

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

Das Forschungsdatenzentrum der Universität Hamburg

Hagen Peukertⁱ Sylvia Melzerⁱⁱ Juliane Jacobⁱⁱⁱ
Iris Vogel^{iv} Kai Wörner^v Steffen Grünler^{vi}
Lisa Brase^{vii} Alexander Heinrich^{viii} Stefan Thiemann^{ix}

2023

Zitiervorschlag

Peukert, Hagen, Sylvia Melzer, Juliane Jacob, Iris Vogel, Kai Wörner, Steffen Grünler, Lisa Brase, Alexander Heinrich und Stefan Thiemann. 2023. Das Forschungsdatenzentrum der Universität Hamburg. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2023: S. 2-13. DOI: [10.17192/bfdm.2023.3.8562](https://doi.org/10.17192/bfdm.2023.3.8562).

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

ⁱUniversität Hamburg. ORCID: [0000-0002-3228-316X](https://orcid.org/0000-0002-3228-316X)

ⁱⁱUniversität zu Lübeck, Universität Hamburg. ORCID: [0000-0002-0144-5429](https://orcid.org/0000-0002-0144-5429)

ⁱⁱⁱUniversität Hamburg. ORCID: [0000-0002-0443-3570](https://orcid.org/0000-0002-0443-3570)

^{iv}Universität Hamburg. ORCID: [0000-0003-2066-3840](https://orcid.org/0000-0003-2066-3840)

^vUniversität Hamburg. ORCID: [0000-0001-8939-4437](https://orcid.org/0000-0001-8939-4437)

^{vi}Universität Hamburg. ORCID: [0000-0001-9852-0626](https://orcid.org/0000-0001-9852-0626)

^{vii}Universität Hamburg. ORCID: [0000-0001-6654-8800](https://orcid.org/0000-0001-6654-8800)

^{viii}Universität Hamburg. ORCID: [0000-0002-5835-0033](https://orcid.org/0000-0002-5835-0033)

^{ix}Universität Hamburg. ORCID: [0000-0001-8300-2519](https://orcid.org/0000-0001-8300-2519)

Abstract

Die Umsetzung und Etablierung eines generischen Forschungsdatenmanagements an der Universität Hamburg folgt dem Ansatz des Kaizen-basierten Continuous Improvement Prozesses (CIP)¹, bei dem induktiv und kleinschrittig aus disziplinspezifischen Best-Practices gelernt wird. Der Forschungsdatenlebenszyklus, in dem sich die Prozesse jeder Disziplin wiederfinden lassen, dient als gemeinsamer Ausgangspunkt. Ebenso werden die strategischen Vorgaben der Universität zum Open Science berücksichtigt. Dieser Erfahrungsbericht stellt ein Ergebnis des CIPs dar, dessen Hauptaufgabe darin besteht, die Prozesse so zu gestalten und in verschiedenen Softwarelösungen abzubilden, dass sich eine stetig wachsende Zahl an disziplinspezifischen Praktiken umsetzen lassen.

1 Einleitung

Die Entwicklung von spezifischen zu generischen Prozessen in öffentlichen Institutionen ist unter der Prämisse der Effizienz durch neue Technologien und den Rechtfertigungsdruck gegenüber der Öffentlichkeit nachweisbar². Die Gestaltung von Abläufen im Forschungsdatenmanagement (FDM) einer Universität unterliegt ebenfalls dieser Tendenz, hängt jedoch auch direkt von ihren Organisationsstrukturen ab. Eine vollständige Umsetzung von disziplinspezifischen zu generischen Prozessen lässt sich deshalb nur bedingt beobachten. Dies spiegelt auch die anhaltende Diskussion in einschlägigen Konferenzen, Workshops und Publikationen wider. Darin zeigt sich, dass unterschiedliche Probleme und Lösungen im Datenmanagement in den Disziplinen vorherrschen, wobei die Natur- und Geisteswissenschaften die Weite des Problemraums aufspannen und die Sozial- und Lebenswissenschaften sich dazwischen positionieren.

Datendienste, die auf die speziellen Bedürfnisse einer Disziplin oder sogar eines Fachs zugeschnitten sind, lassen sich in größerer Zahl den Naturwissenschaften (Physik, Chemie, Geowissenschaften) zuschreiben, während sie für geisteswissenschaftliche Fächer Ausnahmen bilden. Bei Ersteren ist auffällig, dass Skalierungsprobleme den zentralen Arbeitsschwerpunkt bilden. Bei Letzteren können die Heterogenität von relativ geringen Datenmengen und die Abweichung (oder Abwesenheit) von Standards als typische Problemfelder identifiziert werden. In beiden Fällen ist allerdings immer noch ein Effizienzproblem zu lösen: Einerseits müssen riesige Datenmengen und andererseits sehr viele unterschiedliche Anwendungen mit vielfältiger Funktionalität in einem

¹Fryer, Karen J., Jiju Antony, und Alex Douglas. „Critical Success Factors of Continuous Improvement in the Public Sector: A Literature Review and Some Key Findings“. *The TQM Magazine* 19, Nr. 5 (28. August 2007): 497–517. <https://doi.org/10.1108/09544780710817900>

²Moore, Mark H. *Creating Public Value: Strategic Management in Government*. Cambridge, MA: Harvard University Press, 1997. Siffels, Lotje, David van den Berg, Schäfer, Mirko Tobias, und Iris Muis. „Public Values and Technological Change: Mapping how Municipalities Grapple with Data Ethics“. In *New Perspectives in Critical Data Studies*, 2022.

akzeptablen Zeitraum mit begrenzten Ressourcen verarbeitet werden. Hinzu kommt, dass sich die Anforderungen an Heterogenität und Datenvolumen von Geistes- und Naturwissenschaften in den letzten Jahren zunehmend annähern. Es lassen sich nunmehr sehr viele kleinere Projekte mit überschaubaren Datenbeständen mit spezifischer Anwendungsfunktionalität in den naturwissenschaftlichen Fächern ausmachen, die auf ein institutionalisiertes FDM angewiesen sind. Gleichzeitig melden einzelne geisteswissenschaftliche Projekte Speicherbedarfe im Petabyte-Bereich an.

Im vorliegenden Beitrag möchten wir dafür argumentieren, dass Forschungsdatenmanagement unabhängig von der Disziplin als eine gleichwertige Optimierungsaufgabe begriffen wird, welche durch die Bereitstellung geeigneter Tools mit wenigen Anpassungen entlang des Datenlebenszyklus gelöst werden kann. Dieser Ansatz soll jedoch weder die existierenden Datendienste von disziplinspezifischen Einrichtungen ersetzen, noch in Frage stellen, dass ein interdisziplinäres Team notwendige Bedingung für die Umsetzung einer erfolgreichen FDM-Strategie darstellt. In den spezialisierten Datenzentren existiert sehr viel implizites Wissen, das in einem generischen Forschungsdatenzentrum nicht aufgebaut und häufig auch nicht abbildbar ist.

2 Wissenstransfer und Lernen

Daten sind allgegenwärtig und das Zentrum für nachhaltiges Forschungsdatenmanagement (ZFDM) unterstützt Angehörige der Universität Hamburg (UHH), einen kritisch-reflexiven Umgang mit Forschungsdaten im Sinne der Data Literacy zu etablieren. Allerdings sind die Kenntnisstände der UHH-Angehörigen uneinheitlich. Um ein entsprechendes FDM-Angebot aufzubauen, wurde der Bedarf ermittelt. Dieser wird von UHH-Studierenden und Doktorand:innen als hoch eingeschätzt, wobei die FDM-Kenntnisse i. d. R. noch rudimentär sind. Lehrende sehen den Bedarf ebenso. In den ohnehin schon vollen Curricula sind Lernziele und -inhalte genau vorgegeben und zusätzliche, fakultative Angebote werden selten besucht. Das ZFDM integriert daher Angebote auch in bestehende Lehrveranstaltungen und ist in stetigem Austausch mit Lehrenden. FDM-Basiswissen und Strategien im Studium zu implementieren, führt schon bei Abschlussarbeiten zur Qualitätsverbesserung und Arbeitserleichterung (siehe E²D²-Projekt im DDLitLab-Förderprogramm³). Denn FDM verbessert die Data Literacy der Studierenden z. B. in Bezug auf die FAIR-Prinzipien und rechtliche Aspekte, die ggf. für die Erhebung von Forschungsdaten relevant sind. Doktorand:innen müssen selbständiger wissenschaftlich arbeiten und umfangreichere FDM-Kenntnisse sind hilfreich. Sie durchlaufen erstmals den Forschungsdatenlebenszyklus: Erheben oder nutzen fachspezifische Daten, planen, dokumentieren und veröffentlichen ihr Forschungsvorhaben und beenden ihre Doktorarbeit innerhalb einer befristeten Laufzeit. Für Postdocs und Senior Scientists sind weitere FDM-Kompetenzen wie wissenschaftliches

³Early Education for Data Management Decisions, zugegriffen 18.11.2022, <https://www.isa.uni-hamburg.de/ddlitlab/data-literacy-lehrlabor/erste-foerderrunde/fs-e2d2.html>.

Publizieren relevant (z. B. im Forschungsdatenrepositorium, Kap. 3.3.1) und Maßstab für den Forschungs-/Projekterfolg. Für das Einwerben weiterer Fördergelder wird zusätzlich ein Datenmanagementplan (Kap. 3.1) empfohlen.

Die FDM-Inhalte sind gemäß der Lernzielmatrix⁴ auf die Zielgruppen (Qualifikationsstufen Bachelor, Master, PhD und Data Steward) ausgerichtet und spezifische Schulungen (z. B. Datenbereinigung) werden individuell erstellt. Ebenso werden Beratungen oder Tutorials für alle ZFDM-Services (Kap. 3, Abb. 1) angeboten. Die Möglichkeit, via E-Mail oder einer offenen Sprechstunde Kontakt zum ZFDM aufzunehmen, runden das Angebot ab. Somit können UHH-Angehörige ihre FDM-Kompetenzen vielfältig ausbauen und FDM-Services dauerhaft nutzen. So wird das Bewusstsein von FDM und Data Literacy bezüglich der Relevanz für die tägliche Arbeit, aber auch für die langfristige Verfügbarkeit und Nachnutzbarkeit von Forschungsdaten erhöht.

Hilfreich ist nicht nur der Wissenstransfer innerhalb der UHH, sondern auch die Vernetzung mit anderen Trainer:innen u. a. in der UAG Schulungen/Fortbildungen der DINI/nestor-AG Forschungsdaten⁵ und der GO UNITE Initiative⁶. Trainingsmaterialien werden so didaktisch und thematisch fortlaufend weiterentwickelt.

3 Datenlebenszyklus als Orientierung beim Aufbau von Datenservices

Zur Orientierung und Visualisierung einer FDM-Strategie hat sich der Forschungsdatenlebenszyklus bewährt. Abbildung 1 zeigt den auf das Zentrum für nachhaltiges Forschungsdatenmanagement) der Universität Hamburg angepassten Zyklus mit den jeweiligen Softwareanwendungen und Verbindungen zwischen diesen Komponenten. Veranschaulicht wird, welche Werkzeuge miteinander interagieren, welche Daten abgeholt werden sollen und wo sie gespeichert werden. Diese Anwendungen werden in den nachfolgenden Kapiteln eingängiger beschrieben. In allen Phasen - Planung der Datenbeschaffung und Datenerfassung (Kap. 3.1), Verwaltung aktiver Daten (Kap. 3.2), Datenrepositorien und -archive (Kap. 3.3) und Datenkataloge und -verzeichnisse (Kap. 3.4) - werden unterschiedliche, teils implizite, Wissensressourcen benötigt (Kap. 2). Dieses Wissen verändert sich mit Fortschreiten eines Projekts und wird durch den Wissenstransfer verfügbar gemacht. Der innere Kreis in Abbildung 1 symbolisiert die

⁴Petersen, Britta, Engelhardt, Claudia, Hörner, Tanja, Jacob, Juliane, Kvetnaya, Tatiana, Mühlichen, Andreas, Schranzhofer, Hermann, u. a. „Lernzielmatrix zum Themenbereich Forschungsdatenmanagement (FDM) für die Zielgruppen Studierende, PhDs und Data Stewards“, 5. September 2022. <https://doi.org/10.5281/ZENODO.7034477>.

⁵Forschungsdaten.org, zugegriffen 18.11.2022, https://www.forschungsdaten.org/index.php/UAG_Schulungen/Fortbildungen.

⁶Active GO FAIR Implementation Network, zugegriffen 18.11.2022, <https://www.go-fair.org/implementation-networks/overview/dscc/>.

Maßnahmen, um die Data Literacy zu erhöhen. Das Leitbild und die Digitalisierungsstrategie der Universität Hamburg geben vor, dass sich alle Datendienste an der Idee von offenen Daten (Open Data) und Datennachhaltigkeit ausrichten (Kap. 4). Diese sind nicht explizit in Abbildung 1 ersichtlich, lassen sich aber aus den verwendeten Softwarelösungen ableiten. Beschränkungen für die Nutzung von Daten, die aus dem Urheberrecht oder aus Abhängigkeiten von Lizenzierungen Dritter folgen, werden von den hier beschriebenen organisatorischen Prozessen und Data Workflows nicht dargestellt.

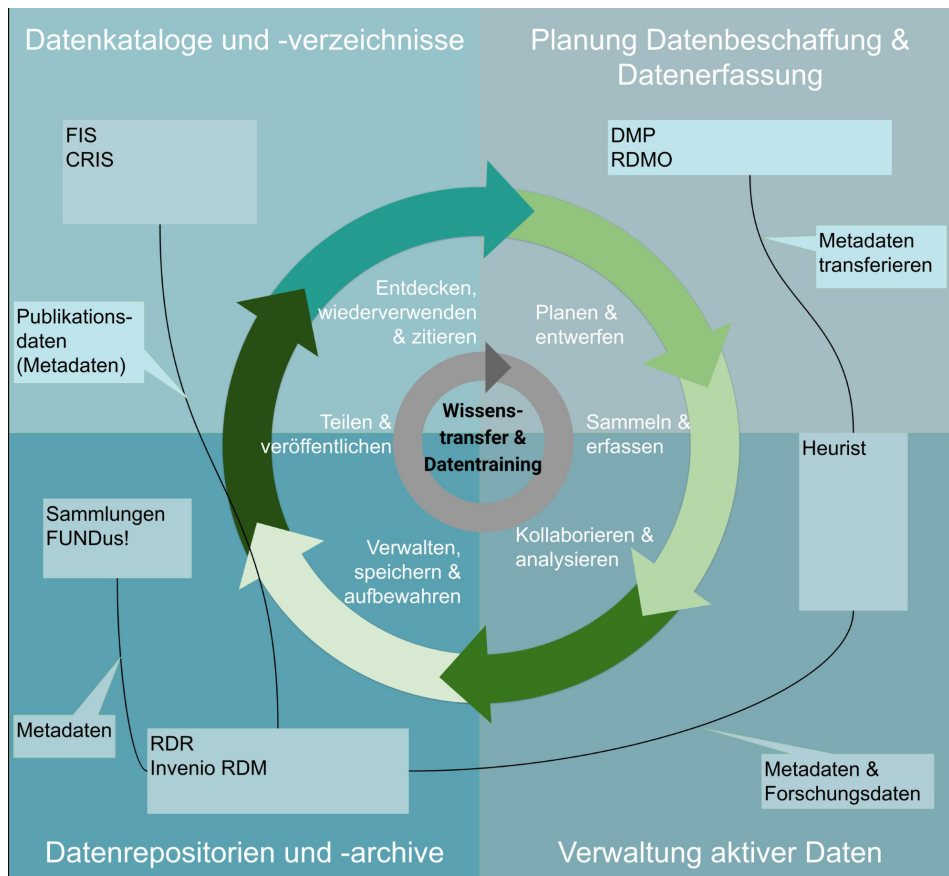


Abbildung 1: Der Forschungsdatenlebenszyklus am ZFDM der Universität Hamburg⁷

3.1 Datenbeschaffung und -aufbewahrung planen und designen

Ein typisches Forschungsprojekt beginnt mit der Planung des Projekts (Abbildung 1 oben rechts). Das ZFDM der UHH unterstützt die Phase der Planung mit einem Werkzeug⁸ zum Erstellen von Datenmanagementplänen (DMP). Darin werden Fragenkataloge hinterlegt, die mit den Richtlinien der Forschungsförderer abgestimmt sind.

⁷Der Forschungsdatenlebenszyklus am ZFDM der Universität Hamburg, zugegriffen 18.11.2022, https://leopard.tu-braunschweig.de/receive/dbbs_mods_00071505.

⁸RDMO, zugegriffen 16.12.2022, <https://rdmorganiser.github.io/>.

Forschende können diese Kataloge ausfüllen und erhalten einen fertigen Datenmanagementplan. Bei Anfrage nach fachspezifischen Fragekatalogen können Vorlagen nachgeladen oder auch selbst erstellt werden. Damit verbunden sind Definitionen von Vokabularen und Sichten, aber auch dazugehörige Bedingungen und Optionen, welche dazu dienen, die Beantwortung der Fragen als Auswahllisten, Freitexte, oder beispielsweise Untermenüs zu präsentieren oder mit noch ausstehenden Aufgaben zu annotieren. Für selbsterstellte Fragebögen übernehmen die Forschenden aus den Fachdisziplinen die Verantwortung zur längerfristigen Pflege und Wartung der Kataloge. Das Werkzeug ist an die Benutzerverwaltung der UHH angebunden, sodass die Nutzung auf die Mitarbeiter:innen der UHH beschränkt bleibt.

Institutionell ist die Erstellung von DMP an der UHH nicht einheitlich geregelt. Jedoch haben sich einige Fakultäten darauf verständigt, Forschungsanträge erst dann abzuzeichnen, wenn die Antragstellenden ein Beratungsgespräch am ZFDM wahrgenommen haben. Auf diese Weise kann die Anfertigung eines DMP empfohlen sowie dessen Vorteile und Relevanz hinsichtlich des Angebots an weiteren Datendiensten erklärt werden. Da die einschlägigen Forschungsförderer einen DMP erwarten, entsteht kein zusätzlicher Aufwand.

Um einen möglichst effizienten Workflow im Forschungsdatenmanagementzyklus zu erreichen, besteht die Möglichkeit, relevante Informationen des DMP direkt in einem nachgelagerten Werkzeug, z. B. Heurist⁹, als Metadaten zu verwenden. Derzeit ist diese Funktion nur über Export und Import möglich. Zukünftig sollen für die zweite Phase im Forschungsdatenlebenszyklus die Metadaten des DMP direkt über eine Schnittstelle integrierbar sein.

3.2 Daten sammeln, Daten analysieren und kollaboratives Arbeiten

Im Falle einer Projektbewilligung werden bei der Durchführung eines Projekts häufig Daten erhoben, ausgewertet, Modelle entworfen, geändert und mit unterschiedlichen Berechtigungen sowie Sichten auf die Daten bearbeitet. Für Projekte mit kleinem bis mittlerem Volumen (im Gigabyte-Bereich) wird für diese Phase eine generische Softwarelösung zur Verfügung gestellt (Heurist). Es handelt sich um ein Open-Source-Datenbankmanagementsystem mit einem Web-Frontend und erlaubt Forschenden verschiedenster Fachrichtungen ohne spezifische IT-Vorkenntnisse, Datenmodelle zu entwickeln, Daten zu sammeln, Suchen zu speichern und Daten auf einer Webseite zu veröffentlichen.

Üblicherweise werden Forschungsdaten in den Formaten DOCX, PDF, JSON und XML repräsentiert, die beispielsweise im Forschungsdatenrepositorium (FDR) archiviert werden. Eine Wiederverwendung dieser archivierten Daten für eine weiterführende Analyse ist oft aufwändig. Die Suche nach bestimmten Informationen in den verschiedenen Dateien erfordert eine benutzerfreundliche Schnittstelle mit einer Suchfunktio-

⁹Heurist, zugegriffen 25.10.2022, <https://heuristnetwork.org/>.

on. Um die Suchergebnisse domänenspezifisch und benutzerfreundlich interpretieren zu können, müssen die Forschungsdaten, die in einem maschinenlesbaren Format repräsentiert werden, zusätzlich in ein für den Menschen lesbares Format überführt werden. Zur Präsentation der Forschungsdaten sind die vielfältigen Anforderungen der Wissenschaftler:innen zu berücksichtigen. Es gibt eine Vielzahl von Werkzeugen und Ansätzen für den Aufbau von Informationssystemen, doch die größte Herausforderung besteht darin, die Nutzer:innen zu befähigen, die verfügbaren Werkzeuge und Ansätze ohne IT-Expertise zu nutzen. Somit können die Informationssysteme auf der Grundlage der projektspezifischen Anforderungen unter Verwendung der archivierten Daten aus Repositorien, z. B. mittels des Prozesses Databasing-on-Demand (DBoD), in Heurist oder anderen Datenbankmanagementsystemen automatisch selbst aufgebaut werden.¹⁰

Der DBoD-Prozess ist ein effizienter Ansatz, der aus archivierten Daten in Forschungsdatenrepositorien Datenbank-Instanzen automatisch auch bereits in wenigen Minuten erstellen kann.^{11 12}

In der Regel dürfen Forschungsdaten nach Beendigung eines Projekts nicht mehr verändert werden. Die Daten aus Heurist können dann in das FDR als langfristige Speicheroption überführt werden. Das FDR skaliert auch bei größeren Datenmengen und ist so für alle abgeschlossenen Projekte gleichermaßen als Archivierungslösung nutzbar.

3.3 Daten veröffentlichen, speichern und archivieren

3.3.1 Forschungsdatenrepositorium (FDR)¹³

Um den Anforderungen an FAIRes FDM gerecht zu werden, ist es von zentraler Bedeutung, einen Ort zu haben, an dem die Forschungsdaten dauerhaft und eindeutig referenzierbar abgelegt werden können.

Für einige wissenschaftliche Disziplinen und Datentypen existieren domänenspezifische Repositorien, die diese Rolle ausfüllen können. Für alle Daten, die in keines dieser Repositorien passen, existieren Dienste wie Zenodo.org, die die Speicherung von

¹⁰Acknowledgement: This section was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – DFG - GEPRIS - EXC 2176: Understanding Written Artefacts, zugegriffen 17. April 2023. <https://gepris.dfg.de/gepris/projekt/390893796>.

¹¹Melzer, S., S. Schiff, F. Weise, K. Harter-Uibopuu, und R. Möller. „Databasing on Demand for Research Data Repositories Explained with a Large Epidoc Dataset“. In CENTERIS 2022 - Conference on ENTERprise Information Systems.

¹²Schiff, Simon, Sylvia Melzer, Eva Wilden, und Ralf Möller. „TEI-Based Interactive Critical Editions“. In Document Analysis Systems, herausgegeben von Seiichi Uchida, Elisa Barney, und Véronique Eglin, 230–44. Cham: Springer International Publishing, 2022.

¹³Research Data Repository (RDR), zugegriffen 24.03.2023, <https://www.fdr.uni-hamburg.de>.

Forschungsdaten nach dem derzeitigen State of the Art anbieten und gleichzeitig ein glaubwürdiges Versprechen für ihre langfristige Verfügbarkeit machen können.

In vielen Fällen wollen oder dürfen Forschende Daten aber nicht auf Systemen speichern, die außerhalb der eigenen Forschungseinrichtung betrieben werden oder möchten sich nicht mit den Bedingungen beschäftigen, zu denen sie ihre Daten in solche Repositorien überführen können.

Für diese Fälle wird an der UHH das Forschungsdatenrepositorium (FDR) zur Verfügung gestellt. Es ermöglicht die sichere und langfristige Ablage von Forschungsdaten in einem weiten Sinne (auch Arbeitspapiere, Projektdokumentation, Quellcode zu Forschungssoftware und weiteres sind explizit erwünscht) mit einer niedrigen Einstiegschürde.

Alle Mitarbeitenden der UHH können das Repositorium mit ihren vorhandenen Anmeldedaten kostenlos nutzen und Daten selbstständig einstellen. Die dazu nötigen Angaben sind weitgehend selbsterklärend und die geringe Zahl an Rückfragen sowie die gute Akzeptanz zeigen, dass niedrige Hürden ein wichtiger Baustein sein können, wenn OS-Praktiken aktiv gefördert werden sollen.

Damit die eingestellten Daten nicht nur als Referenz und Beleg für die Gültigkeit damit erzielter Forschungsergebnisse dienen können, sondern auch auffindbar sind, müssen die Forscher:innen die Daten mit standardisierten Metadaten beschreiben.¹⁴ Das Repositorium sichert die Vollständigkeit der Metadaten, versieht die Datensätze mit einem eindeutigen Identifier¹⁵ und sorgt dafür, dass die Daten in gängigen Indizes erscheinen und für Forschungsdaten-Suchmaschinen zugänglich sind.

Die tatsächliche Nachnutzbarkeit der Daten hängt von den Bedingungen ab, zu denen Interessierte die Daten herunterladen und nutzen dürfen, aber auch von der Dokumentation der Entstehung der Daten und von den verwendeten Dateiformaten. Die Dokumentation des FDR gibt dazu Hilfestellung, und die Mitarbeitenden des ZFDM können zu diesen Aspekten beratend zur Seite stehen.

Die dem Repositorium zugrundeliegende Software¹⁶ ist auf den Umgang mit großen Datenmengen ausgerichtet und das Repositorium ist an den Langzeit-Datenspeicher des Rechenzentrums der UHH angeschlossen. Es ist damit durchaus ausreichend dimensioniert, auch Forschungsdaten weiterer Institutionen aufzunehmen.

¹⁴Die Metadaten im FDR folgen dem Datacite Metadata Schema (<https://schema.datacite.org/>) und lassen sich in eine Reihe anderer Metadatenformate übersetzen.

¹⁵Das FDR verwendet mit DOIs (Digital Object Identifier, <https://www.doi.org/>) den de-facto Standard für PIDs zu Forschungsdaten.

¹⁶Inveniosoftware, zugegriffen 17.04.2023, <https://inveniosoftware.org/products/rdm/>.

3.3.2 Wissenschaftliche Sammlungen

Universitäre Sammlungen beherbergen eine Vielfalt an epistemischen Objekten, die in der Regel dezentral gelagert, administriert und nur zu einem geringen Anteil ausgestellt werden. Die UHH allein beherbergt schätzungsweise 13 Millionen Sammlungsobjekte in 37 wissenschaftlichen Sammlungen. Nur ein kleiner Ausschnitt davon ist in neun Museen der Universität der Öffentlichkeit direkt zugänglich und vereinzelt finden sich digitale Repräsentationen in überregionalen Sammlungskatalogen.

Durch das Sammlungsportal für wissenschaftliche Sammlungen FUNDus!^{17 18} wird der Versuch unternommen, die Schätze und die Expertise, die durch die systematische Erfassung und Digitalisierung der Objekte an der UHH aggregiert wurde, öffentlich zugänglich zu machen. Bei der Konzeption des Sammlungsportals stand die Anforderung im Zentrum, flexibel auf unterschiedliche Sammlungskontexte und damit verbundene Metadatenformate einzugehen. Als Datenquellen überwiegen einfache Tabellen. Auch eingebunden sind Daten anderer gängiger Austauschformate (z. B. Lido-XML¹⁹) bzw. Schnittstellen zu entsprechenden Datenbanken.

Seit der Veröffentlichung des Sammlungsportals 2018 ist die Anzahl der digitalen Sammlungsobjekte von 14072 auf derzeit 31674 digitale Objekte aus 25 Sammlungen erweitert worden (Stand Dezember 2022). Den Anforderungen der Sammlungsleitung entsprechend wird die Darstellung der Sammlungen sukzessive erweitert, beispielsweise zur Präsentation von Teilsammlungen, Objekten zu Unterrichtszwecken oder der Abhängigkeiten zwischen Sammlungsobjekten.

Als Teil der Forschungsdateninfrastruktur²⁰ bietet das Sammlungsportal die Möglichkeit, geeignete Inhalte aus dem FDR direkt in FUNDus! zu präsentieren und dadurch auf Objektebene sichtbar und präziser durchsuchbar zu machen. Eine Erweiterung der Schnittstellen zum Sammlungsportal wird zukünftig auch in Verbindung mit der Zusammenarbeit mit dem Leibniz-Institut zur Analyse des Biodiversitätswandels (LIB)²¹ angestrebt.

¹⁷FUNDus! Sammlungsportal der Universität Hamburg, zugegriffen 01.12.2022, <https://www.fundus.uni-hamburg.de>.

¹⁸Der Name "FUNDus!" weist auf die Schwierigkeiten hin, Finanzierung für diesen Bereich zu erhalten, insbesondere da die meisten Sammlungen von außen nicht sichtbar sind.

¹⁹LIDO - Lightweight Information Describing Objects Version 1.1, zugegriffen 01.12.2022, <http://lido-schema.org/schema/v1.1/lido-v1.1.html>.

²⁰Die Empfehlungen des Wissenschaftsrats machen sehr deutlich, dass in Bezug auf Sammlungen im universitären Kontext der wissenschaftliche Nutzen im Vordergrund stehen muss. Empfehlungen zu wissenschaftlichen Sammlungen als Forschungsinfrastrukturen (Drs. 10464-11), Januar 2011 https://wissenschaftliche-sammlungen.de/files/3213/7096/3073/WR_EmpfehlungenWissSlg2011.pdf, zugegriffen 30.10.2022.

²¹Leibniz-Institut zur Analyse des Biodiversitätswandels (LIB), zugegriffen 16.12.2022, <https://leibniz-lib.de/>.

3.4 Forschungsinformationssystem: Verknüpfung, Verwaltung und Wiederverwendung der Daten

3.4.1 Das Forschungsinformationssystem der Universität Hamburg – Struktur und Datenflüsse

Das Forschungsinformationssystem (FIS) soll den aktuellen Stand der Forschung an der UHH widerspiegeln. Metadaten von Forschungsaktivitäten wie Publikationen, Forschungsprojekten, Aktivitäten und andere Inhalte werden hier zusammengeführt und miteinander verknüpft. Als Basis dienen Personen- und Organisationsdaten aus zuliefernden Systemen der Personalverwaltung (KoPers) und dem Rechnungswesen (SAP). Die Verknüpfung dieser Daten stellt das Gerüst für die Anreicherung mit Metadaten dar, auch für nachgelagerte Systeme und Kataloge außerhalb des FIS, die forschungsbezogene Daten verarbeiten. Als Informationsquelle für Nachnutzung, Zitate und Publikation steht dem FIS eine wichtige Aufgabe im Forschungsdatenlebenszyklus zu (Abb. 1 Datenverzeichnisse und –kataloge). Es übernimmt darüber hinaus weitere Aufgaben, die integraler Bestandteil der Arbeit des ZFDM sind.

Die gesammelten Informationen dienen als Grundlage für die Steuerung der Universität, indem es die Verwaltung von Forschungsprojekten begleitet, die Erstellung von Berichten für unterschiedliche Zwecke (Behörden, Mittelgeber, interne Berichte) unterstützt. Nicht zuletzt schafft die öffentliche Bereitstellung der dort gesammelten Informationen Transparenz. Die Wissenschaftler:innen selbst sind hier für die Eingabe aller relevanten Daten verantwortlich und können diese jederzeit ergänzen, bearbeiten oder löschen.

Das FIS der UHH übernimmt Bestandsdaten aus anderen, älteren internen IT-Systemen, z. B. Metadatenbestände aus der Bibliothek (Pica) oder aus lokalen Datenbanken im RIS oder Bibtex-Format (z. B. aus Citavi, Endnote). Das Einpflegen von aktuellen oder fehlenden Daten kann durch eine direkte Eingabe vorgenommen werden. Dieses manuelle Eintragen ist zeitaufwändig, hat jedoch den Vorteil, dass die Daten eine höhere Qualität aufweisen. Ebenso können auch Daten aus unterschiedlichen Fachdatenbanken automatisiert importiert und synchronisiert werden. Zur eindeutigen Identifikation von Personen werden Autoren-IDs verwendet, wie z. B. ORCID und Scopus-ID.

Alle in das FIS zu importierenden Daten werden zuvor einer Prüfung unterzogen. Auch muss beispielsweise die Zugehörigkeit zu mindestens einer internen Person oder die Zuordnung zu einer internen Organisationseinheit vorhanden sein. Die Sicherung der Datenqualität wird zusätzlich durch die Bedürfnisse der beteiligten Wissenschaftler:innen und durch Schulungen mit den verantwortlichen Personen im Verwaltungsbereich gewährleistet.

Synchronisierte Daten werden aus anderen IT-Systemen in das FIS übernommen. Eine Synchronisation von Daten kann auch mit internen IT-Systemen durchgeführt werden, etwa dem Personalsystem und dem Finanzsystem. Durch diese Synchronisation

werden Nutzende, dazugehörige Forschungsprojekte und Bewilligungen im FIS angelegt.

Die im Projektverlauf anfallenden Metadaten zu Publikationen, weiteren Aktivitäten wie Konferenzteilnahmen oder auch benutzten Gerätschaften können durch die Wissenschaftler:innen und berechtigten Personen der Hochschulverwaltung miteinander verknüpft werden. Dieses implizite Wissen (z. B. wer, was, wann, wo publiziert) und die Verknüpfung mit weiteren Inhalten nachvollziehbar zu sichern, bilden die Grundlage für die nachgelagerte Auswertung und Berichterstattung zu den Forschungsaktivitäten an der UHH.

3.4.2 Digitale Verwaltung des Verlaufs von Forschungsprojekten im FIS

Mit dem FIS stellt die UHH auch ein spezielles Modul zur digitalen Verwaltung von Forschungsprojekten zur Verfügung. Der im Modul abgebildete Prozess umfasst den Verlauf von der Antragsphase über die Bewilligung bis zum Projektabschluss. Dieser Prozess unterliegt hierbei einer holistischen Sichtweise. So wird im FIS ein transparenter, einheitlicher, rechtssicherer und vereinfachter Workflow umgesetzt. Beispielsweise sind bestimmte Berechtigungen zum Lesen und Bearbeiten von Einträgen über mehrere Organisationseinheiten (Institute, Fachbereiche, Fakultäten etc.) und Schritte im Workflow definiert. Dadurch werden die vielfältigen spezifischen Managementpraktiken in den jeweiligen Organisationseinheiten zugelassen, aber auch durch den Prozess als solches vereinfacht. Gleichzeitig erhöht sich bei diesem Vorgehen die Transparenz und Geschwindigkeit des Antragsprozesses und laufende Projekte können effizienter begleitet werden.

In der Regel wird ein Antrag zur Förderung eines Forschungsprojektes durch den Antragsteller selbst im FIS angezeigt. Ebenso kann durch explizit gesetzte Berechtigungen eine Projektanzeige von einer Person in der Fakultäts- und Präsidialverwaltung gestellt werden. Sobald ein Eintrag im FIS erzeugt wurde, wird eine Mitteilung an alle relevanten Stellen der verschiedenen Organisationseinheiten versandt. Angestrebt wird eine möglichst frühzeitige Erfassung des Antragsvorhabens. Gelingen kann dies unter anderem dadurch, dass zunächst nur wenige, notwendige Metadaten wie Antragstellende, Titel und Fördersumme eingegeben werden müssen. Mit der Kenntnis über einen Projektantrag an den relevanten Stellen in der Hochschulverwaltung beginnt die UHH-interne Begleitung eines Forschungsprojektes. Das frühe und transparente Einbinden relevanter Stellen führt dazu, dass ggf. notwendige Beratungsleistungen durch die Hochschulverwaltung in der Antragsphase angeboten werden.

Fragen zur Compliance (z. B. Erfassung von Ethikgutachten) lassen sich ebenso im FIS abbilden und tragen zu einem verbesserten, rechtssicheren Ablauf des Genehmigungsprozesses in den beteiligten Organisationseinheiten bei.

Erst mit einem offiziellen Einreichen des Förderantrags werden in der Regel weitere Metadaten wie z. B. eine Projektbeschreibung, Stichwörter und die Verknüpfung zum

Förderprogramm erfasst. Mit einer Bewilligung durch den Forschungsförderer erlangt ein Antrag den Status Forschungsprojekt. Für die Verwaltung des Projektes stehen in dieser Phase Werkzeuge zur Verfügung, anhand deren der Projektverlauf nachverfolgt und automatisiert den Wissenschaftler:innen mitgeteilt werden kann. So kann z. B. an das Einhalten von Terminen und Fristen erinnert werden.

3.4.3 Nutzung der Daten für das Berichtswesen

Die Zugänglichkeit und die Nutzung der im FIS hinterlegten Daten zu Forschungsaktivitäten spielen auch im Reporting eine wichtige Rolle. Durch die stetige Zunahme der Bedeutung des Berichtswesens über die letzten Jahre an den Universitäten ist eine Sicherstellung der Datenqualität und Vollständigkeit über die wissenschaftlichen Aktivitäten an der UHH in den Fokus gerückt.

Durch das Ablegen der zu den Forschungsaktivitäten zugehörigen Metadaten an zentraler Stelle können diese in Berichten gegenüber den Mittelgebern, Behörden, aber auch für interne Zwecke (z. B. Evaluierung von Tenure Track) genutzt werden. Vor allem bei sich wiederholenden Berichten (z. B. bei Forschungsclustern gegenüber dem Mittelgeber) bietet das FIS der UHH die Möglichkeit, Vorlagen zu erstellen und diese mit den jeweiligen tagesaktuellen Daten automatisiert abzurufen.

Die Qualitätssicherung wird sowohl durch die Wissenschaftler:innen, als auch durch die beteiligte Fakultäts- und Präsidialverwaltung durchgeführt. Durch einen etablierten Validierungsprozess wird zusätzlich sichergestellt, dass Änderungen noch einmal geprüft werden müssen, bevor diese angenommen werden.

Die zentrale Ablage und der Prozess der Validierung hinter den im FIS abgelegten Metadaten zu den Forschungsaktivitäten der UHH stellt somit sicher, dass alle Berichte auf einer gemeinsamen, konsistenten Datenbasis und sich wiederholende Berichte auf denselben Inhaltstypen beruhen, sowie alle Änderungen historisch einsehbar und verfolgbar sind.

3.4.4 Transparenz der Metadaten: Forschungsaktivitäten nach außen sichtbar machen

Neben der internen Verwaltungsnutzung für das Erstellen von Berichten ist die Darstellung der Forschungsaktivitäten für Außenstehende ein weiterer wichtiger Punkt, um Zugänglichkeit und somit Transparenz zu schaffen.

Da eine Universität zum Großteil durch öffentliche Gelder finanziert wird, sollte auch dargelegt werden, in welche Forschung diese Gelder fließen. Eine für alle frei zugängliche Außendarstellung der im FIS hinterlegten Forschungsaktivitäten ist ein geeignetes Mittel, um Interessierten ein umfassendes Bild aufzuzeigen. Gleichzeitig erhöht es die Sichtbarkeit der Wissenschaftler:innen und der gesamten Universität.

Die Veröffentlichung der Forschungsmetadaten fördert auch den wissenschaftlichen Austausch über die Einrichtung hinaus. Die Wissenschaftler:innen der UHH haben zusätzlich die Möglichkeit mittels eines in das Web Content Management System integrierten Werkzeugs ihre im FIS hinterlegten Publikationen, Projekte und Aktivitäten auf ihre persönliche Webseite zu spielen. Ebenso können Instituts- und Fakultätswebseiten nach demselben Prinzip ergänzt werden.

Mittels des FIS-Portals²² präsentiert sich die Forschung der UHH im Gesamtbild. Alle im FIS hinterlegten Metadaten zu Forschungsaktivitäten werden hier aufgezeigt und können je nach Zweck gefiltert und sortiert werden (z. B. nach Organisationsstruktur, Personen, Themenfeldern, Ziele für nachhaltige Entwicklung²³). Auch findet sich ein Link zu den in Repositorien hinterlegten Daten.

4 Open Access und Open Data als strategische Grundausrichtung

Forschungsförderer, Wissenschaftsorganisationen, Politik und Gesellschaft²⁴ fordern und erwarten von Forschungseinrichtungen (FE) eine möglichst große Offenheit (Openness, Open Science). Publikationen sollen kostenfrei verfügbar sein (Open Access, OA) und Forschungsdaten zur Nachnutzung und Replizierung der Forschungsergebnisse verfügbar gemacht werden (Open Data, OD). So einfach diese Forderungen sind, so komplex sind im Detail die Umsetzungen. An einer Volluniversität müssen adäquate Lösungen angeboten und die unterschiedlichen Fachkulturen berücksichtigt werden.

Open Access erfordert Gelder für die Publikationsgebühren und -verträge (z. B. DEAL²⁵). Dabei werden Publikationen in wissenschaftlichen Zeitschriften oft bevorzugt finanziert, was einen Vorteil für die Naturwissenschaften darstellt. Die in anderen Fachwissenschaften häufig üblichen Buchpublikationen können oft gar nicht finanziert werden. Zudem erscheinen OA-Publikationen in der Regel nur in elektronischer Form (Gold OA), gedruckte Bücher werden aber weiterhin in diesen Fächern bevorzugt.

Die Zweitveröffentlichung (Green OA) erfordert seitens der Publizierenden die Speicherung der Publikation in einem Repository. Trotz vorhandener Angebote wird dies wenig genutzt, da meist Unsicherheiten bzgl. des Urheberrechts in Verbindung mit den jeweils relevanten Autorenverträgen bestehen. Auch hier hat sich in den Naturwissenschaften die Publikation als Preprint etabliert mit eigenen Repositorien (z. B.

²²FIS-Portal der UHH, zugegriffen 17.04.2023, <https://www.fis.uni-hamburg.de/>.

²³17 Ziele für nachhaltige Entwicklung, zugegriffen 20.04.2023, <https://unric.org/de/17ziele/>.

²⁴Forschungsfoerderinstitutionen, zugegriffen 16.12.2022, <https://openeconomics.zbw.eu/knowledgebase/haltung-von-forschungsfoerderinstitutionen-zu-open-science/>.

²⁵Projekt DEAL, zugegriffen 16.12.2022, <https://www.projekt-deal.de/>.

arXiv²⁶), was in anderen Fachwissenschaften bisher weitestgehend unüblich ist, ja sogar grundsätzlich abgelehnt wird.

Die Erstveröffentlichung außerhalb eines etablierten Verlags (Diamond OA) oder einer renommierten Zeitschrift wird wenig genutzt, zu wichtig ist der Publikationsort und das Renommee der Zeitschrift (impact factor) oder des Verlags für die wissenschaftliche Karriere. Hinzu kommt auch, dass ein Diamond Journal nicht unerhebliche Arbeitsaufwände erfordert und technische Infrastruktur benötigt, die dann meist mit einer spezifischen Einrichtung verbunden ist und von dieser finanziert wird. Für eine langfristige Perspektive muss sich in der Regel an der entsprechenden Einrichtung eine verantwortliche Person finden.

Open Data erfordert insbesondere technische Infrastruktur, ein leistungsfähiges Repository mit ausreichend Speicherplatz und einem verlässlichen, auf lange Zeit ausgerichteten Betriebsmodell. Das Forschungsdatenrepositorium (FDR) der UHH wurde deshalb als Forschungsinfrastruktur mit den zugehörigen Beratungs- und Schulungsangeboten (siehe Kap. 2) in die Digitalstrategie der Hochschule integriert. Damit wird ermöglicht, Forschungsdaten zu publizieren, soweit kein anderes, fachlich besser geeignetes Datenrepositorium existiert. Berücksichtigt werden unterschiedliche Anforderungen an den Schutz der Daten.

Daten, die bei der Forschung mit und an Menschen entstehen, unterliegen einem besonderen Schutz und dürfen in der Regel nur vollständig anonymisiert oder als aggregierte Daten publiziert werden. Bei berechtigtem Interesse müssen aber auch die Originaldaten langfristig aufbewahrt werden, was aus technischer Sicht eine besondere Herausforderung darstellt. Schützenswert sind aber auch Daten, die Beschränkungen des Urheberrechts (z. B. Digitalisate aus Museen und Archiven) und Vorgaben der Datengebenden (z. B. Wirtschaftsdaten von Firmen) unterliegen oder Daten, die bei einer Veröffentlichung eine Gefährdung bedeuten können. Letzteres ist gegeben, wenn die Publikation von Geokoordinaten und Ortsbezeichnungen Pflanzen, Tiere oder Kulturschätze und Personen oder soziale Gemeinschaften durch Informationen, z. B. in Interviews, gefährdet würden oder die Daten eine Gefährdung für die allgemeine Sicherheit darstellen. Hier entstehen komplexe rechtliche Fragestellungen und widersprüchliche Anforderungen, bei denen Forscher:innen kompetente Unterstützung benötigen.

Eine Verpflichtung zur Openness, wie sie oftmals gefordert und empfohlen wird, wirft deshalb neue Probleme auf. Forschende müssten erklären, warum sie nicht OA und OD publizieren und diese Verpflichtung müsste überprüft und bei Verstößen mit Sanktionen belegt werden. Die UHH fördert deshalb Openness durch entsprechende Beratung und technische Angebote. Ein wichtiger Baustein der Förderung ist dabei, dass die Angebote einfach und kostenfrei zu nutzen sind. Insbesondere wird aber den Forscher:innen vermittelt, dass Daten so offen wie möglich, aber auch so geschlossen wie nötig publiziert werden können. Das FDR der UHH unterstützt dies durch verschiedene Zugriffsrechte der Publikation (open, restricted, embargoed, closed). Openness

²⁶Arxiv.org, zugegriffen 16.12.2022, <https://arxiv.org/>.

kann so auch durch den Nachweis der Daten erreicht werden und entspricht durchaus den FAIR-Prinzipien.²⁷

5 Schluss

Die vom ZFDM angebotenen Dienste orientieren sich am Forschungsdatenlebenszyklus. Dies gewährleistet, dass Forschende für die gesamte Projektlaufzeit und darüber hinaus unterstützt werden. Das Austarieren, welche Projekte und Workflows besser mit einem generischen FDM zu bewältigen sind und welche Dienste in einer domänenspezifischen Lösung behandelt werden sollten, ist ein kontinuierlicher Prozess, der sich in vielen kleineren Schritten vollziehen wird. Voraussichtlich wird sich auch die hier vorgestellte Lösung weiterentwickeln. Dies betrifft die Zahl und Art der Werkzeuge, deren Verzahnung untereinander sowie die Weiterentwicklungen im FIS. Der grundsätzliche Ablauf und die Orientierung am schematischen Forschungsdatenlebenszyklus wird jedoch Bestand haben.

²⁷Forschungsdaten.info, [zugriffen](https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten) 16.12.2022, <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten>.