Carsten Oliver Schmidt (ORCID: [0000-0001-5266-9396](https://orcid.org/0000-0001-5266-9396)), Matthias Schulze (ORCID: [0000-0002-0830-5277](https://orcid.org/0000-0002-0830-5277)), Sebastian C. Semler (ORCID: [0000-0001-8547-6785](https://orcid.org/0000-0001-8547-6785)), Sylvia Thun (ORCID: [0000-0002-3346-6806](https://orcid.org/0000-0002-3346-6806)), Dagmar Waltemath (ORCID: [0000-0002-5886-5563](https://orcid.org/0000-0002-5886-5563)), Lothar H. Wieler (ORCID: [0000-0003-4365-3717](https://orcid.org/0000-0003-4365-3717)), Hajo Zeeb (ORCID: [0000-0001-7509-242X](https://orcid.org/0000-0001-7509-242X)), Iris Pigeot (ORCID: [0000-0001-7483-0726](https://orcid.org/0000-0001-7483-0726)).

## 2 Hintergrund und Mission der NFDI4Health

NFDI4Health<sup>1</sup> – die Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten – fokussiert auf Daten, die in klinischen, epidemiologischen und Public Health-Studien generiert werden. Obwohl diese Daten bereits hohen inhaltlichen Qualitätsstandards genügen, erfüllen sie häufig die Anforderungen der FAIR-Prinzipien<sup>2</sup> nicht: (1) Die Auffindbarkeit der Daten ist aufgrund fehlender internationaler Standards für Registrierung und Publikation häufig eingeschränkt. (2) Möglichkeiten der Datennutzung durch Dritte sind in der Regel unklar. (3) Datenbanken sind oft nicht interoperabel, z. B. aufgrund der großen methodischen Heterogenität bei der Erfassung von Expositionen und Gesundheitsfolgen (s. Use Cases, Kapitel 8). (4) Weiterhin schränken Datenschutzbestimmungen und Einwilligungserklärungen für StudienteilnehmerInnen die Wiederverwendbarkeit der Daten ein.

Die Mission der NFDI4Health ist es, neue Möglichkeiten zur wissenschaftlichen Nutzung von personenbezogenen Gesundheitsdaten unter Wahrung des Datenschutzes zu eröffnen. Basierend auf den FAIR-Prinzipien werden den ForscherInnen Dienste zur Verfügung gestellt, Daten in interoperabler Form dauerhaft zu speichern, semantisch anzureichern und zu teilen und Daten aus verschiedenen Quellen zusammenzuführen. Dies steigert die Leistungsfähigkeit und Qualität der Forschung und hilft somit, die Gesundheit der Bevölkerung zu verbessern. NFDI4Health wird damit die nationale und internationale Sichtbarkeit gesundheitswissenschaftlicher Forschung erhöhen, die Reputation von ForscherInnen stärken, die ihre Daten teilen, und neue Kooperationen fördern. Die daraus entstehenden Infrastrukturen schlagen eine Brücke zwischen NutzerInnen und AnbieterInnen von epidemiologischen, Public Health- und klinischen Studiendaten.

NFDI4Health verfolgt einerseits das Ziel, Metadaten, die Studien und ihre Erhebungen zu charakterisieren und zentral zugänglich zu machen. Andererseits sollen personenbezogene Studiendaten in den verschiedenen Repositorien unterschiedlicher Organisationen und Institutionen verbleiben, statt sie in einem NFDI4Health-spezifischen zentralen Repository zusammenzuführen. Für die Auswertung der Daten wird NFDI4Health Infrastrukturen und Algorithmen nutzen und weiterentwickeln, die einer verteilten Datenspeicherung Rechnung tragen und eine verteilte Datenanalyse erlauben. Daher ist die Aufwertung der dort vorzufindenden Schnittstellen und Strukturen von besonderer Relevanz. Das Konsortium hat nach positiver Begutachtung am 1. Oktober 2020 die Arbeit aufgenommen.

<sup>1</sup>NFDI4Health. <https://www.nfdi4health.de>, letzter Zugriff: 26.03.2021.

<sup>2</sup>Wilkinson, Mark D. et al. „The FAIR Guiding Principles for scientific data management and stewardship.“ *Scientific Data* 3 (2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.

### 3 Datentypen im Fokus der NFDI4Health

NFDI4Health fokussiert sich auf Daten, die in klinischen, epidemiologischen und Public Health-Studien generiert werden. Im Vergleich zu klinischen Routinedaten oder Daten aus der ambulanten Versorgung erfüllen diese Daten bereits hohe inhaltliche Qualitätsstandards. Sie werden nach genau definierten Studienprotokollen generiert, sind hoch strukturiert, qualitätsgeprüft und kuratiert. Sie enthalten detaillierte, standardisiert erhobene phänotypische Informationen der StudienteilnehmerInnen, z. B. zu Erkrankungen, zum Ernährungsverhalten oder zum Lebensstil. Damit eignen sie sich zur Untersuchung der Krankheitslast, -geschichte und -prognose, möglicher Krankheitsursachen und der Wirksamkeit von präventiven und therapeutischen Interventionen auf Individual- und Populationsebene. Die Daten, die in klinischen Studien von vielen PatientInnen generiert werden, unterstützen u. a. die Weiterentwicklung der personalisierten Medizin und helfen neue Therapieoptionen zu finden und die PatientInnenversorgung zu verbessern.

Üblicherweise werden diese Datensammlungen in dezentralen, autonomen epidemiologischen oder klinischen Dateninfrastrukturen gespeichert und sollen dort auch verbleiben. Daher wird NFDI4Health ebenfalls die Studienmetadaten zur Beschreibung einer Studie erfassen, um eine zentrale Suche zu ermöglichen. Zu den Studienmetadaten gehören z. B. der Studientitel, die Studienleitung, eine Beschreibung der Studienpopulation und der Zeitraum der Erhebung.

Darüber hinaus wird NFDI4Health es ermöglichen, Studiendokumente wie Protokolle und Standard Operating Procedures (SOPs) sowie Erhebungsinstrumente unabhängig von den Daten zu publizieren, um Studieninhalte zu beschreiben.

### 4 Zielgruppen der NFDI4Health

Persönliche Gesundheitsdaten werden hauptsächlich von MedizinerInnen, EpidemiologInnen, Public Health-ForscherInnen, SozialwissenschaftlerInnen und BiometrikerInnen erhoben und analysiert. Diese ForscherInnen bilden auch einen großen Teil der User Community der NFDI4Health, wobei die Studiendurchführung, die Datenaufbereitung und -nutzung oftmals von derselben Forschungsgruppe ausgeführt wird.

Klinische Studien werden in der Regel von ForscherInnen initiiert, die in der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF; 182 Mitglieder) vertreten sind. Um diese breite Gemeinschaft möglicher DatenlieferantInnen zu erreichen, bindet NFDI4Health zusätzlich die universitären klinischen Studienzentren als zentrale Kommunikationsknotenpunkte ein. Öffentlich geförderte klinische Studien werden häufig von diesen Zentren betreut und unterstützt.

Zusätzlich bezieht NFDI4Health eine große Anzahl von User Communities ein, die weitere Gesundheitsdateninfrastrukturen beisteuern, wie beispielsweise das Zentrum für Krebsregisterdaten. Bereits in der Phase der Antragstellung wurden die adressierten User Communities durch Befragungen und bei einer Diskussion des NFDI4Health-

Konzepts auf einem ersten Community-Workshop einbezogen. Eine Übersicht der beteiligten Studien bzw. der Institutionen, die Daten anbieten, zeigt [Abbildung 1](#).

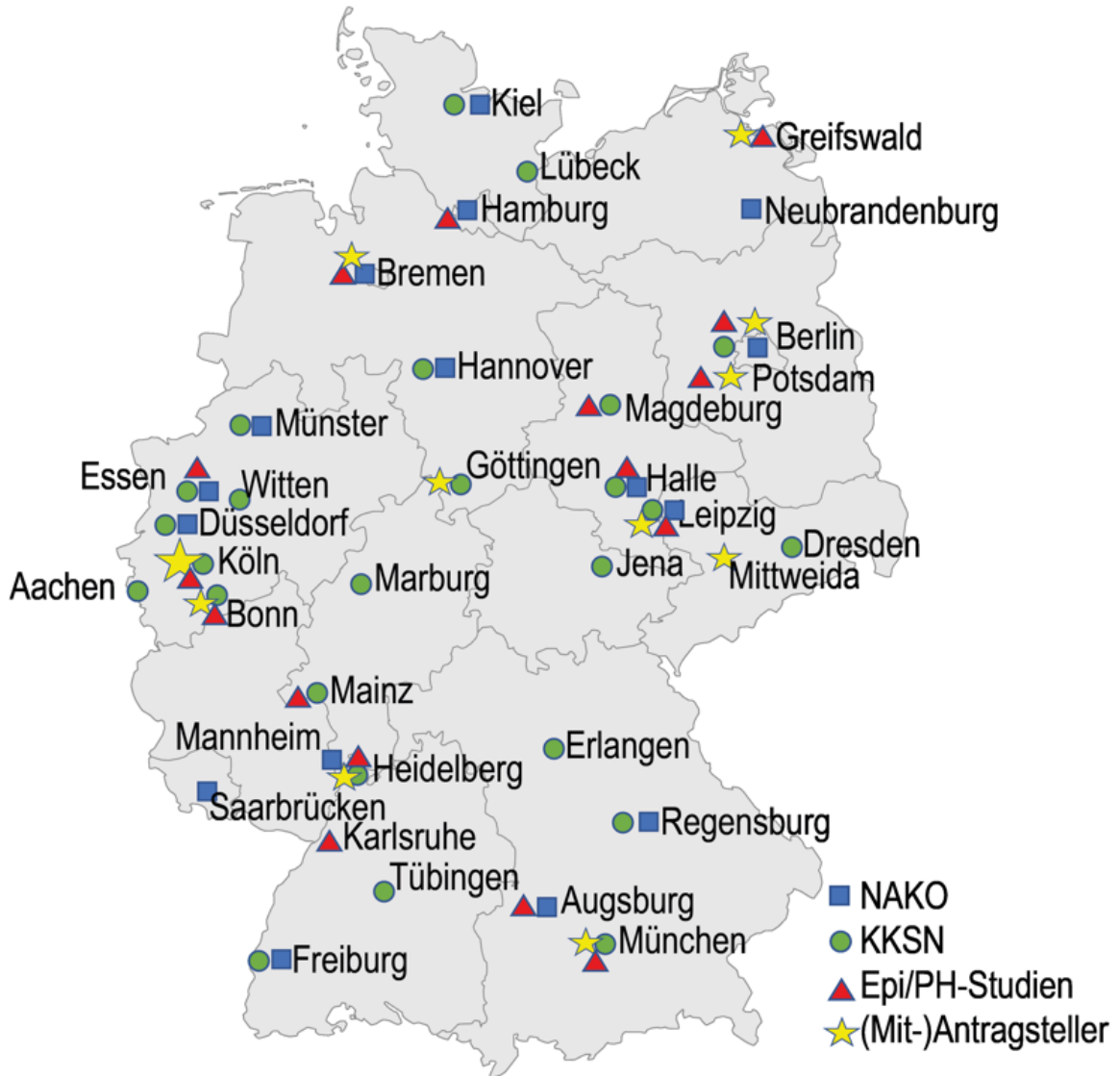


Abbildung 1: Studien bzw. Studienzentren, die sich in der NFDI4Health engagieren.  
 NAKO: NAKO Gesundheitsstudie; KKS: Netzwerk der Koordinierungszentren für klinische Studien; Epi: Epidemiologie; PH: Public Health

## 5 Zentrale Herausforderungen und Lösungsansätze der NFDI4Health

In Bezug auf personenbezogene Gesundheitsdaten lassen sich drei zentrale Herausforderungen erkennen, die im Folgenden mit den entsprechenden Lösungsansätzen in NFDI4Health beschrieben werden.

### 5.1 Auffindbarkeit und Zugang zu strukturierten Gesundheitsdaten

**Herausforderungen:** Die Auffindbarkeit der personenbezogenen Gesundheitsdaten ist aufgrund fehlender internationaler Standards für Registrierung und Publikation häufig eingeschränkt. Zwar existieren klinische Studienregister, aber epidemiologische und Gesundheitsstudien sind dort nicht erfasst. Außerdem werden weder strukturelle Metadaten, die das Design, den Stichprobenplan und die Variablen von klinischen Studien oder Beobachtungsstudien beschreiben, noch deren administrative Metadaten routinemäßig in strukturierter Form ausgetauscht oder veröffentlicht.

Datensätze sind oft nicht interoperabel, z. B. aufgrund der großen methodischen Heterogenität bei der Erfassung von Expositionen (s. Use Case „Ernährungsepidemiologie“) und Gesundheitsfolgen (s. Use Case „Epidemiologie chronischer Krankheiten“). Weiterhin sind die Möglichkeiten des Datenzugriffs in der Regel nicht einfach einsehbar, nicht zentral geregelt und durch vielfältige Varianten von individuellen Webseiten geprägt.

**Der NFDI4Health Lösungsansatz:** Um Studienmetadaten und Studiendokumente persistent identifizierbar zu machen, werden z. B. Publikationsrichtlinien entwickelt, die u. a. auf Publikationsstandards wie dem DataCite Metadata Schema<sup>3</sup> basieren. NFDI4Health wird einen Best-Practice-Entwurf für die Veröffentlichung von durchsuchbaren Metadaten bereitstellen, die mit den entsprechenden Datensätzen der jeweiligen Studien, die in verschiedenen Registern liegen, verknüpft sind. Dies wird die permanente Auffindbarkeit dieser Ressourcen sicherstellen. Um die Interoperabilität der (Meta-)Daten zu gewährleisten, werden internationale medizinische Standards wie z. B. CDISC<sup>4</sup>-Standards und HL7 FHIR<sup>5</sup> genutzt. Die Implementierung der bestehenden Standards in die Use Cases gewährleistet FAIRness sowie deren Anpassung an die Bedürfnisse der NutzerInnen.

Als NFDI-Service-Element wird NFDI4Health ein zentrales Studienportal aufbauen. Das Studienportal bietet semantische Such- und Filterfunktionalitäten und Kohorten-Browsing als zusätzliche innovative Elemente. Der Service baut auf Open-Source-

<sup>3</sup>DataCite Metadata Schema. <https://schema.datacite.org>, letzter Zugriff: 26.03.2021.

<sup>4</sup>CDISC. <https://www.cdisc.org>, letzter Zugriff: 26.03.2021.

<sup>5</sup>HL7 FHIR. <https://www.hl7.org/fhir/>, letzter Zugriff: 23.03.2021.

Komponenten wie FAIRDOM<sup>6</sup>, SEEK<sup>7</sup> und Mica<sup>8</sup> auf. Die Suchfunktionalität des zentralen Studienportals bezieht sich auf Metadaten und ist damit unabhängig vom Zugang zu personenbezogenen Studiendaten. Dies ist bei Gesundheitsdaten entscheidend, da die Daten selbst oft nicht offen zugänglich sind.

NFDI4Health plant, den zentralen Datenzugangspunkt in enger Kooperation mit dem NFDI-Konsortium GHGA<sup>9</sup> und der Medizininformatik-Initiative (MII)<sup>10</sup> einzurichten, um ForscherInnen mit einer einheitlichen Schnittstelle zu unterstützen und die NutzerInnen durch ein gemeinsames Zugangsprotokoll zu führen. Der Zugang wird einen Daten- und Einwilligungsbroker-Service anbieten und den Kontakt der NutzerInnen mit den DatenhalterInnen erleichtern. Im Hinblick auf die Interoperabilität der Dateninfrastrukturen wird NFDI4Health an Standards für Schnittstellen arbeiten, die allen datenhaltenden Organisationen eine einfache Integration in das NFDI4Health-Netzwerk und die Verknüpfung mit anderen Datendiensten erlauben.

## 5.2 Datenanalyse im Einklang mit den Datenschutzbestimmungen

**Herausforderungen:** Datenschutzbestimmungen und spezifische informierte Einwilligungserklärungen der StudienteilnehmerInnen schränken die Wiederverwendbarkeit der Daten ein. Datenschutz und Datensicherheit sind jedoch übergreifende Anforderungen an die Analyse personenbezogener Daten. Die Verarbeitung von Gesundheitsdaten unterliegt besonders strengen Datenschutzerfordernissen, die bei deren Nutzung berücksichtigt werden müssen.

**Der NFDI4Health Lösungsansatz:** Eines der langfristigen Ziele von NFDI4Health ist die Förderung einer harmonisierten Regelung der Forschungsdatenverarbeitung, die mit anderen Initiativen inner- und außerhalb der NFDI abgestimmt werden. NFDI4Health wird Konzepte und Lösungen bereitstellen, die eine DSGVO-konforme Datennutzung insbesondere für verteilte Datenanalysen und für Anwendungen der Künstlichen Intelligenz (KI) unterstützen. Zukünftige Rahmenwerke werden darüber hinaus auch die geistigen Eigentumsrechte (IPR) der DateninhaberInnen berücksichtigen.

Wir werden Werkzeuge bereitstellen, die eine Analyse der Datenschutzrisiken der in die Infrastruktur eingebrachten Datensätze ermöglichen. Eine weitere Option zur Unterstützung der gemeinsamen Nutzung von Daten, die wir in NFDI4Health explorieren,

<sup>6</sup>FAIRDOM. <https://fair-dom.org>, letzter Zugriff: 26.03.2021.

<sup>7</sup>Wolstencroft, Katherine et al. „SEEK: A systems biology data and model management platform.“ *BMC Systems Biology* 9 (2015): 33. <https://doi.org/10.1186/s12918-015-0174-y>.

<sup>8</sup>Doiron, Dany et al. „Software Application Profile: Opal and Mica: Open-source software solutions for epidemiological data management, harmonization and dissemination.“ *International Journal of Epidemiology* 46, Nr. 5 (2017): 1372-1378. <https://doi.org/10.1093/ije/dyx180>.

<sup>9</sup>GHGA. <https://ghga.dkfz.de/>, letzter Zugriff: 16.03.2021.

<sup>10</sup>Semler, Sebastian C., Frank Wissing und Ralf Heyder. „German Medical Informatics Initiative: A national approach to integrating health data from patient care and medical research.“ *Methods of Information in Medicine* 57 Suppl 1 (2018): e50. <https://doi.org/10.3414%2FME18-03-0003>.

ist die Generierung synthetischer Daten, bei der hinreichend realistische aber nicht personenbezogene Daten aus einem Eingabedatensatz abgeleitet werden. Synthetische Daten ermöglichen Data Mining und Modellentwicklung ohne die rechtlichen Zugangsbeschränkungen realer personenbezogener Daten.

Für die Auswertung von Daten wird NFDI4Health die Infrastrukturen DataSHIELD<sup>11</sup> und Personal Health Train (PHT)<sup>12</sup> für verteilte Datenanalysen einsetzen und der Forschungsgemeinschaft zur Verfügung stellen. Beide Infrastrukturen ermöglichen die verteilte Analyse von studien- (bzw. standort-) spezifischen Individualdaten, ohne dass diese ihren Ursprungsort verlassen müssen, d. h. ein physikalisches Pooling der Individualdaten findet nicht statt. Dies setzt besondere FAIRness der zu untersuchenden Daten voraus. NFDI4Health wird die bereits in DataSHIELD und PHT verfügbaren Prozeduren zur Datenanalyse bezogen auf die Anforderungen von NutzerInnen erweitern, die aus den Use Cases erwachsen.

Dabei hat DataSHIELD verschiedene Mechanismen implementiert, um einen Zugriff auf strukturierte Individualdaten, insbesondere bei kleinen Fallzahlen, zu verhindern. Zudem stellt DataSHIELD bereits jetzt eine Auswahl an Prozeduren zur Datenanalyse bereit, deren Ergebnisse qualitativ ähnlich sind zu denen gepoolter Analysen<sup>13</sup>. Der PHT kann darüber hinaus für die Auswertung von Bild- und anderen großvolumigen Daten eingesetzt werden. Daher nutzt NFDI4Health die PHT-Infrastruktur insbesondere im Use Case „Radiomics/KI“, um Bilddaten mithilfe maschinellen Lernens und KI auszuwerten.

### 5.3 Zusammenarbeit mit Universitäten und Bibliotheken

**Herausforderungen:** Eine bislang weitgehend unbeantwortete Frage im Rahmen der NFDI insgesamt ist die Vernetzung der Fachkonsortien mit den Universitäten bzw. den Hochschulen. Zentrale Fragen sind hierbei, wie die erarbeiteten Standards und Services in den fachlich relevanten Fakultäten bekannt gemacht werden können und Feedback dazu eingeholt werden kann, um letztlich eine Anwendung durch die ForscherInnen an den Hochschulen zu erreichen.

**NFDI4Health sieht hierzu im Wesentlichen zwei Lösungsansätze vor:** Zum einen wird die Verbreitung der Ergebnisse über Ausbildung und Trainings angestrebt. Hier ist das Netzwerk der Koordinierungszentren für Klinische Studien (KKS-Netzwerk) aktiv und wird flächendeckend für alle klinischen Studienzentren Schulungen anbieten. Darüber hinaus ist NFDI4Health gemeinsam mit der U Bremen Research Alliance und Akteuren der weiteren drei in Bremen vertretenen NFDI-Konsortien (NFDI4BioDiversity, KonsortSWD, NFDI4Ing) am Aufbau eines interdisziplinären Graduiertenausbildungspro-

<sup>11</sup>DataShield. <https://www.datashield.ac.uk/>, letzter Zugriff: 26.03.2021.

<sup>12</sup>Beyan, Oya et al. „Distributed analytics on sensitive medical data: The Personal Health Train.“ *Data Intelligence 2*, Nr. 1-2 (2020): 96-107. [https://doi.org/10.1162/dint\\_a\\_00032](https://doi.org/10.1162/dint_a_00032).

<sup>13</sup>Amadou Gaye et al. „DataSHIELD: Taking the analysis to the data, not the data to the analysis.“ *International Journal of Epidemiology* 43 Nr. 6 (2014): 1929-1944. <https://doi.org/10.1093/ije/dyu188>.



gramms zu Forschungsdatenmanagement und Data Science „Data Train“<sup>14</sup> beteiligt. Der Lehrplan und die Module werden in einem ersten Durchlauf erprobt und nach einer Anpassung allen NFDI-Konsortien zur Verfügung gestellt.

Zum anderen werden Universitäts- bzw. Medizinbibliotheken als Multiplikatoren für die Beratung sowie für Schulungs- und Trainingsangebote zum Management von Daten aus epidemiologischen und klinischen Studien eingesetzt. Sie sollen als Knotenpunkte dienen, um lokal ForscherInnen und InfrastrukturanbieterInnen/Servicestellen mit NFDI4Health zu verbinden. Vor dem Hintergrund der Arbeitsteilung und der Suche nach lokalen MultiplikatorInnen wird ein Konzept für eine Zusammenarbeit zwischen spezialisierten Data Stewards für Fakultäten und weiteren Servicebereichen der Bibliothek entwickelt. Data Stewards unterstützen die ForscherInnen im Management der Daten aus epidemiologischen und klinischen Studien gemäß der in NFDI4Health erarbeiteten Ergebnisse, während die Bibliothek generische Aspekte des Forschungsdatenmanagements vermittelt und eine Plattform für die universitätsweite Vernetzung bietet. Nach einer Pilotphase an der Universität Köln soll das Data Steward-Konzept ggf. deutschlandweit auf weitere Fakultäten ausgeweitet werden<sup>15</sup>. Bei der Umsetzung dieses Konzepts greift das Konsortium auf die bereits existierenden Vernetzungsstrukturen wie die Forschungsdatenkompetenzzentren an Hochschulen oder die Landesinitiativen zum Forschungsdatenmanagement<sup>16</sup> zurück.

## 6 Zusammensetzung des NFDI4Health-Konsortiums

Das NFD4Health Konsortium<sup>17</sup> zeichnet sich durch seine transdisziplinäre Zusammenarbeit von KlinikerInnen, EpidemiologInnen, JuristInnen, (biomedizinischen) InformatikerInnen, ErnährungswissenschaftlerInnen, StatistikerInnen und InfrastrukturvertreterInnen aus. Juliane Fluck (ZB MED – Informationszentrum Lebenswissenschaften) ist Sprecherin und Iris Pigeot (Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS) stellvertretende Sprecherin von NFDI4Health. Die zentralen Dienste der NFDI4Health werden bei der ZB MED als Sprecherorganisation gehostet. Der Bereich Public Health und Epidemiologie ist durch Forschungsinstitute wie das Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, das Robert Koch-Institut (RKI), die Universitätsmedizin Greifswald, das Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE) in Leipzig, das Institut für

<sup>14</sup>Data Train. <https://www.uni-bremen.de/research-alliance/forschungsdaten/data-train>, letzter Zugriff: 26.03.2021.

<sup>15</sup>Shutsko, Aliaksandra und Birte Lindstädt. „Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten – NFDI4Health: Pilotprojekt zu Bibliotheken und Forschungsdatenkompetenzzentren als Multiplikatoren („Data Steward“).“ *GMS Medizin – Bibliothek – Information* 20 Nr. 3 (2020): Doc27. <https://doi.org/10.3205/mbi000484>.

<sup>16</sup>Vernetzungsorganisationen: Landesinitiativen. <https://www.forschungsdaten.org/index.php/FDM-Kontakte#Landesinitiativen>, letzter Zugriff: 26.03.2021.

<sup>17</sup>Fluck, Juliane et al. „National Research Data Infrastructure for Personal Health Data (NFDI4Health) Proposal“ (October 15, 2019). <https://doi.org/10.4126/FRL01-006421856>.

Ernährungsepidemiologie an der Universität Bonn und das Deutsche Institut für Ernährungsforschung Potsdam-Rehbrücke (DIfE) vertreten. Diese bilden als DatenproduzentInnen das Rückgrat der Infrastruktur für persönliche Gesundheitsdaten, die in der NFDI zur Verfügung stehen werden.

Der Bereich der klinischen Studien wird durch das KKS Netzwerk – Koordinierungszentrum für klinische Studien (KKSZ) als Partner der NFDI4Health abgedeckt. Über das KKSZ sind die universitären klinischen Studienzentren als zentrale Kommunikationsknotenpunkte in NFDI4Health eingebunden.

Standardisierungsprozesse werden vom Heidelberger Institut für Theoretische Studien (HITS) in Kooperation mit der Universitätsmedizin Göttingen und Greifswald, der ZB MED sowie dem Berlin Institute of Health (BIH) und dem IMISE koordiniert.

Die zu entwickelnden Services werden von der ZB MED als koordinierende Einrichtung in Kooperation mit zahlreichen weiteren PartnerInnen etabliert. Unterstützt werden alle Prozesse in Hinblick auf die Einbeziehung der User Communities insbesondere von der Universität Köln, der TMF Technologie- und Methodenplattform für die vernetzte medizinische Forschung, Berlin, und dem BIPS. Bei der FAIRifizierung von personenbezogenen Gesundheitsdaten müssen insbesondere ethische und juristische Aspekte beachtet werden, was u. a. in der Verantwortung der Universität Bremen und der Universitätsmedizin Göttingen liegt. Die sich daraus ergebenden methodischen Herausforderungen in der Datenanalyse werden von den Fraunhofer-Instituten FIT, SCAI und MEVIS, dem BIH, dem DIfE, dem Max-Delbrück-Centrum für Molekulare Medizin und der Hochschule Mittweida adressiert.

Die eingebundenen ProjektpartnerInnen zeichnen sich durch eine hohe IT-Kompetenz, hohe Datenschutzstandards und die Entwicklung etablierter Nutzungs- und Zugriffsverfahren aus. Außerdem besitzt das Konsortium langjährige Erfahrung im Bereich Forschungsdatenmanagement. NFDI4Health ist in ein breites Netzwerk von (inter-)nationalen Initiativen und Organisationen zum FAIRen Teilen personenbezogener Gesundheitsdaten eingebunden und an entsprechenden Standardisierungsgremien wie das Deutsche Institut für Normung (DIN)<sup>18</sup>, die International Organization for Standardization (ISO)<sup>19</sup>, der Research Data Alliance (RDA)<sup>20</sup>, European Open Science Cloud (EOSC)<sup>21</sup> oder der Health Level Seven International (HL7)<sup>22</sup> beteiligt. Somit ist NFDI4Health hervorragend aufgestellt, um eine nachhaltige Infrastruktur für den Datenaustausch unter ForscherInnen zu schaffen.

<sup>18</sup>DIN. <https://www.din.de/de>, letzter Zugriff: 22.03.2021.

<sup>19</sup>ISO. <https://www.iso.org/home.html>, letzter Zugriff: 22.03.2021.

<sup>20</sup>RDA. <https://www.rd-alliance.org/>, letzter Zugriff: 23.03.2021.

<sup>21</sup>EOSC. <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>, letzter Zugriff: 23.03.2021.

<sup>22</sup>HL7. <https://www.hl7.org/index.cfm>, letzter Zugriff: 23.03.2021.

## 7 Arbeitsprogramm der NFDI4Health

NFDI4Health hat ein Arbeitsprogramm konzipiert, das sich auf sechs Task Areas (TA) verteilt (s. Abbildung 2).

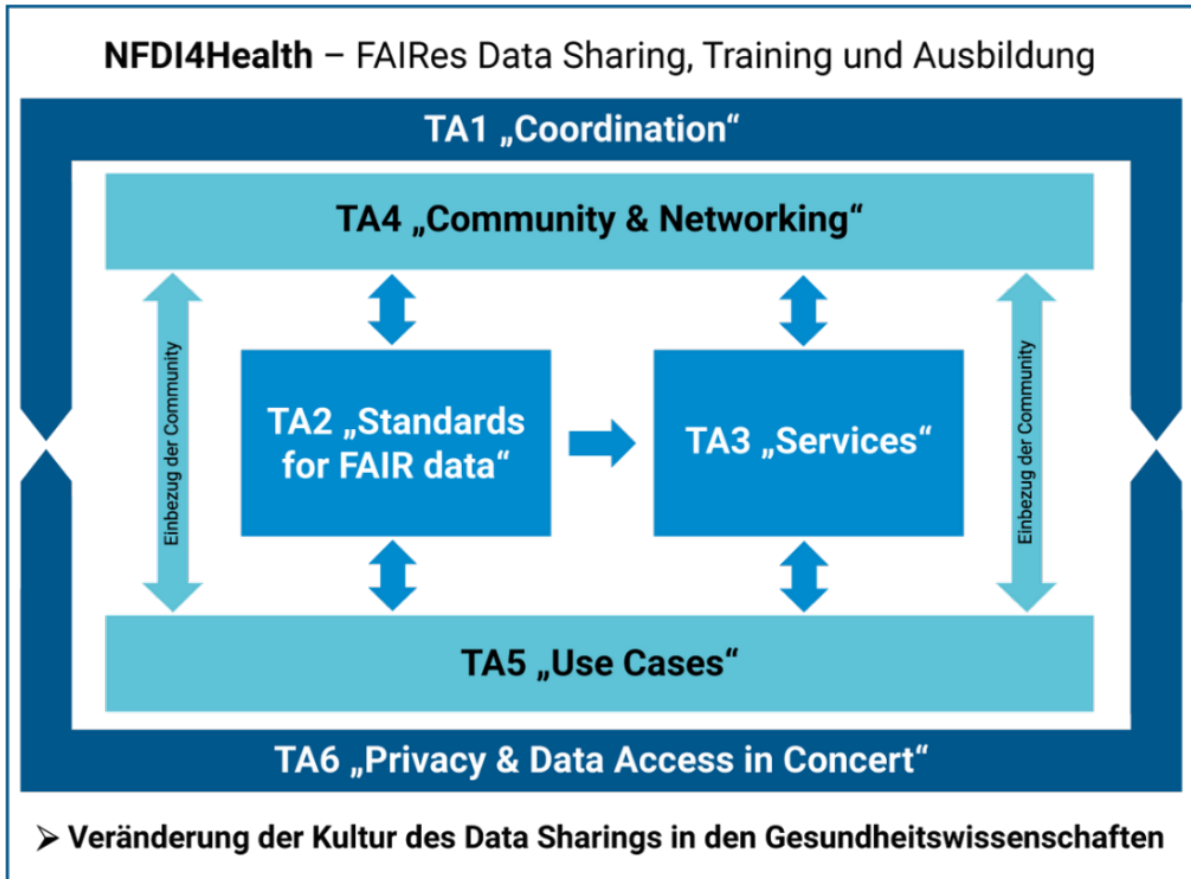


Abbildung 2: Zusammenarbeit der Task Areas (TA) in der NFDI4Health

**TA1** konzentriert sich auf Maßnahmen zur Projektkoordination, einschließlich des finanziellen Controllings des Gesamtprojekts, der Verbreitung, der Öffentlichkeitsarbeit und des Kontakts zu politischen EntscheidungsträgerInnen.

**TA2** bewertet und entwickelt Standards für Auffindbarkeit und Interoperabilität. Dies betrifft das Datenmanagement und Veröffentlichungsrichtlinien, die Harmonisierung von (Meta-)Datenstandards, die Bewertung der Datenqualität und Datenherkunft sowie die Standardisierung des Zugangs zu Gesundheitsdaten.

**TA3** nutzt die von TA2 erarbeiteten Standards, um NFDI4Health-Dienste, Software und die notwendige technische Unterstützung bereitzustellen. Dazu gehören ein zentraler Suchdienst, Werkzeuge zur Unterstützung von Datenqualitätsbewertungen und Metadatenannotationen sowie Dienste für Datennutzung und -zugriff, Daten-/Software-Repositoryn und Archivierungsdienste. Darüber hinaus bietet TA3 verteilte Datenanalyse Dienste an.

**TA4** befasst sich mit den Bedürfnissen und Anforderungen der ForscherInnengemeinschaft, wird potenzielle NutzerInnen informieren, schulen und deren Beteiligung organisieren, indem Werkzeuge und Wissen an die NutzerInnen weitergegeben werden und Feedback dazu eingeholt wird.

**TA5** gibt thematische Anwendungsfelder vor, für die TA2 und TA3 primär die notwendigen Standards und Dienste entwickeln, um diese dann in Use Cases zu implementieren. Dabei dient jeder Use Case als Beispiel für eine spezifische Herausforderung der FAIRifizierung von persönlichen Gesundheitsdaten.

**TA6** befasst sich mit den rechtlichen Rahmenbedingungen des Datenschutzes nach EU-Datenschutz-Grundverordnung (DSGVO), der Datensicherheit, der sicheren Verknüpfung von Daten über Studien und Register hinweg (Record Linkage) sowie mit Einwilligungserklärungen für eine breite Datennutzung und der Entwicklung von Werkzeugen zur Analyse und Minimierung der Datenschutzrisiken.

## 8 Use Cases der NFDI4Health

Die Use Cases (vgl. [Abbildung 3](#)) adressieren exemplarische Fälle zur Implementation der intendierten Infrastrukturen. Als Beispiel für expositionsorientierte Epidemiologie adressiert der Use Case „**Ernährungsepidemiologie**“ die Notwendigkeit, komplexe Expositionsdaten wie das Ernährungsverhalten über eine Vielzahl von Datenquellen hinweg zu standardisieren.

Der Use Case „**Epidemiologie chronischer Krankheiten**“ hingegen befasst sich mit chronischen Erkrankungen als Beispiel für Gesundheitsfolgen. Dieser Use Case konzentriert sich auf die harmonisierte Erfassung von Gesundheitsparametern, die im Zusammenhang mit chronischen Krankheiten relevant sind. Beide Use Cases implementieren NFDI4Health-Infrastrukturkomponenten in großen epidemiologischen Studien in Deutschland. Dazu gehören die NAKO-Gesundheitsstudie<sup>23</sup>, die EPIC-Studien Potsdam und Heidelberg<sup>24</sup>, KORA<sup>25</sup>, SHIP<sup>26</sup> und andere.

Der Zugang zu Daten, Metadaten, Dokumenten und Data Dictionaries aus klinischen Studien ist ein zentrales Ziel im Use Case „**Klinische Studien**“, in dem Infrastrukturen für Datenrepositorien sowie Suche und Zugriff im Kontext akademischer, öffentlich geförderter klinischer Studien implementiert werden.

Für die Auswertung medizinischer Bilddaten werden im Use Case „**Radiomics/KI**“ NFDI4Health-Infrastrukturkomponenten als prototypisches Analyse-Testbett für Künstliche Intelligenz (KI) implementiert. Radiomics/KI bietet dabei einen attraktiven Ansatz, um mit der Kombination strukturierter und bildgebungsbasierter Merkmale

<sup>23</sup>NAKO Gesundheitsstudie. <http://www.nako.de/>, letzter Zugriff: 26.03.2021.

<sup>24</sup>European Prospective Investigation into Cancer and Nutrition. <https://www.dife.de/forschung/kooperationen/epic-studie> und [https://www.dkfz.de/de/epidemiologie-krebserkrankungen/arbeitsgr/ernaerepi/EPIC\\_p02\\_EPIC\\_Europa.html](https://www.dkfz.de/de/epidemiologie-krebserkrankungen/arbeitsgr/ernaerepi/EPIC_p02_EPIC_Europa.html), letzter Zugriff: 26.03.2021.

<sup>25</sup>Kora-Studie. <https://www.helmholtz-muenchen.de/kora/index.html>, letzter Zugriff: 26.03.2021.

<sup>26</sup>SHIP-Studie. <https://www2.medizin.uni-greifswald.de/cm/fv/ship/>, letzter Zugriff: 26.03.2021.

neues Wissen zu generieren.

Der Use Case „**Sekundärdaten/Record Linkage**“ erforscht Verfahren zur Verknüpfung von Primärdaten speziell mit Sekundärdaten, wie Abrechnungsdaten der Krankenkassen, aber auch mit weiteren Primärdatenquellen wie Sequenzier- und Bilddaten, soziodemographischen und psychosozialen Daten.

Der ursprünglich geplante Use Case „Surveillance“ wurde im Zuge der Budgetkürzungen aus dem Arbeitsprogramm gestrichen. Stattdessen wurde der zusätzlich finanzierte Use Case „**Task Force COVID-19**“<sup>27</sup> aufgenommen, der u. a. die Erstellung eines Studienportals von deutschen Gesundheitsstudien zu COVID-19 und den dabei eingesetzten Erhebungsinstrumenten zum Ziel hat. Inhaltlich beziehen sich die Studien auf die Verbreitung und die Konsequenzen von COVID-19 in der Allgemeinbevölkerung und bei an COVID-19 Erkrankten.

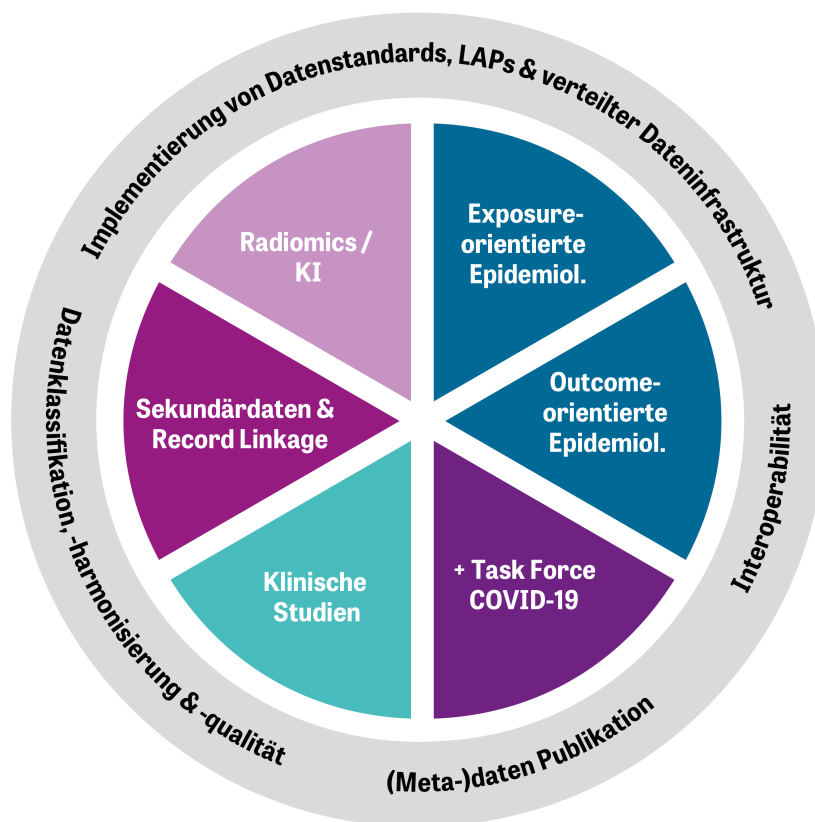


Abbildung 3: Use Cases der NFDI4Health; Exposure-orientierte Epidemiologie wird in dem Use Case „Ernährungsepidemiologie“ und Outcome-orientierte Epidemiologie im Use Case „Epidemiologie chronischer Krankheiten“ abgedeckt.

LAP: Local Access Point

<sup>27</sup>NFDI4Health Task Force COVID-19: Integration und Harmonisierung von Forschungsanstrengungen zum besseren Verständnis von COVID-19 und seinen Folgen. <https://gepris.dfg.de/gepris/projekt/451265285?context=projekt&task=showDetail&id=451265285&>, letzter Zugriff: 26.03.2021.

## 9 Fazit

NFDI4Health wird die Integration von Epidemiologie, Public-Health- und klinischer Forschung vorantreiben. Durch die Einrichtung von gemeinsamen Such- und Datenzugriffsdiensten sowie von Datenanalyseplattformen für verteilte Datenanalysen, die Definition gemeinsamer Standards für (Meta-)Daten und die Bereitstellung gemeinsamer Datenstrukturen sowie gemeinsamer Einwilligungserklärungen wird es gelingen, die bestehenden Datensilos zu öffnen und die wissenschaftliche Nutzung von persönlichen Gesundheitsdaten effizienter und nachhaltiger im Sinne der FAIR-Prinzipien zu gestalten.

## Danksagung

Diese Arbeit ist im Rahmen des NFDI4Health-Konsortiums entstanden (<https://www.nfdi4health.de>). Wir danken der Deutschen Forschungsgemeinschaft (DFG) für die finanzielle Unterstützung – NFDI 13/1.