

B 17 Forschungsdaten

1 Definition und Bedeutung

Der Begriff Forschungsdaten wird mittlerweile völlig selbstverständlich in verschiedenen Kontexten genutzt, trotzdem ist es schwierig, eine einheitliche und abgestimmte Definition zu finden.

In den *Leitlinien zum Umgang mit Forschungsdaten* der Deutschen Forschungsgemeinschaft (DFG 2015) findet sich eine umfassende Beschreibung, welche Forschungsobjekte Gegenstand der wissenschaftlichen Tätigkeit sein können und als Forschungsdaten gelten:

Forschungsdaten sind eine wesentliche Grundlage für das wissenschaftliche Arbeiten. Die Vielfalt solcher Daten entspricht der Vielfalt unterschiedlicher wissenschaftlicher Disziplinen, Erkenntnisinteressen und Forschungsverfahren. Zu Forschungsdaten zählen u. a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden. (S. 1)

Bei dieser Beschreibung wird deutlich, dass jede Fachkultur ein anderes Verständnis über Forschungsdaten hat und diese Vielfalt nicht in einer prägnanten Definition umfassend abgedeckt werden kann. Hinzu kommen Fragestellungen und Forschungsobjekte im Rahmen von (inter-)nationalen Projekten sowie interdisziplinäre Fragestellungen, die den Fokus auf Forschungsdaten verschieben können. Je nach Forschungsfragen, Forschungsobjekten als Untersuchungsgegenstände sowie verwendeter Methode variieren die Daten, die erhoben, bereinigt, prozessiert oder angereichert werden und aus ehemals heterogen verteilten Rohdaten ergeben sich hoch aggregierte Forschungsdaten, die mittels digitaler, wissenschaftlicher Werkzeuge und Dienste in andere Umgebungen überführt werden. Gemeinsam ist all diesen Forschungsdaten, dass sie in digitaler Form repräsentiert sind und das Aufkommen an digitalen Daten mit zunehmender Digitalisierung der Forschungsprozesse rasant wächst. Diese Digitalisierung erreicht mehr und mehr Fachdisziplinen: Während zu Beginn die Naturwissenschaften durch die gemeinsame Nutzung teurer Großgeräte eher gezwungen waren, diese Daten größeren Forschungsverbänden z. T. auch aus benachbarten (Teil-)Disziplinen zur Verfügung zu stellen, haben die Geisteswissenschaften z. B. mit der Neugründung der Disziplin Digital Humanities¹ 2013 der zunehmenden Digitalisierung ebenfalls Rechnung getragen.

Mit der wachsenden Bedeutung von Forschungsdaten für die Wissenschaft, aber auch für die Gesellschaft² hat die DFG konsequenterweise auch die förderpolitischen Rahmenbedingungen für Projektanträge deutlich präzisiert. So sind Angaben zum Umgang mit Forschungsdaten mittlerweile integraler Bestandteil eines jeden Förderantrags. Es wird deutlich gemacht, dass die Ausführungen zum Umgang mit Forschungsdaten Teil der Begutachtung und Teil der Berichtspflicht nach Abschluss des Projekts sind (DFG 2022).

¹ <https://dig-hum.de/ueber-dhd>.

² Dies wird in der heutigen Zeit eindrucksvoll durch die Corona-Pandemie belegt.

Eine von der DFG bereitgestellte *Checkliste zum Umgang mit Forschungsdaten* (DFG 2021a) umfasst die Bereiche „Datenbeschreibung“, „Dokumentation und Datenqualität“, „Speicherung und technische Sicherung während des Projektverlaufs“, „Rechtliche Verpflichtungen und Rahmenbedingungen“, „Datenaustausch und dauerhafte Zugänglichkeit der Daten“ sowie „Verantwortlichkeiten und Ressourcen“. Es wird darauf hingewiesen, dass diese Checkliste bereits mit der Planung eines Forschungsvorhabens angewendet werden sollte. Dies ist ein klarer Paradigmenwechsel im Vergleich zu früheren Zeiten, wo Forschungsdaten – wenn überhaupt – erst mit dem Ende eines Projekts (vielleicht) in einen wie auch immer gearteten Fokus gerieten.

Noch deutlicher wird dieser Paradigmenwechsel mit der Veröffentlichung des DFG-Kodex *Leitlinien zur Sicherung guter wissenschaftlicher Praxis* in Version 1.0 (DFG 2019) bzw. in Version 1.1 (DFG 2021b). Zum ersten Mal zielen die Standards der guten wissenschaftlichen Praxis (GwP) nicht nur auf die Publikation von Artikeln ab und mahnen ein wissenschaftliches Fehlverhalten z. B. bei Plagiaten an, sondern inkludieren alle am Forschungsprozess beteiligten Personen (s. u.), Objekte (Forschungsdaten), Werkzeuge (digitale Tools, Dienste) und Software bzw. Code. So heißt es in den Erläuterungen zu der Leitlinie 7 „Phasenübergreifende Qualitätssicherung“:

Die Herkunft von im Forschungsprozess verwendeten Daten, Organismen, Materialien und Software wird kenntlich gemacht und die Nachnutzung belegt; die Originalquellen werden zitiert. Art und Umfang von im Forschungsprozess entstehenden Forschungsdaten werden beschrieben. Der Umgang mit ihnen wird, entsprechend den Vorgaben im betroffenen Fach, ausgestaltet. Der Quellcode von öffentlich zugänglicher Software muss persistent, zitierbar und dokumentiert sein. (DFG 2021b, S. 14–15)

Bezüglich aller am Forschungsprozess beteiligten Personen berücksichtigt der Kodex in Leitlinie 14 „Autorschaft“ explizit auch Beteiligte, die einen „genuinen, nachvollziehbaren Beitrag zu dem Inhalt einer wissenschaftlichen Text-, Daten- oder Softwarepublikation geleistet [haben]“ (DFG 2021b, S. 19). Dies umfasst auch einen Beitrag zu „der Entwicklung und Konzeption des Forschungsvorhabens“, „der Erarbeitung, Erhebung, Beschaffung, Bereitstellung der Daten, der Software, der Quellen“, „der Analyse/Auswertung oder Interpretation der Daten, Quellen und an den aus diesen folgenden Schlussfolgerungen“ oder ein Beitrag „am Verfassen des Manuskripts“ (DFG 2021b, S. 20). Auch hier wird deutlich, dass Forschungsdaten ins Zentrum gerückt sind und alle Personen, die sich in erheblicher Weise um auch nur einen Aspekt im Lebenszyklus der Forschungsdaten wissenschaftlich verdient gemacht haben, ein Recht auf Nennung bei einer (Daten-)Publikation haben.

Abschließend werden in der Leitlinie 17 „Archivierung“ die Eckpunkte genannt, die sich dezidiert mit Forschungsdaten befassen und verdeutlichen, welche Bedeutung Forschungsdaten gewonnen haben und dass die archivierten Daten einen Wert an sich darstellen, also nicht nur bloßes Beiwerk zu Artikelpublikationen sind. Die Regel ist, die Forschungsdaten zu sichern und in Ausnahmefällen zu begründen, warum dies nicht möglich ist:

Wissenschaftlerinnen und Wissenschaftler sichern öffentlich zugänglich gemachte Forschungsdaten beziehungsweise Forschungsergebnisse sowie die ihnen zugrunde liegenden, zentralen Materialien und gegebenenfalls die eingesetzte Forschungssoftware, gemessen an den Standards des betroffenen Fachgebiets, in adäquater Weise und bewahren sie für einen angemessenen Zeitraum auf. Sofern nachvollziehbare Gründe dafür existieren, bestimmte Daten nicht aufzubewahren, legen die Wissenschaftlerinnen und Wissenschaftler dies dar. Hochschulen und außerhochschulische Forschungseinrichtungen stellen sicher, dass die erforderliche Infrastruktur vorhanden ist, die die Archivierung ermöglicht. (DFG 2021b, S. 22)

2 FAIRe Forschungsdaten

Parallel zur Selbstverpflichtung der Wissenschaft zur Sicherung der guten wissenschaftlichen Praxis, die im DFG-Kodex (s. o.) festgehalten ist, gibt es weitreichende internationale Diskussionen, Initiativen und Entwicklung, die maßgeblich zur Stärkung der Transparenz, Reproduzierbarkeit (Bollen et al. 2015) und Nachnutzbarkeit von Forschungsdaten auch im Sinne von Open Science (s. Kapitel E 9 Open Science) beitragen sollen. Angesichts weltweit rasant wachsender digitaler Datenmengen bei gleichzeitig steigender Heterogenität der Datenarten kommt dem Auffinden qualitätsgeprüfter Daten und der rechtssicheren sowie einfachen Nachnutzung von Daten eine große Bedeutung zu. Die FAIR-Prinzipien (FORCE11 2016; Wilkinson et al. 2016) mit den 15 FAIR-Kriterien bilden international die Basis für ein gemeinsames Verständnis über den Umgang mit Forschungsdaten und verdeutlichen, dass Forschungsdaten (zukünftig) auffindbar (**F**indable), zugänglich (**A**ccessible), interoperabel (**I**nteroperable) und wiederverwendbar (**R**eusable) sein sollen. Sie bilden die Grundlage für eine disziplinen- und länderübergreifende Nachnutzung der Daten und folgen dem Motto „As open as possible, as closed as necessary“ (Landi et al., 2020). Die FAIR-Prinzipien beziehen alle digitalen Daten, die im Zuge wissenschaftlicher Prozesse entstehen, ein und berücksichtigen dabei auch die Metadaten, Algorithmen, Werkzeuge und Software. FAIRe Daten müssen dabei – im Gegensatz zu Open Data (s. Kapitel E 11 Open Data) – nicht automatisch sofort und kostenfrei verfügbar sein, sondern erlauben eine dokumentierte Zugangsbeschränkung. Die FAIR-Prinzipien betonen nicht nur die intellektuelle Nachnutzung der Daten, sondern legen auch besonderen Wert auf die Maschinenlesbarkeit, d. h. der (teil-)automatischen Nachnutzung im Sinne der Weiterverarbeitung der Daten. Publisso (n. d.) gibt einen guten Überblick über die vier Säulen der FAIR-Prinzipien in deutscher Sprache:

Auffindbarkeit (to be Findable):

Daten und Metadaten sollten sowohl von Menschen als auch von Maschinen leicht zu finden sein. Grundlegende, maschinenlesbare und beschreibende Metadaten ermöglichen die Entdeckung interessanter Datensätze. Um dies zu gewährleisten sollten die Forschungsdaten durch Metadaten erläutert werden, z. B. durch einen Titel, den Autor, eine Inhaltsangabe oder die Erhebungsmethode. Die Vergabe eines persistenten Identifikators für die (Meta-) Daten trägt ebenfalls in hohem Maße zur Auffindbarkeit von Daten bei. Ein Beispiel hierfür ist der Digital Object Identifier (DOI).

Zugänglichkeit (to be Accessible):

Daten und Metadaten sollten verfügbar gemacht und langzeitarchiviert werden, sodass sie leicht von Menschen und Maschinen heruntergeladen und genutzt werden können. Dies erfolgt über Standard-Kommunikationsprotokolle wie zum Beispiel https. Eine weitere Bedingung für die Zugänglichkeit ist, dass die Metadaten verfügbar sind, auch wenn die eigentlichen Forschungsdaten beispielsweise aus Gründen des Datenschutzes nicht direkt abrufbar sind.

Interoperabilität (to be Interoperable):

Die Daten sollten derart vorliegen, dass sie mit anderen Datensätzen von Menschen und Maschinen verknüpft werden können. Dies wird dadurch erreicht, dass z. B. Ontologien oder Thesauri wie Medical Subject Headings (MESH) oder der AGROVOC bzw. maschinenlesbare Formate für Metadaten wie XML verwendet werden. Wird in den Metadaten mittels des persistenten Identifikators auf andere Datensätze verwiesen, indem beispielsweise Angaben wie „ist Teil von“ oder „ist eine Version von“ erfolgen, trägt dies ebenfalls zur Verknüpfbarkeit von Datensätzen bei.

Wiederverwendbarkeit (to be Reusable):

Zur Wiederverwendbarkeit trägt eine Beschreibung der Datensätze über Metadaten bei, sodass sie für weitere Forschungen nachnutzbar und mit anderen Datensätzen vergleichbar sind. Dabei spielt die Entstehungsgeschichte (Provenienz) eine wichtige Rolle: welche Methoden oder Geräte wurden für die Datengenerierung benutzt? Ein ordnungsgemäßes Zitieren der Daten muss möglich sein, in der Regel durch Nutzung des persistenten Identifikators wie dem Digital Object Identifier DOI. Außerdem sollte eine eindeutige Lizenz (z. B. eine Creative Commons Lizenz) die Bedingungen für die Nachnutzung für Menschen und Maschinen eindeutig kenntlich machen.

Abb. 1: Forschungsdaten referenzierbar und auffindbar machen (nach Publisso n. d.)

Die 15 FAIR-Prinzipien werden sehr anschaulich in deutscher Sprache bei Kraft (2017) und in englischer Sprache bei der Förderorganisation der Schweiz (Swiss National Science Foundation – SNF n. d.) erklärt. Dabei werden nicht nur die einzelnen Kriterien erläutert, sondern auch die jeweiligen Konsequenzen sowohl für die Forschenden als auch die Betreibenden von Forschungsdaten-Repositoryn abgeleitet.

Die FAIR-Prinzipien sollen helfen, den folgenden Herausforderungen zu begegnen, die nicht nur auf Grund der vermehrten Bedeutung von digitalen Daten insgesamt aufkommen, sondern auch darin begründet sind, dass Daten nicht mehrfach erhoben werden müssen:

- Wo finde ich geeignete Forschungsdaten für meine Fragestellung?
- Woher weiß ich, wie diese Daten entstanden sind und erhalte notwendige Kontextinformationen (z. B. Erhebung, Weiterverarbeitung, Anreicherung)?
- Wie kann ich diesen Daten bzw. der Quelle vertrauen?
- Darf ich diese Daten in meinem Kontext rechtssicher nachnutzen?
- Kann ich die Daten in meine (virtuelle) Forschungsumgebung automatisiert einlesen und dort weiterprozessieren?

Gleichzeitig tragen die Kriterien dazu bei, die Nachvollziehbarkeit (DFG 2021b, S. 17–18) – inklusive Validität, Objektivität, Reliabilität bzw. Reproduzierbarkeit – von Forschungsergebnissen zu erhöhen, indem sowohl Zwischenergebnisse als auch publizierte Ergebnisse (DFG 2021b, S. 18–19), bezogen auf den gesamten Forschungsprozess, nachvollziehbar dokumentiert und reproduzierbar sind. 1997 ereignete sich einer der bis heute größten Forschungsskandale Deutschlands: Dem Krebsforscher Friedhelm Herrmann wurde systematische Datenmanipulation in mehreren Fällen angelastet. Ihm (und einigen Mitarbeitenden) wurden nicht nur freie Erfindung von Daten ohne zugrunde liegende Laborversuche vorgeworfen, sondern auch die Fälschung von Daten. Die DFG veröffentlichte nach dem Einsetzen einer 2-jährig arbeitenden Untersuchungskommission im Juni 2000 eine Pressemitteilung,³ nach der von insgesamt 347 untersuchten Veröffentlichungen in 94 Fällen konkrete Hinweise auf Datenmanipulation vorlagen. Auch wenn in Zukunft die Manipulation von Forschungsdaten sicherlich nie ganz ausgeschlossen werden kann, erhöht die Anwendung der FAIR-Kriterien die Transparenz der Entstehungsgeschichte dieser Daten und trägt somit zur weiteren wissenschaftlichen Integrität bei.

Im Kontext der FAIR-Prinzipien entstanden in den letzten Jahren zahlreiche Initiativen, deren wichtigste im Folgenden genannt werden.

Die *Nationale Forschungsdateninfrastruktur* (NFDI): Die Bundesregierung und die Regierungen der Länder haben durch die Gemeinsame Wissenschaftskonferenz (GWK) die 10-jährige Förderung der NFDI im November 2018 beschlossen (GWK 2018). Startend mit der ersten Ausschreibung im Jahr 2019 und wissenschaftsgeleitet durch die DFG, sollen im Endausbau bis zu 30 Fach-Konsortien FAIRe Daten anbieten und entsprechende Dienste aufbauen. Dazu wurde im Oktober 2020 auch der NFDI Verein⁴ gegründet, der die Ziele formuliert, wertvolle Datenbestände von Wissenschaft und Forschung systematisch zu erschließen und nutzbar zu machen, Forschungsdatenmanagement (s. u.) nach FAIRen Kriterien zu etablieren und sich international zu vernetzen.⁵ Die NFDI wird als deutscher Beitrag zur European Open Science Cloud (s. u.) verstanden.

³ https://www.dfg.de/service/presse/pressemitteilungen/2000/pressemitteilung_nr_26/index.html.

⁴ <https://www.nfdi.de/>.

⁵ <https://www.nfdi.de/verein/>.

Die *European Open Science Cloud* (EOSC): Im November 2018 erfolgte die offizielle Auftaktveranstaltung zum Start der EOSC, die von der Europäischen Kommission mit dem Ziel initiiert wurde, u. a. den europäischen Forschenden eine föderierte und multi-disziplinäre Umgebung zur Verfügung zu stellen, in der Daten, Dienste und Werkzeuge veröffentlicht und nachgenutzt werden können.⁶ Auch hier spielen die FAIR-Prinzipien eine entscheidende Rolle, da die zuverlässige Nachnutzung von Forschungsdaten und allen anderen digitalen Objekten, die entlang des Forschungslebenszyklus entstehen (z. B. Methoden, Software und Publikationen), als zentrales Vorhaben im Sinne eines „Web of FAIR Data and Services“ betont werden.

GAIA-X⁷ wurde 2019 von den Bundesministerien für Wirtschaft und Energie (BMWi) und dem Bundesministerium für Bildung und Forschung (BMBF) initiiert. Mittlerweile handelt es sich um ein europäisches Projekt mit Vertreter*innen aus Wirtschaft, Politik und Wissenschaft. Ziel ist eine föderierte und sichere Dateninfrastruktur, welche auch die Öffentlichkeit adressiert und die Datenhoheit gewährleistet. Die Architektur von GAIA-X basiert auf dem Prinzip der vernetzten Dezentralisierung, folgt also nicht dem Cloud-Prinzip. Heute ist GAIA-X eine gemeinnützige Organisation nach belgischem Recht,⁸ der über 300 weltweite Mitglieder⁹ angehören (Stand Februar 2022).

Die *GO FAIR Initiative*¹⁰ ist ein bottom-up getriebener Ansatz, in dem sich Deutschland, Frankreich und die Niederlande gemeinsam zum Ziel gesetzt haben, die EOSC im Bereich der FAIRen Daten und Dienste zu unterstützen (Linne et al. 2021). In drei Säulen gliedert sie sich in GO Change (kulturellen Wandel), GO Train (Aus- und Weiterbildung) sowie GO Build (interoperable und föderierte Dateninfrastrukturen).

In der internationalen *Research Data Alliance* (RDA)¹¹ haben sich derzeit mehr als 12 000 Expertinnen und Experten aus 145 Ländern versammelt, um in über 100 sog. Working Groups und Interest Groups (RDA 2022) (fach-)spezifische Themen zu bearbeiten und als eine Art Quasi-Standard zu verabschieden. So hat z. B. die FAIR Data Maturity Model Working Group Empfehlungen mit dem *FAIR Data Maturity Model – core criteria to assess the implementation level of the FAIR data principles*¹² (RDA FAIR Data Maturity Model Working Group 2020) vorgelegt, die einen gemeinsamen Satz von Kernbewertungskriterien für die Bewertung der FAIRness in Form von Leitlinien und einer Checkliste für die Umsetzung der Indikatoren beinhalten. Das geförderte BMBF-Projekt EcoDM¹³ hat diese Empfehlungen 2021 in die deutsche Sprache übersetzt (RDA FAIR Data Maturity Model Working Group 2021), um die Verbreitung von FAIRness in den Bereichen Wissenschaft, Wirtschaft, Verwaltung und Qualifizierung in Deutschland zu erhöhen. Mit RDA-DE gibt es seit 2017 auch einen Verein¹⁴ für die deutsche RDA Community, die sich mindestens einmal jährlich zu einer großen Konferenz¹⁵ inklusive Workshops und Schulungen zusammenfindet.

6 https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en.

7 <https://www.data-infrastructure.eu/>.

8 <https://www.gaia-x.eu/who-we-are/association>.

9 <https://gaia-x.eu/members>.

10 <https://www.go-fair.org/>.

11 <https://www.rd-alliance.org/>.

12 <https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0>.

13 <https://www.ecodm.de/>.

14 <https://www.rda-deutschland.de/>.

15 2022: <https://www.rda-deutschland.de/events/rda-de-2022>.

Das EU-Projekt *Fostering Fair Data Practices in Europe (FAIRsFAIR)*¹⁶ zielt darauf ab, praktikable und nutzbare Lösungen zur Umsetzung der FAIR-Prinzipien im gesamten Lebenszyklus von Daten zur Verfügung zu stellen. Besonderer Fokus liegt auf der Förderung einer FAIRen Datenkultur, unterstützt durch die Veröffentlichung von Best Practices im Bereich der Nutzung und Umsetzung der FAIR-Prinzipien. Dazu gehören auch das aktuell erschienene Handbuch für die Qualifizierung im Bereich der FAIRen Datenkultur (Engelhardt et al. 2022).

Im Januar 2021 veröffentlichte die Bundesregierung ihre Datenstrategie (*Datenstrategie Bundesregierung*) mit dem Ziel, die „Datenbereitstellung und Datennutzung insbesondere in Deutschland und Europa signifikant [zu] erhöhen – in der Wirtschaft, der Wissenschaft, der Zivilgesellschaft und der Verwaltung“ (Bundeskanzleramt 2021). Über 240 Maßnahmen (Bundeskanzleramt 2021, S. 64–103), zum Teil (zukünftig) mit konkreten Förderprogrammen unterlegt, sollen den Aufbau von Dateninfrastrukturen, die Steigerung der Datennutzung, den Ausbau von Datenkompetenz und Datenkultur sowie die Digitalisierung der öffentlichen Verwaltungen unterstützen.

Das Registry of Research Data Repositories (*re3data*)¹⁷ ist ein internationales Nachweissystem für (fachspezifische) Forschungsdatenrepositorien, in denen (standardisierte) Beschreibungen zu Datensammlungen dauerhaft gespeichert und mittels (persistenter) Identifikatoren hinterlegt werden können. Das 2012 von der DFG geförderte Projekt ist mittlerweile unter dem Dach von DataCite¹⁸ angesiedelt, einer Non-Profit-Organisation im Bereich von Persistenten Identifikatoren (DOI). Re3data weist mit Stand Februar 2022 mehr als 2 700 Forschungsdatenrepositorien aus fast 100 Ländern nach, wobei aus Deutschland über 460 Repositorien beschrieben sind. In re3data ist es möglich, die Suche nach zertifizierten Repositorien einzuschränken, wobei das Zertifikat eine Art Gütesiegel darstellt. Das international bekannteste Zertifikat ist das CoreTrustSeal,¹⁹ welches klare Anforderungen für die Vergabe definiert (CoreTrustSeal 2019a).

3 Forschungsdatenmanagement und Datenmanagementpläne

Forschungsdatenmanagement (FDM) bezeichnet einen aktiven Prozess (Neuroth et al. 2018) und nicht etwa eine einmalige Verwaltungsaktivität. Kindling & Schirnbacher definieren bereits 2013:

Der gesamte Prozess, der die Allokation, die Generierung, die Bearbeitung und Anreicherung, die Archivierung und Veröffentlichung von digitalen Forschungsdaten selbst oder von einer entsprechenden klassischen Textproduktion unterstützt, wird inzwischen meist unter dem Begriff „Forschungsdatenmanagement“ zusammengefasst [...]. (Kindling & Schirnbacher S. 130)

FDM ist ein wesentlicher Baustein für die langfristige Nachnutzung von Forschungsdaten und unterstützt die wissenschaftliche Integrität²⁰ (DFG 2019) maßgeblich.

¹⁶ <https://www.fairsfair.eu/>.

¹⁷ <https://www.re3data.org/>.

¹⁸ <https://datacite.org/>.

¹⁹ <https://www.coretrustseal.org/>.

²⁰ <https://wissenschaftliche-integritaet.de/>.

Datenmanagementpläne (DMP) sind ein wichtiges Instrument, um Forschungsdatenmanagement konkret umzusetzen. Als zentrales Dokument zu den Forschungsdaten beschreibt ein DMP den Umgang mit diesen Daten und zwar sowohl vor, während als auch nach dem Forschungsvorhaben (Leendertse et al. 2019). Wichtig ist, dass der DMP entlang des Forschungsdatenlebenszyklus (Baur 2021) stets aktualisiert wird und somit systematisch wichtige Prozessschritte von der Planung, Erhebung bzw. Nachnutzung über die Weiterverarbeitung bis hin zur Publikation der Daten inklusive rechtliche Zugangsinformationen, die zugrunde liegenden Methoden und die Verwendung bestimmter Software, Algorithmen etc. festgehalten werden. Erst diese Dokumentation der Genese der Forschungsdaten erlaubt eine vertrauensvolle Nachnutzung der Daten, da Fragen wie „of what to keep and why [...] of who, how, why, for whom and how long [...]“ (Borgman 2015, S. 271) beantwortet werden können. Mittlerweile gibt es zahlreiche Leitfäden (z. B. Forschungsdaten.info n. d.; Universität Hannover 2020) und Checklisten (DFG 2021a) für die Erstellung von DMPs.

International gibt eine Reihe von digitalen *Werkzeugen*, die bei der Erstellung und Aktualisierung von DMPs unterstützen. Beispiele im anglo-amerikanischen Raum sind das DMPTool²¹ der California Digital Library (CDL)²² oder DMPonline²³ des Digital Curation Centers (DCC).²⁴ Beide Werkzeuge bieten strukturierte Fragenkataloge und ermöglichen das kollaborative Arbeiten über Institutions- und Ländergrenzen hinweg. Auch ist es möglich, eigene, dem jeweiligen Projekt angepasste Fragenkataloge zu hinterlegen. Bei Open Science FAIR²⁵ wird der Data Stewardship Wizard²⁶ angeboten, dem ein fachspezifisches und erweiterbares Wissensmodell inklusive Validierung²⁷ zugrunde liegt. In Deutschland hat sich, ausgehend von einem DFG-geförderten Projekt, der Research Data Management Organiser (RDMO)²⁸ durchgesetzt, der ein aktives FDM unterstützt. Während in den letzten Jahren der Fokus all dieser Werkzeuge darin lag, einen intellektuell lesbaren DMP zu generieren, liegt der Fokus nun vermehrt auf der Maschinenlesbarkeit der DMPs (Miksa et al. 2019). Dies ermöglicht z. B. die (teil)automatische Generierung bestimmter Bereiche von DMPs, die Integration in eigene Arbeitsumgebungen, den Austausch über Systemgrenzen hinweg und minimiert somit den Aufwand bei der Erstellung von DMPs.

Insgesamt ist zu beobachten, dass Forschungsdatenmanagement und Datenmanagementpläne in den letzten Jahren nicht nur international, sondern auch in Deutschland als Thema und neues Aufgabenspektrum deutlich sichtbarer geworden sind und sich auch in die Breite bzw. Fläche ausgeweitet haben. Beigetragen²⁹ dazu haben u. a.:

- die Verabschiedungen *institutioneller* Policies (z. B. Hiemenz & Kuberek 2019) oder Richtlinien (z. B. Universität Osnabrück 2021; Freie Universität Berlin 2021), Strategien (z. B. Universität Potsdam 2020) und Handreichungen (z. B. Arbeitsgruppe Forschungsdaten 2018; Hausen et al. 2018),

²¹ <https://dmptool.org>.

²² <https://cdlib.org/>.

²³ <http://www.dcc.ac.uk/dmponline>.

²⁴ <https://dcc.ac.uk/>.

²⁵ <https://www.opensciencefair.eu/>.

²⁶ <https://ds-wizard.org>.

²⁷ <https://github.com/ds-wizard/ds-km>.

²⁸ <https://rdmorganiser.github.io>.

²⁹ Vgl. Übersicht bei: https://www.forschungsdaten.org/index.php/Data_Policies.

- *fachspezifische* Stellungnahmen und Empfehlungen von Fachgesellschaften, die in Fachkollegien der DFG Verwendung finden³⁰ und damit auch Grundlage für Begutachtungsverfahren sind (z. B. für die Psychologie: Gollwitzer et al. 2020),
- das Bewusstsein zahlreicher *Bundesländer*, FDM strategisch auf Landes- und Hochschulebene zu adressieren (z. B. Wissenschaftliche Kommission Niedersachsen 2021),
- die Etablierung von *Landesinitiativen* zu Forschungsdatenmanagement (z. B. Brandenburg,³¹ Nordrhein-Westfalen³²),
- die zunehmende Angebotspalette³³ im Bereich *Schulungen, Workshops, Materialien* (z. B. Train2Dacar³⁴ bzw. Lonati et al. 2018; Train-the-Trainer Konzept³⁵ bzw. Biernacka et al. 2021),
- die Veröffentlichung von *Handbüchern* (z. B. Putnings et al. 2021) oder spezifischen Ratgebern z. B. zu Rechtsfragen³⁶ (Baumann et al. 2021),
- *Studiengänge* im Bereich (Forschungs-)Datenmanagement wie z. B. der Weiterbildende Masterstudiengang Digitales Datenmanagement (DDM)³⁷ (Petras et al. 2019),
- erste *Handreichungen für Studierende* zum Umgang mit Forschungsdaten im Kontext von Abschlussarbeiten, inklusive Template eines DMPs und einem ausgefüllten Beispiel,³⁸ angeboten von der Humboldt-Universität zu Berlin seit Dezember 2021.

Auch wenn diese zahlreichen Aktivitäten vielversprechend sind und das Fundament für den Paradigmenwechsel im Bereich der digitalen Transformation darstellen sowie der Digitalisierung der Wissenschaftsdisziplinen mit ihrer rasant steigenden Anzahl an Forschungsdaten und deren Nachnutzungspotenzialen Rechnung tragen, so gibt es im Bereich der Etablierung neuer Berufsfelder mit verlässlichen Karrierewegen noch signifikante Defizite. Zwar scheinen sich die Begriffe Data Steward und Data Stewardship (CoreTrustSeal 2019b) mehr und mehr durchzusetzen (Scholtens et al. 2019) und z. B. in den Niederlanden (Teperek & Plomp 2019) oder Dänemark (Wildgaard 2020) als feste Berufsbezeichnungen mit definierten Verantwortungs- und Tätigkeitsbeschreibungen zu etablieren, in Deutschland dagegen gibt es noch einigen Nachholbedarf zur Professionalisierung der Tätigkeiten im Umfeld von Forschungsdaten. Dabei gilt, was bereits die Expertengruppe FAIR Data der europäischen Kommission 2018 in ihrem Bericht geschrieben hat:

New job profiles need to be defined and education programs put in place to train the large cohort of data scientists and data stewards required to support the transition to FAIR. Since the skillsets required for data science and data stewardship are varied and rapidly evolving, multiple formal and

30 https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/empfehlungen/index.html.

31 <https://fdm-bb.de/>.

32 <https://www.fdm.nrw/>.

33 Vgl. z. B. Materialsammlung der UAG Schulungen/Fortbildungen der DINI/nestor AG Forschungsdaten, https://rs.cms.hu-berlin.de/uag_fdm/.

34 http://campus.hesge.ch/id_bilingue/projekte/train2dacar/.

35 <https://www.fdm.nrw/index.php/veranstaltungen/train-the-trainer-workshop-zum-forschungsdatenmanagement/>.

36 <https://www.ajbd.de/publikationen/arbeitshefte/ah-28-baumann-krahn-lauber%E2%80%90roensberg-forschungsdatenmanagement-und-recht/>.

37 <https://www.ddm-master.de/>.

38 <https://www.ibi.hu-berlin.de/de/studium/rundumdasstudium/fdm-fuer-studierende>.

informal pathways to learning are required. (European Commission, Directorate-General for Research and Innovation 2018, S. 13)

Auch die Empfehlungen des Rats für Informationsinfrastrukturen (RFII) rund um die „Köpfe“, die mit den vermehrt auftretenden Stellenausschreibungen nicht nur im Kontext der NFDI adressiert werden, haben bisher nicht an Bedeutung verloren (RFII 2016, S. 3).

4 Literaturverzeichnis

- Arbeitsgruppe Forschungsdaten (2018). *Forschungsdatenmanagement. Eine Handreichung*. Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen. <https://doi.org/10.2312/allianz0a.029>.
- Baumann, P., Krahn, P. & Lauber-Rönsberg, A. (2021). *Forschungsdatenmanagement und Recht. Datenschutz-, Urheber- und Vertragsrecht*. W. Neugebauer.
- Baur, J. (2021). FDM erklärt – Was ist eigentlich ein Daten-Lebenszyklus? (16. April 2021). *RWTH Blog*. <https://blog.rwth-aachen.de/forschungsdaten/2021/04/16/fdm-erklart-was-ist-eigentlich-ein-daten-lebenszyklus/>.
- Biernacka, K., Buchholz, P., Danker, S. A., Dolzycka, D., Engelhardt, C., Helbig, K., Jacob, J., Neumann, J., Odebrecht, C., Petersen, B., Slowig, B., Trautwein-Bruns, U., Wiljes, C. & Wuttke, U. (2021). *Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement*. (Version 4.0). <https://doi.org/10.5281/zenodo.5773203>.
- Bollen, K., Cacioppo, J. T., Kaplan, R. M., Krosnick, J. & Olds, J. L. (2015). *Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science*. Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences. https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf.
- Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press.
- Bundeskanzleramt (2021). *Datenstrategie der Bundesregierung: Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum* (Kabinettfassung, 27. Januar 2021). <https://www.bundesregierung.de/breg-de/suche/datenstrategie-der-bundesregierung-1845632>.
- CoreTrustSeal Standards and Certification Board. (2019a). *CoreTrustSeal Trustworthy Data Repositories Requirements 2020-2022*. <https://doi.org/10.5281/zenodo.3638211>.
- CoreTrustSeal Standards and Certification Board. (2019b). *CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2020-2022*. <https://doi.org/10.5281/zenodo.3632563>.
- Deutsche Forschungsgemeinschaft (2000). *Task Force legt Abschlußbericht vor* (Pressemitteilung Nr. 26). https://www.dfg.de/service/presse/pressemitteilungen/2000/pressemitteilung_nr_26/index.html.
- Deutsche Forschungsgemeinschaft (2015). *Leitlinien zum Umgang mit Forschungsdaten*. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf.
- Deutsche Forschungsgemeinschaft (2019). *Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex* (Version 1.0). <https://doi.org/10.5281/zenodo.3923602>.
- Deutsche Forschungsgemeinschaft (2021a). *Umgang mit Forschungsdaten: Checkliste für Antragstellende zur Planung und zur Beschreibung des Umgangs mit Forschungsdaten in Forschungsvorhaben*. https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf.
- Deutsche Forschungsgemeinschaft (2021b). *Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex* (Version 1.1). https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf.
- Deutsche Forschungsgemeinschaft (2022). *Umgang mit Forschungsdaten*. https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/.
- Engelhardt, C., Biernacka, K., Coffey, A., Cornet, R., Danciu, A., Demchenko, Y., Downes, S., Erdmann, C., Garbuglia, F., Germer, K., Helbig, K., Hellström, M., Hettne, K., Hibbert, D., Jetten, M., Karimova, Y.,

- Kryger H., Karsten, K., Mari, E. & Zhou, B. (2022). *How to be FAIR with your data. A teaching and training handbook for higher education institutions*. (V1.2 DRAFT). <https://doi.org/10.5281/zenodo.5905866>.
- European Commission, Directorate-General for Research and Innovation (2018). *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2777/54599>.
- FORCE11 (2016). *The FAIR Data Principles*. <https://www.force11.org/group/fairgroup/fairprinciples>.
- Forschungsdaten.info (n. d.). *Der Datenmanagementplan: Eine Wegbeschreibung für Daten*. <https://www.forschungsdaten.info/themen/informieren-und-planen/datenmanagementplan/>.
- Freie Universität Berlin (2021). *Forschungsdaten-Policy der Freien Universität Berlin*. <http://dx.doi.org/10.17169/refubium-30560>.
- Gemeinsame Wissenschaftskonferenz (2018). *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur vom 26. November 2018 (BAnz AT 21.12.2018 B10)*. <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf>.
- Gollwitzer, M., Abele-Brehm, A., Fiebach, C., Ramthun, R., Scheel, A. M., Schönbrodt, F. D. & Steinberg, U. (2020). *Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen*. <https://doi.org/10.31234/osf.io/hcxtm>.
- Hausen, D., Trautwein-Bruns, U. & von der Ropp, S. (2018). Handreichung zur Beratung im Kontext des Forschungsdatenmanagements an der RWTH Aachen University. *Bausteine Forschungsdatenmanagement*, 1, 43–49. <https://doi.org/10.17192/bfdm.2018.1.7815>.
- Hiemzen, B. & Kuberek, M. (2019). *Strategischer Leitfaden zur Etablierung einer institutionellen Forschungsdaten-Policy*. <http://dx.doi.org/10.14279/depositonce-8412>.
- Kindling, M. & Schirmbacher, P. (2013). „Die digitale Forschungswelt“ als Gegenstand der Forschung. *Information: Wissenschaft & Praxis*, 64(2–3), 127–136. <https://doi.org/10.1515/iwp-2013-0017>.
- Kraft, A. (2017). Die FAIR Data Prinzipien für Forschungsdaten (12. September 2017). *TIB Blog*. <https://blogs.tib.eu/wp/tib/2017/09/12/die-fair-data-prinzipien-fuer-forschungsdaten/>.
- Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., Olavo Bonino da Silva Santos, L. & Roos, M. (2020). The „A“ of FAIR – As Open as Possible, as Closed as Necessary. *Data Intelligence*, 2 (1–2), 47–55. https://doi.org/10.1162/dint_a_00027.
- Leendertse, J., Mocken, S. & von Suchodoletz, D. (2019). Datenmanagementpläne zur Strukturierung von Forschungsvorhaben. *Bausteine Forschungsdatenmanagement*, 2(4–9), 4–9. <https://doi.org/10.17192/bfdm.2019.2.8003>.
- Linne, M., Drefs, I., Dörrenbächer, N., Siegers, P. & Bug, M. (2021). GO FAIR und GO CHANGE: Chancen für das deutsche Wissenschaftssystem. In M. Putnings, H. Neuroth & J. Neumann (Hrsg.), *Praxishandbuch Forschungsdatenmanagement* (S. 215–238). De Gruyter Saur. <https://doi.org/10.1515/9783110657807-013>.
- Lonati, S. S., Sommet, M. & Pugin, L. (2018). *Kursbuch Forschungsdatenmanagement*. Hochschule für Technik und Wirtschaft Chur.
- Miksa, T., Simms, S., Mietchen, D. & Jones, D. (2019). Ten Principles for Machine-Actionable Data Management Plans. *PLOS Computational Biology*, 15(3), Article e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>.
- Neuroth, H., Engelhardt, C., Klar, J., Ludwig, J. & Enke, H. (2018). Aktives Forschungsdatenmanagement, *ABI Technik*, 38(1), 55–64. <https://doi.org/10.1515/abitech-2018-0008>.
- Petras, V., Kindling, M., Neuroth, H. & Rothfritz, L. (2019). Digitales Datenmanagement als Berufsfeld im Kontext der Data Literacy. *ABI Technik*, 39(1), 26–33. <https://doi.org/10.1515/abitech-2019-1005>.
- Publisso. (n. d.). *Die FAIR-Prinzipien für Forschungsdaten*. Open-Access-Publikationsportal für die Lebenswissenschaften. <https://www.publisso.de/forschungsdatenmanagement/fair-prinzipien/>.
- Putnings, M., Neuroth, H. & Neumann, J. (Hrsg.) (2021). *Praxishandbuch Forschungsdatenmanagement*. De Gruyter Saur. <https://doi.org/10.1515/9783110657807>.
- Rat für Informationsinfrastrukturen (2016). *Leistung aus Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. <http://www.rfii.de/?p=1998>.
- Research Data Alliance (2022). *RDA Groups*. <https://www.rd-alliance.org/groups>.
- Research Data Alliance FAIR Data Maturity Model Working Group (2020). *FAIR Data Maturity Model: Specification and Guidelines*. <https://doi.org/10.15497/rda00050>.

- Research Data Alliance FAIR Data Maturity Model Working Group (2021). *Das FAIR Data Maturity Model. Spezifikation und Leitlinien*. <https://doi.org/10.5281/zenodo.5834115>.
- Scholtens, S., Jetten, M., Böhmer, J., Staiger, C., Slouwerhof, I., van der Geest, M. & van Gelder, C. W. G. (2019). *Final report: Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise* (Version 3). <https://doi.org/10.5281/zenodo.3474789>.
- Swiss National Science Foundation (n. d.). *Explanation of the FAIR data principles*. http://www.snf.ch/Site-CollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf.
- Teperek, M. & Plomp, E. (2019). *The role and value of data stewards in Universities: A TU Delft case study on data stewardship*. <https://doi.org/10.5281/zenodo.2684278>.
- Universität Hannover (2020). *Leitfaden zur Erstellung eines Datenmanagementplans* (Version 2.3). https://www.fdm.uni-hannover.de/fileadmin/fdm/Dokumente/Leitfaden_DMP_LUH_v2.3.pdf.
- Universität Osnabrück (2021). *Richtlinie für das Forschungsdatenmanagement an der Universität Osnabrück („Forschungsdaten-Policy“)* (Amtliches Mitteilungsblatt der Universität Osnabrück, 05/2021). https://www.ub.uni-osnabrueck.de/fileadmin/documents/public/Startseite/4_Publizieren_und_Archivieren/4_4_Forschungsdaten/UOS_AmtlMitteilungen_Nr05-2021_FDM_Policy.pdf.
- Universität Potsdam (2020). *Forschungsdatenstrategie 2019–2022*. Universitätsverlag Potsdam. <https://doi.org/10.25932/publishup-44436>.
- Wildgaard, L. (2020). *Reframing Data Stewardship Education in Denmark and abroad*. <https://doi.org/10.5281/zenodo.3628375>.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R. & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Wissenschaftliche Kommission Niedersachsen (2021). *Forschungsdatenmanagement in Niedersachsen. Bericht der Arbeitsgruppe*. <https://www.wk.niedersachsen.de/download/177260>.