Jens Dierkes

4.1 Planung, Beschreibung und Dokumentation von Forschungsdaten

Abstract: Forschungsdaten sind ein zentraler Bestandteil von Forschungsprozessen und rücken zunehmend in den Fokus der wissenschaftlichen Kommunikation. Dabei gibt es eine Reihe von Anforderungen von verschiedenen Agierenden an das Forschungsdatenmanagement. Zum Zwecke der Sicherung der guten wissenschaftlichen Praxis, aber auch um die interne/offene Nachnutzung, die Referenzierung oder Reproduzierbarkeit zu ermöglichen, ist ein planvoller Umgang mit den Forschungsdaten nötig. Im folgenden Kapitel werden der Forschungsprozess und der Datenlebenszyklus eingeführt und die darin enthaltenen Schritte der Planung und Dokumentation beschrieben. Dabei spielen auf der einen Seite Metadaten eine entscheidende Rolle, auf der anderen Seite aber auch das Zusammenspiel, im Sinne einer Aufgabenteilung und dem Zusammenbringen unterschiedlicher Kompetenzen, zwischen den verschiedenen Stakeholdern. Dies gilt insbesondere für die Forschenden und die forschungsunterstützenden Dienstleister.

1 Einleitung

Um die vielfachen Anforderungen an Forschungsdaten zu erfüllen, ist ein strukturierter und systematischer, kurz ein planvoller Umgang mit diesen erforderlich. Als Referenzwerke im deutschsprachigen Raum können die Arbeiten von Büttner, Hobohm und Müller¹ sowie Ludwig und Enke² herangezogen werden. In diesem Kapitel werden wesentliche Aspekte beleuchtet, wie die *Beschreibung* und *Dokumentation* von Forschungsdaten. Damit können die Daten auch von Dritten verstanden und es kann ein Kontext hergestellt werden. Neben u. a. der Einhaltung der Leitlinien zur Sicherung der guten wissenschaftlichen Praxis (Kodex)³ und der Reproduzierbarkeit ist dies gerade für die Nachnutzung der Forschungsdaten in anderen Kontexten (interdisziplinäre Forschungsfragen, wie z. B. in der Klimaforschung) von großer Bedeutung. Digitale Arbeitsweisen und Methoden unterscheiden sich zwi-

¹ Vgl. Büttner, Hobohm und Müller 2011.

² Vgl. Ludwig und Enke 2013.

³ S. https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/index.html. Letztes Abrufdatum der Internet-Dokumente ist der 15.11.2020.

schen den einzelnen Disziplinen.⁴ Diese Diversität spiegelt sich auch in dem Umgang mit Forschungsdaten wider.⁵

Abb. 1: Darstellung des Forschungsprozesses, der nicht-linearen Entwicklung von Forschungsprojekten und den zahlreichen Stufen, bei denen Daten gesammelt werden (basierend auf Jahnke und Asher 2012, 10)

Die (schriftliche) *Planung* des Umgangs mit Forschungsdaten während einer Projektphase und darüber hinaus wird mittlerweile von vielen Forschungsförderern gefordert.⁶ Die daraus resultierenden Datenmanagementpläne (DMP) sollen dokumentieren, dass sich die Beteiligten des Forschungsprojektes Gedanken über die anfallenden Forschungsdaten, die Datentypen, die Menge an Daten und deren Speicherung, Sicherheitsaspekte, Dokumentation der Verarbeitungsschritte, mögliche Nachnutzungsszenarien und insbesondere auch über die dafür notwendigen technischen und personellen Ressourcen gemacht haben.

⁴ Vgl. Meyer und Schroeder 2015, 197–208.

⁵ Vgl. Borgman 2015, 81–200.

⁶ S. a. Beitrag von Putnings, Kap. 1.3 in diesem Praxishandbuch.

2 Prozesse und Akteure

2.1 Prozesse und Lebenszyklen

Sowohl Forschung als auch der Umgang mit Forschungsdaten lassen sich als Prozesse bzw. Lebenszyklen verstehen.⁷ Dies hilft, Aufgaben des Forschungsdatenmanagements (FDM) gewissen Phasen der Prozesse zuzuordnen. Abb. 1 zeigt eine Darstellung des Forschungsprozesses. Hierbei fällt auf, dass Daten nicht nur in einer bestimmten Phase eines Projektes gesammelt werden, sondern zu ganz unterschiedlichen Zeiten anfallen. Jahnke, Asher und Keralis beschreiben, dass es sein kann, dass Daten bzw. Verfahren zur Verarbeitung von Daten bereits im Vorfeld einer Projektförderung entwickelt werden, diese aber im Laufe des Projektes modifiziert werden.⁸ Kooperationspartner stoßen zum Projekt dazu und bringen ihre eigenen Forschungsdaten ein. Manchmal kann es bis in die aktive Forschungsphase hinein dauern, bis Daten in systematischer Form verarbeitet werden. Oder das Projekt entwickelt sich in eine ganz andere Richtung. Forschende können z. T. nicht abschätzen, welche Daten in Zukunft nützlich sein könnten. Das führt oftmals zu einer Unsicherheit darin, welche Daten und Metadaten (kontextuelle Daten) bewahrt werden sollen.

Obwohl der Forschungsprozess keine einfache lineare Aneinanderreihung von kategorisierbaren Aktivitäten ist, ist der Zugang über den Datenlebenszyklus sinnvoll, weil damit wesentliche Phasen des FDM und damit verbunden Aufgaben, Rollen und Verantwortlichkeiten adressiert werden.⁹ Eine stärker auf den Datenlebenszyklus ausgerichtete Darstellung zeigt die Abb. 2.

Neben den Aufgaben und Aktivitäten, die einzelnen Phasen zuzuordnen sind, gibt es Themenbereiche, die als Querschnittsthemen in allen Phasen von Bedeutung sind. Hierzu gehören die Bereiche Organisation, Recht, Finanzierung, Metadaten und Identifikatoren.¹⁰ Eine weitere Sichtweise ist das sog. Data Curation Continuum von Treloar, Groenewegen und Harboe-Ree mit einem Update von Treloar und Klump.¹¹ Hierbei werden eher Nutzungsprofile und damit verbundene Kuratierungsaufgaben betrachtet. Forschungsdaten entstehen häufig durch Aktivitäten einzelner Forschender oder Gruppen (entspricht der „privaten Domäne“). Diese werden in weiteren Schritten in Kooperationen geteilt (entspricht der „Gruppendomäne“). Je nach Art der Daten und Ziele des Forschungsvorhabens werden die Daten oder Teile davon in die dauerhafte Domäne verschoben, wo eine (quasi-)öffentliche Nachnut-

⁷ Vgl. z. B. Borgman 2019, 3–6; Jahnke, Asher und Keralis 2012, 9–11.

⁸ Vgl. Jahnke, Asher und Keralis 2012, 9.

⁹ Vgl. Michener 2015; Ludwig und Enke 2013, 14–18.

¹⁰ Vgl. Ludwig und Enke 2013, 16–17.

¹¹ Vgl. Treloar, Groenewegen und Harboe-Ree 2007; für das Update vgl. Treloar und Klump 2019.

zung möglich wird (entspricht der „Publikationsdomäne“).¹² Diese zwei Sichtweisen können genutzt werden, um mit den Agierenden und Stakeholdern über Erwartungen, Aufgaben, Rollen und Verantwortlichkeiten zu sprechen.

Abb. 2: Datenlebenszyklus nach Michener (vgl. Michener 2015)

2.2 Stakeholder

Im FDM gibt es zahlreiche Sichtweisen und Interessengruppen, die in komplexer Art und Weise zusammenspielen.¹³ Dazu gehören:

- Vertreterinnen und Vertreter der akademischen Gemeinschaft,
- die Universitätsleitung,
- professionelle Dienstleister mit Interesse am FDM, wie z. B. Bibliotheken, Rechenzentren, IT- und Forschungsadministratorinnen und -administratoren,
- Forschungsförderer,
- wissenschaftliche Communities,
- kommerzielle Partner (z. B. Verlage).

Es ist nicht einfach, diese Sichtweisen voneinander losgelöst zu betrachten. Der Fokus in diesem Kapitel liegt bei den Forschenden und forschungsunterstützenden Dienstleistern (Informationsinfrastruktureinrichtungen).

¹² Vgl. z. B. Klar und Enke 2013, 13–57.

¹³ Vgl. Cox und Verbaan 2018, 71.

2.3 Rollen

Die Herausforderung in einem Forschungsprojekt oder bei der Entwicklung einer unterstützenden Infrastruktur liegt nun darin, Rollen und Aufgaben zu ermitteln und zu definieren, Arbeitsabläufe und digitale Werkzeuge weiterzuentwickeln oder neue zu designen und zu etablieren. Dabei gibt es viele Möglichkeiten, Rollen und Aufgaben zu verteilen. Im Zuge immer komplexerer Forschungsvorhaben und der Anforderungen an die Informationsinfrastruktur scheint eine Arbeitsteilung notwendig.¹⁴ Arbeitsteilung kann hier insbesondere die Zusammenarbeit von Forschenden und Infrastruktureinrichtungen bedeuten. Beispiele sind Sonderforschungsbereiche der Deutschen Forschungsgemeinschaft (DFG), die explizit Infrastrukturentwicklung erlauben.¹⁵ Dies kann dann in Kooperation mit einem Rechenzentrum und/oder einer Bibliothek erfolgen.¹⁶

Zentrale Rollen nehmen Datenproduzierende und Datennutzende sowie die Infrastruktureinrichtungen und die Institution ein. Tab. 1 nennt die zentralen Rollen, ihre Rechte, Verantwortlichkeiten sowie mögliche Beziehungen zu anderen Rollen bzw. Agierenden.

Tab. 1: Rollen, Rechte und Verantwortlichkeiten beim FDM¹⁷

Rolle	Rechte	Verantwortlichkeiten	Beziehungen zu
Forschende/Datenproduzierende	Erstnutzung anerkannt werden Training und Beratung erhalten	Daten während des Projektes managen gute wissenschaftliche Praxis erfüllen Förderer und institutionelle Anforderungen beachten Daten für Nachnutzung vorbereiten	der Institution als Angestellte der Fachcommunity den Infrastruktureinrichtungen der Förderinstitution
Forschende/Datennutzende	Daten nachnutzen zu können hochqualitative Metadaten zu Einschätzung erhalten	Lizenzbedingungen einzuhalten die Datenerzeugenden zu zitieren abgeleitete Daten managen	den Infrastruktureinrichtungen als den Datenlieferanten den Institutionen als Datenlieferanten

¹⁴ Vgl. z. B. Cremer, Engelhardt und Neuroth 2015; Pampel, Bertelmann und Hobohm 2010; Treloar und Wilkinson 2008; Lyon 2007, 43–52; Brand & Dierkes 2020.

¹⁵ Hierzu zählen z. B. Sonderforschungsbereiche, in denen sogenannte INF-Teilprojekte zur Entwicklung von Informationsinfrastruktur beantragt werden können. S. https://www.dfg.de/foerderung/programme/koordinierte_programme/sfb/antragsteller/programmelement_inf/index.html.

¹⁶ Vgl. Curdt et al. 2019, 65; Cremer, Engelhardt und Neuroth 2015.

¹⁷ Basierend auf Abb. 3 bei Pampel, Bertelmann und Hobohm 2010, 11; Tab. 1 in Lyon 2007, 9.

Rolle	Rechte	Verantwortlichkeiten	Beziehungen zu
Datenmanager (sowohl im Forschungsvorhaben als auch auf institutioneller Seite)	anerkannt werden (z. B. Ko-Autorenschaft)	Management der Daten auf Koordinierungsebene	den Datenerzeugende dem Data Librarian der Institution dem Datenzentrum
Data Librarian	anerkannt werden (z. B. Ko-Autorenschaft)	FDM-Arbeitsabläufe fachnah begleiten Entwicklung von Standards	den Datenerzeugenden der Datenmanagerin/ dem Datenmanager der Institution dem Datenzentrum
Institution	darf eine Kopie der Daten erhalten	Vorgabe der internen FDM-Policy Management der Daten auf mittelfristiger Basis Einhaltung der guten wissenschaftlichen Praxis Ko-Finanzierung zentraler Infrastruktur	den Forschenden als Arbeitgeber den Infrastrukturexpertinnen und -experten (Rechenzentrum, Bibliothek usw.)
Rechenzentrum/Bibliothek/Kompetenzzentrum	anerkannt werden durch Forschende Rückhalt durch die Institution	Bereitstellung von Training und Beratung zur Unterstützung der Forschenden Fördern eines Repositorium-Dienstes Bereitstellung nachhaltiger Infrastruktur	den Forschenden als Datenproduzierenden und -nutzenden der Institution der Datenmanagerin/ dem Datenmanager

Hier wird deutlich, dass FDM Teamarbeit ist. Im Forschungs- bzw. Datenlebenszyklus kommen die einzelnen Agierenden zusammen und tragen gemeinsam dazu bei, dass die Daten entsprechend nachhaltig verwaltet werden. Forschende sind nicht allein, wenn es darum geht, geeignete Maßnahmen zu ergreifen, um entsprechende Metadaten zu den entsprechenden Zeitpunkten zu erfassen und aufzubereiten.

Für ein Forschungsvorhaben sind nach Michener folgende zehn Fragen relevant, die für eine sinnvolle Datenstrategie beantwortet werden sollten:¹⁸

- Was sind die Anforderungen der Förderorganisationen zum FDM?¹⁹
- Welche Daten werden gesammelt?
- Wie werden die Daten organisiert?
- Wie werden die Daten dokumentiert?²⁰

¹⁸ Vgl. Michener 2015.

¹⁹ S. a. Beitrag von Putnings, Kap. 1.3 in diesem Praxishandbuch.

²⁰ S. a. Beitrag von Dierkes, Kap. 4.1 in diesem Praxishandbuch.

- Wie wird die Datenqualität gewährleistet?²¹
- Wie sieht die Datenspeicherungs- und Archivierungsstrategie aus?²²
- Wie wird mit Daten im Forschungsvorhaben und darüberhinausgehend umgegangen (Daten-Policy)?
- Wie werden die Daten disseminiert?
- Welche Rollen und Verantwortlichkeiten gibt es?
- Wie sieht ein realistisches Budget für das FDM aus?

Diese Fragen können in Verbindung mit dem Lebenszyklus oder dem Domänenmodell Forschenden wie auch Infrastruktureinrichtungen die Tür zu einem strukturierter Umgang mit Forschungsdaten im jeweiligen Forschungsvorhaben öffnen.

Im Folgenden wird auf die Bereiche Planung, Beschreibung und Dokumentation von Forschungsdaten näher eingegangen. Details zu (Metadaten-)Datenqualität, infrastruktureller Lösungen und Archivierung bzw. langfristiger Verfügbarmachung von Forschungsdaten sind Themen nachfolgender Kapitel.

3 Planung

Ein wesentlicher Aspekt des Planens – besonders bei der Konzeption und der Antragstellung – ist für alle Beteiligten (Stakeholder), das ganze Forschungsvorhaben hinsichtlich des Umgangs mit Forschungsdaten über ihren gesamten Lebenszyklus in den Blick zu nehmen. Es geht darum, möglichst frühzeitig wesentliche Aspekte bzw. Herausforderungen des Umgangs mit den Daten im Forschungsprojekt zu identifizieren und zu berücksichtigen.²³ In der Regel sollen dabei die FDM-Rahmenbedingungen, die durch Förderinstitutionen,²⁴ durch Fachgesellschaften,²⁵ Fachverlage,²⁶ durch Gesetzgebung²⁷ oder andere institutionelle Ebenen (z. B. Daten-Polici-

²¹ S. a. Beitrag von Brase und Király, Kap. 4.3 in diesem Praxishandbuch.

²² S. a. Beitrag von Weber und Piesche, Kap. 4.2 in diesem Praxishandbuch.

²³ Vgl. z. B. Leendertse, Mocken und Suchodoletz 2019.

²⁴ S. z. B. Bundesministerium für Bildung und Forschung, <https://www.bildung-forschung.digital/de/forschungsdatenmanagement-2835.html>, DFG https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/, Horizon 2020 Programm der Europäischen Union, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm.

²⁵ S. z. B. Richtlinien zum Umgang mit Forschungsdaten in der Biodiversitätsforschung, https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten_biodiversitaetsforschung.pdf; Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie: Stellungnahme des Vorstands und Konzils der DGS, https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/dgs_stellungnahme_forschungsdaten.pdf.

²⁶ S. z. B. <https://journals.plos.org/plosone/s/data-availability>.

²⁷ Vgl. z. B. die europäische Datenschutz-Grundverordnung oder das deutsche Gentechnikgesetz.

es an Hochschulen,²⁸ Promotionsordnungen) vorgegeben werden, eingehalten werden. Obwohl die Planung des Umgangs mit Forschungsdaten häufig noch als zusätzlicher administrativer Aufwand gesehen wird, lohnt sich ein planvoller Umgang mit Forschungsdaten.²⁹ Auch macht FDM dann Sinn, wenn es nicht vorrangig um die mittlerweile häufig geforderte Publikation und Nachnutzung von Forschungsdaten in der oben genannten Publikationsdomäne geht. Ein Beispiel sind Daten, die in der Arbeitsgruppe eines Lehrstuhls anfallen und die strukturiert sowie gut dokumentiert gesichert werden sollen, damit beispielsweise sukzessive aufeinander aufbauende Dissertationen effektiv und effizient realisiert werden können. Neben der Selbstverpflichtung der Wissenschaft zur Sicherung der guten wissenschaftlichen Praxis³⁰ findet eine internationale Entwicklung zur Stärkung von Transparenz und Reproduzierbarkeit im Sinne von Open Science³¹ statt. Die FAIR-Prinzipien³² als eine Komponente von Open Science³³ stellen hierbei eine Art Leitfaden dar, um Forschungsergebnisse, auch im Sinne von Maschinenlesbarkeit, nachnutzbar zu machen. Jedoch hängt die Art der Umsetzung auch in besonderem Maße von den Gepflogenheiten und Anforderungen der jeweiligen Fachgebiete ab.

3.1 Datenmanagementplan

Zurzeit wird das Thema Planung des FDM in Forschungsvorhaben in der Förderlandschaft noch recht uneinheitlich behandelt. Es zeigt sich ein Spektrum von einfachen Hinweisen zum FDM über Empfehlungen zum Umgang mit Forschungsdaten inklusive der Möglichkeit dezidierte Mittel fürs FDM zu beantragen, bis hin zu konkreten Anforderungen, auf die in einem DMP eingegangen werden soll. Der DMP ist ein mitunter formales Dokument, das sowohl aus allgemeingültigen Fragen zum Umgang mit Forschungsdaten als auch solche, die sich aus den Anforderungen ergeben, besteht. Dazu gibt es umfassende Fragenkataloge wie z. B. die WissGrid-Checkliste³⁴ und Vorlagen, die neben Themen und Fragen auch eine Form vorgeben.³⁵

²⁸ S. https://www.forschungsdaten.org/index.php/Data_Policies.

²⁹ Z. B. Goodman et al. 2014.

³⁰ S. https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp.

³¹ Vgl. Boulton et al. 2012, 7–12.

³² FAIR steht für Findable, Accessible, Interoperable, Re-usable, vgl. Wilkinson et al. 2016.

³³ S. Beitrag von Linne et al., Kap. 3.2 in diesem Praxishandbuch.

³⁴ Vgl. Ludwig und Enke 2013.

³⁵ Zwei Beispiele sind der DMP des Horizon 2020 Förderprogramms, s. https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx, und die Vorlage der e-Infrastructures Austria Initiative, s. <http://phaidra.univie.ac.at/o:459774>.

Daneben gibt es eine Reihe digitaler Werkzeuge, mit denen sich DMP automatisch erstellen lassen. Etablierte Beispiele im englischsprachigen Raum sind das DMPTool des California Digital Curation Centers³⁶ und DMPonline des Digital Curation Centers.³⁷ Diese Tools enthalten gängige Fragenkataloge und erlauben zudem eigene Kataloge zu hinterlegen und kollaboratives Arbeiten. Ein interessantes Beispiel im Bereich FAIR Open Science ist der Data Stewardship Wizard.³⁸ Diesem Tool liegt ein komplexes, erweiterbares Wissensmodell zugrunde.³⁹ Ganz ähnlich entwickelt sich zurzeit der Research Data Management Organizer (RDMO)⁴⁰ für den deutschsprachigen Raum.⁴¹ Hier liegt der Fokus auf dem aktiven Datenmanagement. Wie eingangs erwähnt ist der Forschungsprozess und damit auch die Planung des FDM komplex und ein nichtlinearer, i. d. R. iterativer Prozess. Daher macht es Sinn die digitalen Werkzeuge wie RDMO kontinuierlich einzusetzen und den Planungsstand regelmäßig zu aktualisieren, ganz im Sinne eines „lebenden Dokuments“. Diese Vorgehensweise wird mittlerweile auch von einigen Förderorganisationen in ihren Programmen wie EU-Horizon 2020 und dem European Research Council (ERC) erwartet. Damit kann auch auf dieser Ebene eine Transparenz für die Nachvollziehbarkeit geschaffen werden. Die Wahl des geeigneten Werkzeuges hängt von mehreren Faktoren ab. Dazu zählen die unterstützten Sprachen, die Verfügbarkeit von spezifischen Fragenkatalogen sowie die Verfügbarkeit von lokalen bzw. institutionellen Angeboten. Beratungsangebote machen häufig deutlich, dass Forschende lokale Angebote bevorzugen, bei denen die in den DMP enthaltenen Daten auf Servern der Institution gespeichert werden.

Bisher sind DMP im Wesentlichen von Menschen lesbare Dokumente. Eine neuere Entwicklung geht in die Richtung der Maschinenlesbarkeit.⁴² Hier wird es den beteiligten Agierenden (Stakeholdern) einfacher möglich gemacht, Erfahrungen und Informationen über Forschungswerkzeuge und Systeme hinweg auszutauschen und DMP in existierende Arbeitsabläufe einzubetten. Teile von DMP können automatisch generiert und geteilt werden, womit administrativer Aufwand verringert und die Qualität von Informationen in einem DMP erhöht werden. Mögliche Vorteile für einige der oben genannten Stakeholder sind nach Miksa et al.:⁴³

- *Forschende*: Ermöglicht die Vernetzung mit Expertinnen und Experten während eines Forschungsprojektes, um Beratung und Unterstützung für das Datenmanagement zu erhalten. Automatisierte Prozesse können die Erstellung von DMP

³⁶ S. <https://dmptool.org>.

³⁷ S. <http://www.dcc.ac.uk/dmponline>.

³⁸ S. <https://ds-wizard.org>.

³⁹ S. <https://github.com/ds-wizard/ds-km>.

⁴⁰ S. <https://rdmorganiser.github.io>.

⁴¹ Vgl. Neuroth et al. 2018.

⁴² Z. B. Miksa et al. 2019.

⁴³ Vgl. Miksa et al. 2019.

erleichtern, andere dazu dienen, DMP zu aktualisieren, die Datensicherung zu optimieren und die Berichterstattung zu automatisieren. DMP stellen auch eine bedeutende Informationsquelle für die Planung und Durchführung von Experimenten dar.

- *Infrastrukturbetreiber*: Informationen können zwischen Systemen ausgetauscht und müssen nicht mehrfach eingegeben werden; sie können von den entsprechenden Beteiligten im Namen der Forschenden aktualisiert (was auch die Qualität der Informationen verbessert) und für die Prozessoptimierung aggregiert werden.
- *Mitarbeiterinnen und Mitarbeiter der Forschungsunterstützung*: Sie können die Qualität der in einem DMP enthaltenen Informationen beurteilen und Feedback geben. Automatisierte Benachrichtigungen an zentralen Stellen (z.B. Förderungszusage, Datenhinterlegung, Berichterstattung) können zur Unterstützung eingesetzt werden. Erleichtert die Programmentwicklung für Beratungs- und Unterstützungsleistungen.

Miksa et al. entwickelten einen Leitfaden von zehn Prinzipien für den Einsatz von maschinenlesbaren DMP.⁴⁴ Zwei wesentliche Bausteine sind standardisierte Vokabulare und ein einheitliches Datenmodell. Damit wird sowohl ein gemeinsames Verständnis über die Inhalte zwischen verschiedenen Stakeholdern als auch Interoperabilität zwischen Werkzeugen und Services möglich. Das Datenmodell sollte modular aufgebaut sein und eine nachnutzbare, strukturierte Repräsentation maschinenlesbarer Informationen zum FDM darstellen.⁴⁵ Die oben genannten Beispiele RDMO und der Data Stewardship Wizard entwickeln sich entlang dieser Linien.

3.2 Vernetzung und Zusammenarbeit

Wie kann eine Infrastruktureinrichtung auf die dynamische Entwicklung des Bereichs FDM reagieren? An zahlreichen Hochschulen gibt es zurzeit Initiativen zum Aufbau von Kompetenzzentren,⁴⁶ bei denen institutionelle DMP-Angebote aufgebaut werden. Diese werden durch Beratungs- und Schulungsangebote ergänzt. Häufig arbeiten Bibliothek, Rechenzentrum und der Drittmittel-Service der Universität zusammen, um weite Bereiche des Datenlebenszyklus abzudecken. Zur besseren Vernetzung, dem Erfahrungsaustausch und dem Finden von Synergien wurde eine

⁴⁴ Vgl. Miksa et al. 2019.

⁴⁵ Die Research Data Alliance hat hierzu eine Arbeitsgruppe gebildet, s. <https://www.rd-alliance.org/groups/dmp-common-standards-wg>.

⁴⁶ Vgl. z. B. die Liste von FDM-Kontakten auf [forschungsdaten.org](https://www.forschungsdaten.org/index.php/FDM-Kontakte), s. <https://www.forschungsdaten.org/index.php/FDM-Kontakte>.

Unterarbeitsgruppe der DINI/nestor AG Forschungsdaten⁴⁷ zu diesem Thema gegründet, die hilfreiche Informationen aus Sicht der Infrastrukturanbieter liefert.⁴⁸ Komplementär hat sich eine deutschlandweite RDMO-Community und auf Ebene von Landesinitiativen (z. B. in NRW⁴⁹) thematische Arbeitsgruppen gebildet. Hier werden z. B. arbeitsteilig bzw. entsprechend der Kompetenzen einzelner Einrichtungen spezifische Fragenkataloge erstellt und geteilt. Es zeigt sich, dass DMP-Werkzeuge häufig ausschließlich im Beratungskontext eingesetzt werden. D. h., dass Forschende noch nicht selbständig DMP erstellen, sondern oft auf Unterstützung der Infrastrukturanbieter angewiesen sind. Insbesondere ist bei der Einführung maschinenlesbarer DMP-Werkzeuge ein teils erheblicher technischer infrastruktureller Aufwand nötig. Nicht zuletzt ist die Frage danach, was einen guten DMP ausmacht, nicht einfach zu beantworten.⁵⁰ Hier ist also noch ein Lern- und Erfahrungsaufbau vonnöten. Ein äußerer Trigger, der den Kompetenzaufbau fördert, ist, dass einige Förderprogramme dazu übergehen, in regelmäßigen Abständen – die z. T. mit üblichen Berichtszeiträumen zusammenfallen – aktualisierte DMP einzufordern (z. B. BMBF, Horizon 2020).⁵¹

Damit Forschungsdaten, ihre Genese und der Kontext verstanden werden können, ist die Beschreibung der Forschungsdaten und die Dokumentation der Prozesse, die mit Daten arbeiten, eine wesentliche Voraussetzung. Neben der Beschreibung der Forschungsdaten sollten im DMP zusätzlich die Dokumentationsabläufe mit den dazugehörigen Rollen und Verantwortlichkeiten beschrieben werden. Im folgenden Kapitel wird darauf näher eingegangen.

4 Beschreibung und Dokumentation

Forschungsdaten können unter mehreren Gesichtspunkten (Anforderungen) charakterisiert werden. Dazu zählen u. a. Datentypen, Formate, Organisation, Herkunft, Authentizität, Qualität, Kontext. Diese Informationen helfen den Forschenden, die

47 S. <https://dini.de/ag/dininestor-ag-forschungsdaten/>. DINI ist die Deutsche Initiative für Netzwerkinformation e. V. und nestor ist ein Kooperationsverbund zum Thema digitale Langzeitarchivierung, s. <https://www.langzeitarchivierung.de/>.

48 S. https://www.forschungsdaten.org/index.php/UAG_Datenmanagementpl%C3%A4ne.

49 S. <https://www.fdm.nrw/index.php/fdm-nrw/dmp/>.

50 Vgl. den Workshop „Supporting and reviewing Data Management Plans“, IDCC 2016, insbesondere „Analysing DMPs to inform and empower academic librarians in providing research data support: lessons from the DART project“ von A. Whitmire, s. <https://www.dcc.ac.uk/events/workshops/supporting-and-reviewing-data-management-plans>. Die Teammitglieder haben jeweils eine Vielzahl von DMP analysiert und die Ergebnisse dann miteinander verglichen.

51 S. a. Beitrag von Putnings, Kap. 1.3 in diesem Praxishandbuch.

Daten und ihren Kontext einzuordnen bzw. zu bewerten, und ermöglichen es digitalen Werkzeugen (z. B. Analyseprogramme, virtuelle Forschungsumgebungen), Daten verarbeiten zu können oder Aufgaben der längerfristigen Verfügbarmachung (Veröffentlichung in einem Repositorium, Archivierung) zu übernehmen. Diese charakterisierenden Daten werden als sog. Metadaten⁵² den eigentlichen Forschungsdaten hinzugefügt. Sie sind sozusagen strukturierte Informationen über eine digitale Ressource. Metadaten sind für alle Phasen des Lebenszyklus von Forschungsdaten relevant. Damit wird das generelle Verständnis und das Nachnutzungspotenzial der Forschungsdaten erhöht.⁵³

Ein digitales Objekt kann durch drei Ebenen charakterisiert werden:⁵⁴

- Als physisches Objekt sind es die Bits, die auf einem physischen Trägermedium (z. B. einem magnetischen oder optischen Speicher) gespeichert sind.
- Als logisches Objekt, das von einer Software erkannt und verarbeitet werden kann (z. B. ASCII Text, Dokument einer Textverarbeitungssoftware, aber auch zusammengesetzte Objekte).
- Als konzeptionelles Objekt, das von einer Person oder einer Software erkannt und verstanden werden kann (z. B. ein Buch, eine Karte, ein Foto).

4.1 Arten von Metadaten

Ausgehend von dieser Vorstellung lassen sich Metadaten prinzipiell in folgende vier Kategorien aufteilen:

- deskriptive Metadaten,
- administrative Metadaten,
- technische Metadaten und
- strukturelle Metadaten.⁵⁵

Tab. 2 beschreibt die vier Kategorien näher und gibt Beispiele.

⁵² Vgl. „Data about data“ in Gartner 2016, 6–8.

⁵³ Man könnte fast so weit gehen zu sagen, dass Forschungsdaten ohne Metadaten nicht interpretierbar sind (z. B. eine Zeitreihe von Temperaturmessungen ohne Informationen darüber, dass es sich um eine Zeitreihe oder um welche Art von Temperaturen es sich handelt).

⁵⁴ Vgl. Thibodeau 2002, 6–10.

⁵⁵ Vgl. z. B. Corrado und Sandy 2017, 64–74; Gartner 2016, 8; Caplan 2003, 158–166.

Tab. 2: Grundlegende Kategorien von Metadaten⁵⁶

Art von Metadaten	Beschreibung	Beispiele
Beschreibende Metadaten	Inhaltliches Objekt: Beschreibung von Objekteigenschaften zum Auffinden, Identifizieren, Auswählen und Zugreifen	Titel, Autorin/Autor oder Erzeugerin/Erzeuger, Name, Thema, Identifier (DOI, ORCID)
Administrative Metadaten	Verwendung von Inhaltselementen, Dateien und administrativen Metadaten: Informationen über Rechte, Provenienz dokumentieren; für die Archivierung relevante Informationen bereitstellen; Informationen über die Erzeugung von Surrogaten	Urheberrecht, Zugriffsrechte, Lizenzen usw.
Technische Metadaten	Elektronische Datei: Dokumentation von Informationen, die nötig sind um auf das beschriebene Element zuzugreifen	Dateigröße, Dateiformat, Datenkompression, Prüfsummen usw.
Strukturelle Metadaten	Satz von elektronischen Dateien Dokumentation der inneren Struktur des Elements; zeichnet Informationen über Beziehungen zu anderen Elementen auf; beschreibt die physische und logische Struktur eines digitalen Objektes	Beziehung zu anderen Dateien; Beschreibung von Tabellen und Views in einer relationalen Datenbank

Zwischen den Kategorien gibt es eine gewisse Überlappung, sodass manchmal Kategorien zusammengefasst oder erweitert werden⁵⁷ (z. B. nennen Treloar und Wilkinson die zusätzlichen Kategorien Provenienz- und Archivierungsmetadaten⁵⁸). Metadaten bestehen aus drei grundlegenden Komponenten:⁵⁹

- Eine semantische Komponente, die die Bedeutung der Felder bzw. Elemente beschreibt,
- eine syntaktische, die der Art der Kodierung der Metadaten entspricht (z. B. Tabelle, XML usw.) und

⁵⁶ Vgl. Tab. 5.1. in Corrado und Sandy 2017, 66.

⁵⁷ Vgl. Gartner 2016; Treloar und Wilkinson 2008.

⁵⁸ Vgl. Treloar und Wilkinson 2008, 784.

⁵⁹ Vgl. Gartner 2016, 53.

- inhaltliche Regeln, d. h. welche Werte bzw. Inhalte sind als Metadatenelemente erlaubt bzw. gefordert.

Ein Beispiel aus der Astronomie soll dies illustrieren.⁶⁰ Beispielsweise sollen in einem Foto von einer bestimmten Region am Nachthimmel weitere astronomische Objekte aufgefunden werden (s. Abb. 3).

Abb. 3: Suche in einem bestimmten Bereich des Nachthimmels nach astronomischen Objekten. Hier mittels des Werkzeuges Aladin (s. <https://aladin.u-strasbg.fr/>)

Hierbei geht es darum, das Bild des Nachthimmels mit den Positionen von bekannten Objekten zu überlagern bzw. zu markieren. Hierfür müssen die Koordinaten/Positionen in dem Bild bekannt sein, sodass ein Crossmatch mit einem Querkatalog vorgenommen werden kann. Positionen am Nachthimmel werden üblicherweise in Rektaszension und Deklination angegeben.⁶¹ Hierfür sind die Metadatenfelder RA und DE vorgesehen (unterer Bereich in Abb. 3). Da es mehrere Koordinatensysteme zur Beschreibung von Positionen am Himmel gibt, muss eindeutig sein, welches System hier benutzt wird. Um die Beschreibung von astronomischen Beobachtungen

⁶⁰ Für Beispiele aus dem Bereich der Geisteswissenschaften siehe Gartner 2016, 53–63.

⁶¹ S. https://de.wikipedia.org/wiki/%C3%84quatoriales_Koordinatensystem.

gen zu vereinheitlichen und Interoperabilität zu ermöglichen, wurden von der IVOA Metadatenstandards entwickelt.⁶² Das Beispiel aus Abb. 3 verweist in der Beschreibung der Koordinaten auf diesen Standard (UCD: *pos.eq.ra*, *pos.eq.dec*). UCD bedeutet Unified Content Descriptor. Hierbei handelt es sich um ein formales Vokabular für astronomische Daten.⁶³ Abbildung 4 zeigt einen maschinenlesbaren XML-Ausschnitt für diese Datenfelder.

```
<FIELD name="s_ra" datatype="double" ucd="pos.eq.ra" unit="deg"
  utype="obscore:Char.SpatialAxis.Coverage.Location.Coord.Position
  2D.Value2.C1" xtype="adql:DOUBLE">
  ...<DESCRIPTION>RA of central coordinates</DESCRIPTION>
</FIELD>
<FIELD name="s_dec" datatype="double" ucd="pos.eq.dec"
  unit="deg"
  utype="obscore:Char.SpatialAxis.Coverage.Location.Coord.Position
  2D.Value2.C2" xtype="adql:DOUBLE">
  ...<DESCRIPTION>DEC of central coordinates</DESCRIPTION>
</FIELD>
```

Abb. 4: XML-Ausschnitt für die Position von astronomischen Objekten basierend auf Metadatenstandards der IVOA (s. <http://ivoa.net/documents/ObsCore/index.html>)

Dabei ist noch die Bedeutung von (physikalischen) Einheiten (hier *deg*) und Datenformaten (hier *adql:DOUBLE*) zu betonen.

Heute sind Metadaten in maschinenlesbarer Form oftmals in Form von XML⁶⁴ kodiert. Inhaltliche Regeln sind oft in Form von kontrollierten Vokabularen definiert (z. B. der Getty Thesaurus von geografischen Namen⁶⁵ oder das Basel Register of Thesauri, Ontologies & Classifications⁶⁶).

Standardisierte Metadaten unterstützen bzw. ermöglichen die Interoperabilität, Wiederverwendbarkeit, Transparenz und Integrationsfähigkeit von Daten und digitalen Werkzeugen. Standards werden meist innerhalb einer Community⁶⁷ entwickelt und etabliert.⁶⁸ Es gibt eine Reihe recht generischer Metadatenstandards wie z. B. Dublin Core⁶⁹ (Beschreibung von Dokumenten und anderen Objekten im Internet),

⁶² International Virtual Observatory Alliance, s. <http://ivoa.net/documents/ObsCore/index.html>.

⁶³ S. <http://ivoa.net/documents/latest/UCD.html>.

⁶⁴ eXtensible Markup Language, s. <https://www.w3.org/standards/xml/>.

⁶⁵ S. <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>.

⁶⁶ S. <http://www.bartoc.org>.

⁶⁷ Dies können durchaus unterschiedliche Communities sein, z. B. eine wissenschaftliche Community, eine bibliothekarische, eine wirtschaftliche, eine technische usw.

⁶⁸ S. z. B. <https://phaidra.univie.ac.at/o:441215>.

⁶⁹ S. <https://dublincore.org/>.

DataCite⁷⁰ (Publikation und Zitation von Forschungsdaten), RDF⁷¹ (Beschreibung von Webressourcen) usw. Fachspezifische Metadatenstandards finden sich z. B. in den Sozialwissenschaften (DDI⁷²), in den Geisteswissenschaften (TEI⁷³), in den Geowissenschaften (ISO 19115⁷⁴), in der Astrophysik (IVOA-Standards⁷⁵), um einige zu nennen. Die Webseite von Jenn Riley zeigt einen Überblick über existierende Standards auf dem Stand von 2010.⁷⁶ Die Abstimmungsprozesse für Standards innerhalb einer Community sind oft aufwendig und langwierig, was sowohl die Adaption von fremden Standards als auch die Entwicklung von community-übergreifenden Standards erschwert.⁷⁷ An dieser Stelle können jedoch Infrastruktureinrichtungen wie wissenschaftliche Bibliotheken oder Rechenzentren Serviceangebote entwickeln, was zumindest das Auffinden und die Adaption von Standards für Forschungsprojekte betrifft.⁷⁸

4.2 FAIR-Prinzipien

Eine Grundlage für die Umsetzung der FAIR-Prinzipien stellen maschinenlesbare Metadaten dar. Diese erlauben z. B. die eindeutige Referenzierbarkeit (z. B. DOI,⁷⁹ ORCID⁸⁰), die Nachnutzung (Lizenzbedingungen) und Interoperabilität (Daten- und Dateiformate) von Forschungsdaten.⁸¹

Die Umsetzung ist allerdings nicht ganz einfach, insbesondere für Forschende, die sich vermehrt mit entsprechenden Anforderungen seitens der Förderorganisationen konfrontiert sehen.⁸² Bei den FAIR-Prinzipien handelt es sich um einen Leitfaden und nicht um einen Standard. Zurzeit gibt es noch kein FAIR-Ökosystem, das einfach anzuwendende Praktiken und Werkzeuge für die alltägliche Arbeit bietet.⁸³ Es lässt sich relativ einfach testen, wie FAIR die eigenen Daten und Services sind. Aber entsprechende Best Practices, wie diese dann FAIRer gemacht werden können, existieren derzeit nur sporadisch. Doch ist das Feld sehr agil und es gibt eine Reihe

⁷⁰ S. <https://datacite.org/>.

⁷¹ S. <https://www.w3.org/2001/sw/wiki/RDF>.

⁷² S. <https://ddialliance.org>.

⁷³ S. <https://tei-c.org>.

⁷⁴ S. <http://www.dcc.ac.uk/resources/metadata-standards/iso-19115>.

⁷⁵ S. <http://ivoa.net/documents>.

⁷⁶ S. <http://jennriley.com/metadatamap>.

⁷⁷ Vgl. Gartner 2016, 37–39.

⁷⁸ Vgl. z. B. Curdt et al. 2019; Tenopir et al. 2017.

⁷⁹ Digital Object Identifier, s. <https://www.doi.org/>.

⁸⁰ Open Researcher and Contributor Identifier, s. <https://orcid.org/>.

⁸¹ S. a. Beitrag von Linne et al., Kap. 3.2 in diesem Praxishandbuch; Wilkinson et al. 2016.

⁸² S. a. Beitrag von Putnings, Kap. 1.3 in diesem Praxishandbuch.

⁸³ Vgl. Thompson et al. 2019.

von Aktivitäten zu praktischen Umsetzungen.⁸⁴ Insbesondere zeigen diese Ansätze Wege auf, wie sich schrittweise Daten und Dienste FAIRer gestalten lassen. Da diese z. T. sehr infrastruktur- bzw. IT-lastig sind, bildet auch hier die Kooperation mit Infrastruktureinrichtungen einen Ansatz für eine Arbeitsteilung mit verteilten Rollen, Aufgaben und Verantwortlichkeiten.⁸⁵

4.3 Workflows

Ein wesentlicher Aspekt für die praktische Umsetzung des Einsatzes von Metadaten ist die Frage, wann welche Metadaten von wem und auf welche Art und Weise erfasst werden. Generell lässt sich empfehlen, Metadaten möglichst zeitnah (zeitgleich) mit den entsprechenden Arbeitsschritten, bei denen Forschungsdaten erzeugt oder verarbeitet werden, aufzunehmen. Entsprechende Zeitpunkte sind z. B. die semiautomatische Erzeugung von Metadaten bei Experimenten. Auch sind die Kurationsgrenzen des Data Curation Continuum⁸⁶ (private → Kollaborations- → öffentliche Domäne) Zeitpunkte im Forschungsprozess, zu denen entsprechend der Anforderungen der neuen Domäne gewisse Metadaten angereichert werden sollten. Eine retrospektive Erfassung von Metadaten birgt immer die Gefahr der Unvollständigkeit, weil relevante Informationen fehlen, keine Zeit (Motivation) oder finanzielle/personelle Ressourcen dafür vorhanden sind. Um die Hürde bzw. den Aufwand für Forschende zu minimieren, können lokale Supportangebote seitens der Infrastruktureinrichtungen oder entsprechend geschulten Personals in den Forschungsvorhaben entwickelt und genutzt werden. Hierzu gehören

- Beratung durch professionelle Serviceeinrichtungen (z. B. Drittmittel-Services) bei der anfänglichen Planung von Forschungsvorhaben,
- Schulungsangebote durch die Infrastruktureinrichtungen (Kompetenzaufbau der Forschenden),
- Bereitstellung/Entwicklung/Nachnutzung von Werkzeugen, die die zeitnahe Erfassung von Metadaten ermöglichen/vereinfachen,
- Personal, das die Kuratierung übernimmt (z. B. Datenmanager oder Datenmanagerin).

⁸⁴ Z. B. Thompson et al. 2019; Griffin et al. 2018; Mons et al. 2017; Wilkinson et al. 2017.

⁸⁵ Vgl. Schirrwagen et al. 2019.

⁸⁶ Vgl. Treloar & Klump 2019.

Für die Umsetzung gibt es mehrere Möglichkeiten.⁸⁷ Im Bereich des Long Tail⁸⁸ können Forschungsvorhaben universelle Werkzeuge⁸⁹ für die Datenbeschreibung und Schnittstellen zu interdisziplinären Forschungsdatenrepositorien (z. B. Zenodo⁹⁰) angeboten werden. Für größere Projekte mit spezifischeren Anforderungen lassen sich spezialisierte Arbeitsabläufe entwickeln. Die Werkzeuge, die zur Verfügung stehen und deren Nutzbarkeit sind auch entscheidend dafür, wieviel Engagement Forschende in das FDM investieren.

Bei der Betrachtung der Kategorien von Metadaten, lassen sich entlang des Forschungsprozesses ebenfalls Zuordnungen zu den in Tab. 1 genannten Rollen herstellen. Forschende werden im Wesentlichen für die beschreibenden Metadaten verantwortlich sein, denn diese erlauben die Identifizierung sowie die Referenzierung und stellen den Kontext her. Auch im Bereich der administrativen Metadaten, insbesondere im Bereich der Provenienz, d. h. die Dokumentation der Forschungs- bzw. Verarbeitungsprozesse, werden sich Forschende mit einbringen müssen, weil hier disziplinspezifische Kenntnisse erforderlich sind. Die anderen Kategorien von Metadaten werden in Abhängigkeit von den lokalen organisatorischen Gegebenheiten eher im Bereich Forschungsunterstützung oder Infrastruktureinrichtungen verortet sein. Ein enger Austausch zwischen den Agierenden ist wichtig, damit entsprechende Metadaten im Rahmen einer Datenstrategie aus der einen Phase des Projektes in die nächste Phase migriert bzw. transformiert werden können, z. B. wenn es darum geht gewisse Daten und Metadaten für die Langzeitarchivierung vorzubereiten.⁹¹

Ein mögliches Szenario für die Erfassung von Metadaten in einem Forschungsprojekt könnte sein, dass globale Metadaten, die das Projekt beschreiben (Namen der Forschenden, Projekttitel, Vorhabenbeschreibung usw.) sowie eine erste Iteration der Beschreibung von anfallenden Forschungsdaten, Software usw., in der Planungs- bzw. Antragsphase erfasst und dokumentiert werden. Hier können Beratung, Schulung und Werkzeuge zur Erstellung von DMP, die z. B. durch die zentralen professionellen Unterstützungsstrukturen an Hochschulen bereitgestellt werden, helfen. Eine sinnvolle Strategie zur Beschreibung könnte nach Michener wie folgt aussehen:⁹²

- Identifizierung der Arten von Informationen, die benötigt werden, die Forschenden erlauben sollen, die Daten zu finden, auf die Daten zuzugreifen, die Daten zu interpretieren und zu zitieren.

⁸⁷ Vgl. Ribeiro et al. 2018.

⁸⁸ Vgl. Borgman et al. 2016.

⁸⁹ S. z. B. <https://library.stanford.edu/research/data-management-services/data-best-practices/creating-metadata/metadata-tools>.

⁹⁰ S. <https://zenodo.org>.

⁹¹ S. a. Beitrag von Weber und Piesche, Kap. 4.2 in diesem Praxishandbuch.

⁹² Vgl. Michener 2015, 4.

- Herausfinden, ob es fachspezifische Metadatenschemas oder Standards gibt, die zusammen mit kontrollierten Vokabularen angewendet werden können.
- Identifizierung von Softwarewerkzeugen, die eingesetzt werden können, um Metadaten zu erzeugen und zu managen. Zusätzlich können z. B. Textdateien (readme), welche relevante Metadaten enthalten, hinzugezogen werden oder sogar in Dateihedern integriert werden.

Außerdem macht es Sinn, eine verantwortliche Person zu benennen, die alle Projektdetails bspw. in einem elektronischen Laborbuch verwaltet. Hier könnte ein Prozess aufgesetzt werden, in dem in regelmäßigen Abständen ein weiteres Team-Mitglied die Aufzeichnungen überprüft. Die Metadaten können dann zu Datenprodukten hinzugefügt werden, die gespeichert, nachgenutzt und geteilt werden. Ein weiterer wichtiger Aspekt ist die Nutzung offener, nicht-proprietärer Datenformate und Werkzeuge. Bei nicht-offenen Formaten handelt es sich dann oft um „Black Boxes“, die nur innerhalb der kommerziellen Umgebung nutzbar sind. Beim Export in offene Formate gehen dann meist wesentliche Metadaten verloren.

Komplexer stellt sich die Situation für die Dokumentation während der aktiven Forschungsphase dar, die oftmals iterativ und „on-the-fly“ (weiter-)entwickelt und durchgeführt wird (vgl. Abb. 1). Wie bereits erwähnt könnte eine Zusammenarbeit zwischen Forschenden und Infrastruktureinrichtungen innerhalb von Projekten stattfinden. Neben der Arbeits- und Aufgabenteilung besteht für die Infrastruktureinrichtungen die Möglichkeit des Kompetenzaufbaus bzw. -ausbaus und der Entwicklung einer nachhaltigen (lokalen) Infrastruktur von relevanten und nachgefragten Werkzeugen (z. B. mittels eingebetteten Personals⁹³). Über das konkrete Projekt hinaus können Forschende am lokalen Standort zukünftig von diesen Werkzeugen profitieren. Hierbei ist zu beachten, dass aufgrund der diversen Arbeitskulturen immer auch eine Einpassung in etablierte Forschungsprozesse vonnöten ist.⁹⁴ Eine Adaption von Forschungsprozessen an neue Technologien scheint am ehesten schrittweise erreichbar zu sein, um mögliche Widerstände abzufedern. Das Konzept der Rampen von Atkinson et al.⁹⁵ beschreibt in diesem Zusammenhang, wie ausgehend von aktuellen Arbeitspraktiken, aktuell genutzten (digitalen) Werkzeugen usw. mit kleinen in sich logischen Schritten Anpassungen der Arbeitspraktiken vorgenommen werden. Dabei sollen detaillierte technische Aspekte zunächst möglichst ausgeblendet werden. Ein Beispiel könnte die Nutzung eines Cloudspeichers wie z. B. sciebo⁹⁶ sein. Die Client-Software lässt sich nahtlos in das lokale Dateimanagement integrieren, sodass das bekannte Aussehen des Dateixplorers erhalten bleibt. Der

⁹³ Vgl. Cremer, Engelhardt und Neuroth 2015, 15–17.

⁹⁴ Vgl. z. B. Meyer und Schroeder 2015, 197–208.

⁹⁵ Vgl. Atkinson et al. 2010.

⁹⁶ S. <https://www.sciebo.de/>.

Client erlaubt dann aber Dateien über mehrere Endgeräte hinweg zu synchronisieren und mit Kooperationspartnern teilen.

Im Sinne der Qualitätssicherung entlang des Datenlebenszyklus spielt die Qualität der Metadaten eine wichtige Rolle. Die Qualität der in ihnen enthaltenen Informationen bestimmt die (Nach-)Nutzbarkeit der Forschungsdaten. Die Qualität von Metadaten kann unter folgenden Gesichtspunkten betrachtet werden: Vollständigkeit, Genauigkeit, Provenienz, Erwartungskonformität, logische Konsistenz und Kohärenz, Aktualität und Zugänglichkeit.⁹⁷

Fazit

Die Planung, Beschreibung und Dokumentation beim Umgang mit Forschungsdaten ist ein komplexes Feld mit vielen Stakeholdern. Die Zusammenarbeit zwischen Forschenden, professionellen Dienstleistern und Infrastruktureinrichtungen hat das Potenzial, verschiedene Expertisen zusammenzubringen, um insbesondere so für die Forschenden eine bedarfsgerechte, effektive und nachhaltige Arbeitsumgebung zu ermöglichen.

Zurzeit lässt sich eine gewisse Stagnation bzw. eine Verschiebung der Schwerpunkte bei der Entwicklung und Bereitstellung von Informationsinfrastrukturen für die Dissemination und Vernetzung von Forschungsergebnissen beobachten.⁹⁸ Hinzu kommt der große Bedarf an Datenkompetenzen in allen Stakeholdergruppen.⁹⁹ Solange größere Initiativen wie z. B. die Nationale Forschungsdateninfrastruktur (NFDI),¹⁰⁰ die European Open Science Cloud (EOSC)¹⁰¹ oder GO-FAIR¹⁰² sich gerade finden oder noch in der Entwicklung befinden, scheint es durchaus sinnvoll, sich lokal bzw. in Projekten, ganz im Sinne von Atkinson et al.,¹⁰³ schrittweise weiterentwickelnden Beschreibungs- und Dokumentationsworkflows zu nähern.¹⁰⁴ Dann wird es ebenfalls einfacher, unterschiedliche Bedarfe bzw. Anforderungen wie Dokumentation zur internen Verwendung, Dokumentation und Kuratierung für eine Veröffentlichung von Forschungsdaten zu erfüllen bzw. ineinander zu überführen.

⁹⁷ Der Beitrag von Király und Brase, Kap. 4.3 in diesem Praxishandbuch geht näher darauf ein.

⁹⁸ Vgl. Neuroth 2019, 167–169.

⁹⁹ Vgl. z. B. Fournier 2017.

¹⁰⁰ S. <https://www.dfg.de/foerderung/programme/nfdi>.

¹⁰¹ S. <https://www.eosc-portal.eu>.

¹⁰² S. <https://www.go-fair.org>.

¹⁰³ Vgl. Atkinson et al. 2010.

¹⁰⁴ Vgl. Mons et al. 2017, 53.

Literatur

Letztes Abrufdatum der Internet-Dokumente ist der 15.11.2020.

- Atkinson, Malcolm, David De Roure, Jano van Hemert und Danus Michaelides. 2010. „Shaping Ramps for Data-Intensive Research.“ Conference gehalten auf der UK e-Science All Hands Meeting 2010, Juni. <http://eprints.soton.ac.uk/id/eprint/271235>.
- Borgman, Christine L. 2015. *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT press.
- Borgman, Christine L. 2019. „The Lives and After Lives of Data.“ *Harvard Data Science Review*, Juni. doi:10.1162/99608f92.9a36bdb6.
- Borgman, Christine L., Milena S. Golshan, Ashley E. Sands, Jillian C. Wallis, Rebekah L. Cummings, Peter T. Darch und Bernadette M. Randles. 2016. „Data Management in the Long Tail: Science, Software, and Service.“ *International Journal of Digital Curation* 11 (1): 128–149. doi:10.2218/ijdc.v11i1.428.
- Boulton, Geoffrey, Philip Campbell, Brian Collins, Peter Elias, Wendy Hall, Graeme Laurie, Onora O'Neill et al. 2012. „Science as an Open Enterprise.“ The Royal Society. <http://royalsociety.org/policy/projects/science-public-enterprise/report/>.
- Brand, Ortrun und Jens Dierkes. 2020. „Failures and major issues“. Bausteine Forschungsdatenmanagement, in press
- Büttner, Stephan, Hans-Christoph Hobohm und Lars Müller, Hg. 2011. *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen.
- Caplan, Priscilla. 2003. *Metadata Fundamentals for All Librarians*. Chicago: American Library Association.
- Corrado, Edward M. und Heather Moulaison Sandy. 2017. *Digital Preservation for Libraries, Archives, and Museums*. Lanham: Rowman & Littlefield.
- Cox, Andrew und Eddy Verbaan. 2018. *Exploring Research Data Management*. London: Facet Publishing.
- Cremer, Fabian, Claudia Engelhardt und Heike Neuroth. 2015. „Embedded Data Manager – Integriertes Forschungsdatenmanagement: Praxis, Perspektiven und Potentiale.“ *Bibliothek Forschung und Praxis* 39 (1): 13–31.
- Curdtt, Constanze, Dirk Hoffmeister, Tanja Kramm, Ulrich Lang und Georg Bareth. 2019. „Etablierung von Forschungsdatenmanagement-Services in geowissenschaftlichen Sonderforschungsbereichen am Beispiel des SFB/Transregio 32, SFB 1211 und SFB/Transregio 228.“ *Bausteine Forschungsdatenmanagement*, 2 (September): 61–67. doi:10.17192/bfdm.2019.2.8103.
- Fournier, Johannes. 2017. „Zum qualifizierten Umgang mit Forschungsdaten. Ein Bericht über den Workshop ‚Wissenschaft im digitalen Wandel‘ am 6. Juni 2017 in der Universität Mannheim.“ *o-bib. Das offene Bibliotheksjournal* 4 (3): 88–93. doi:10.5282/o-bib/2017H3S88-93.
- Gartner, Richard. 2016. *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*. New York: Springer.
- Goodman, Alyssa, Alberto Pepe, Alexander W. Blocker, Christine L. Borgman, Kyle Cranmer, Merce Crosas, Rosanne Di Stefano u. a. 2014. „Ten Simple Rules for the Care and Feeding of Scientific Data.“ *PLoS Comput Biol* 10 (4): e1003542. doi:10.1371/journal.pcbi.1003542.
- Griffin, Philippa C., Jyoti Khadake, Kate S. LeMay, Suzanna E. Lewis, Sandra Orchard, Andrew Pask, Bernard Pope u. a. 2018. „Best practice data life cycle approaches for the life sciences.“ *F1000Research* 6 (Juni). doi:10.12688/f1000research.12344.2.
- Jahnke, Lori, Andrew Asher und Spencer DC Keralis. 2012. „The problem of data.“ http://digitalcommons.bucknell.edu/fac_pubs/52/.

- Klar, Jochen und Harry Enke. 2013. „Organisation und Struktur.“ DFG-Projekt RADIESCHEN – Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur. Konsortien der DFG-Projekte Radieschen, EWIG, re3data.org, KomFor und BoKeLa. doi:10.2312/RADIESCHEN_005.
- Leendertse, Jan, Susanne Mocken und Dirk von Suchodoletz. 2019. „Datenmanagementpläne zur Strukturierung von Forschungsvorhaben.“ *Bausteine Forschungsdatenmanagement* 2 (Mai): 4–9. doi:10.17192/bfdm.2019.2.8003.
- Ludwig, Jens und Harry Enke. 2013. „Leitfaden zum Forschungsdaten-Management“. <http://www.univerlag.uni-goettingen.de/handle/3/isbn-978-3-86488-032-2>.
- Lyon, Dr Liz. 2007. „Dealing with Data: Roles, Rights, Responsibilities and Relationships. Consultancy Report.“ Bath: UKOLN, University of Bath.
- Meyer, Eric T. und Ralph Schroeder. 2015. *Knowledge machines: Digital transformations of the Sciences and Humanities*. Cambridge, MA: MIT Press.
- Michener, William K. 2015. „Ten Simple Rules for Creating a Good Data Management Plan.“ *PLOS Comput Biol* 11 (10): e1004525. doi:10.1371/journal.pcbi.1004525.
- Miksa, Tomasz, Stephanie Simms, Daniel Mietchen und Sarah Jones. 2019. „Ten Principles for Machine-Actionable Data Management Plans.“ *PLOS Computational Biology* 15 (3): e1006750. doi:10.1371/journal.pcbi.1006750.
- Mons, Barend, Cameron Neylon, Jan Velterop, Michel Dumontier, da Silva Santos, Luiz Olavo Bonino und Mark D. Wilkinson. 2017. „Cloudy, Increasingly FAIR; Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud“. *Information Services & Use* 37 (1): 49–56. doi:10.3233/ISU-170824.
- Neuroth, Heike. 2019. „Haben Metadaten ihre Unschuld verloren?“ In *Bibliotheksentwicklung im Netzwerk von Menschen, Informationstechnologie und Nachhaltigkeit: Festschrift für Achim Oßwald*. Bad Honnef: Bock + Herchen. https://www.th-koeln.de/mam/downloads/deutsch/studium/studiengaenge/f03/bib_inf_ma/festschrift_osswald.pdf.
- Neuroth, Heike, Claudia Engelhardt, Jochen Klar, Jens Ludwig und Harry Enke. 2018. „Aktives Forschungsdatenmanagement.“ *ABI Technik* 38 (1): 55–64. doi:10.1515/abitech-2018-0008.
- Pampel, Heinz, Roland Bertelmann und Hans-Christoph Hobohm. 2010. „Data Librarianship“ – *Rollen, Aufgaben, Kompetenzen*. Bd. 144. Berlin: RatSWD Working Paper Series
- Ribeiro, Cristina, João Rocha da Silva, João Aguiar Castro, Ricardo Carvalho Amorim, João Correia Lopes und Gabriel David. 2018. „Research Data Management Tools and Workflows: Experimental Work at the University of Porto.“ *IASSIST Quarterly* 42 (2): 1–16. doi:10.29173/iq925.
- Schirrwagen, Jochen, Philipp Cimiano, Vidya Ayer, Christian Pietsch, Cord Wiljes, Johanna Vompras und Dirk Pieper. 2019. „Expanding the Research Data Management Service Portfolio at Bielefeld University According to the Three-Pillar Principle Towards Data FAIRness.“ *Data Science Journal* 18 (1): 6. doi:10.5334/dsj-2019-006.
- Tenopir, Carol, Sanna Talja, Wolfram Horstmann, Elina Late, Dane Hughes, Danielle Pollock, Birgit Schmidt, Lynn Baird, Robert Sandusky und Suzie Allard. 2017. „Research Data Services in European Academic Research Libraries.“ *LIBER Quarterly* 27 (1/14. Februar): 23–44. doi:10.18352/lq.10180.
- Thibodeau, Kenneth. 2002. „Overview of technological approaches to digital preservation and challenges in coming years.“ *The state of digital preservation: an international perspective*, Marcum, Deanna, Kenneth Thibodeau, Margaret Hedstrom, Meg Bellinger, Laura Campbell, Titia van der Werf, Colin Webb, Donald Waters, 4–31. Washington: Council on Library and Information Resources.
- Thompson, Mark, Kees Burger, Rajaram Kaliyaperumal, Marco Roos und Luiz Olavo Bonino da Silva Santos. 2019. „Making FAIR Easy with FAIR Tools: From Creolization to Convergence.“ *Data Intelligence* November: 87–95. doi:10.1162/dint_a_00031.

- Treloar, Andrew, David Groenewegen und Cathrine Harboe-Ree. 2007. „The Data Curation Continuum: Managing Data Objects in Institutional Repositories.“ *D-Lib Magazine* 13 (9/10). doi:10.1045/september2007-treloar.
- Treloar, Andrew und Jens Klump. 2019. „Updating the Data Curation Continuum“. *International Journal of Digital Curation* 14 (September): 87–101. doi:10.2218/ijdc.v14i1.643.
- Treloar, Andrew und Ross Wilkinson. 2008. „Rethinking Metadata Creation and Management in a Data-Driven Research World.“ In *2008 IEEE Fourth International Conference on eScience*, 782–789. Indianapolis: IEEE. doi:10.1109/eScience.2008.41.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg u. a. 2016. „The FAIR Guiding Principles for Scientific Data Management and Stewardship.“ *Scientific Data* 3 (März): 160018. doi:10.1038/sdata.2016.18.
- Wilkinson, Mark D., Ruben Verborgh, Luiz Olavo Bonino da Silva Santos, Tim Clark, Morris A. Swertz, Fleur D. L. Kelpin, Alasdair J. G. Gray u. a. 2017. „Interoperability and FAIRness through a Novel Combination of Web Technologies.“ *PeerJ Computer Science* 3 (April): e110. doi:10.7717/peerj-cs.110.