

Forschungsdatenmanagement und Publizieren von Forschungsdaten – Aufbau von Services am Beispiel der TU Wien

Barbara Sánchez Solís¹, Paloma Marín-Arraiza², Christiane Stork³, Magdalena Andrae⁴

Schlagwörter: Forschungsdatenmanagement, Forschungsdatenpublikation, Publikationsunterstützung, Datenmanagementplan, FAIR-Prinzipien, Open Data

Keywords: research data management, research data publication, publishing support, data management plan, FAIR principles, open data

Einleitung

In den letzten Jahren hat sich die Forschungspraxis substanziell geändert. Zurzeit spricht man von einem datengetriebenen wissenschaftlichen Paradigma, in dem die Forschungsdaten von der Planungsphase bis zur allfälligen Veröffentlichung der Forschungsergebnisse eine wesentliche und zentrale Rolle spielen. Der Forschungsoutput wird immer vielfältiger. Neben Artikeln werden vermehrt auch Code und Daten als echte Forschungsergebnisse berücksichtigt und nicht nur als Zusatzmaterialien betrachtet.

Eine geregelte Aufbewahrung und der Austausch von Forschungsdaten sind für die Wissenschaft von entscheidender Bedeutung. Sie ermöglichen den Forschenden, ihre eigenen Ergebnisse und die von anderen langfristig zu nutzen, zu analysieren und zu reproduzieren, um sie bewusst zu replizie-

1 Zentrum für Forschungsdatenmanagement, TU Wien

2 Services Bibliothek, TU Wien Bibliothek

3 Zentrum für Forschungsdatenmanagement, TU Wien

4 Fachgruppe Publikationsservices, TU Wien Bibliothek

ren oder zu widerlegen. Voraussetzung für diese Maßnahmen ist ein korrektes Forschungsdatenmanagement (FDM).

Innerhalb von Forschungseinrichtungen erfordert die Verwaltung von Forschungsdaten die gemeinsame Arbeit von Bibliotheken, Forschungssupport, IT-Abteilungen und Ansprechpersonen für Ethikfragen. Nur so können die organisatorischen, rechtlichen, ethischen und technischen Anforderungen im Managementprozess erfüllt werden. Die Zusammenarbeit zwischen verschiedenen Akteuren unterstreicht den interdisziplinären Charakter des FDM, erfordert aber gleichzeitig klar definierte Workflows. Modelle für institutionelle FDM-Strategien sind z. B. DIAMANT⁵ (Design an Information Architecture for Data MANagement Technologies) des deutschen BMBF-Projekts PODMAN und RISE-DE⁶ aus dem deutschen BMBF-Projekt FDMentor.

Ziel dieses Beitrags ist es, wesentliche Aspekte und wichtige Grundsätze des Forschungsdatenmanagements vorzustellen (Abschnitt 1), einen Einblick in die Möglichkeiten zur Veröffentlichung von Forschungsdaten zu geben (Abschnitt 2) und anhand eines Beispiels zu zeigen, wie institutionelle Services zum Thema FDM in der Praxis aussehen können (Abschnitt 3).

1 Forschungsdatenmanagement

Als Forschungsdaten gilt jegliches Material, das im Zuge wissenschaftlicher Arbeit z. B. durch Digitalisierung, Aufzeichnungen, Quellenforschung, Experimente, Messungen, Erhebungen und Interviews entsteht. Das beinhaltet z. B. Mess-, Analyse- und Umfragedaten, Bilder, Film- und Audioaufnahmen, aber auch Software und Code.

Das Forschungsdatenmanagement (FDM) erstreckt sich über den gesamten Lebenszyklus der Forschungsdaten und umfasst alle Prozesse der Planung, Sammlung, Auswahl, Transformation und Archivierung von Forschungsdaten, immer mit dem Ziel, diese langfristig und unabhängig vom

5 Siehe Katarina Blask und André Förster, »Designing an Information Architecture for Data Management Technologies: Introducing the DIAMANT Model«, *Journal of Librarianship and Information Science*, 2020, doi:10.1177/0961000619841419

6 Siehe Niklas K. Hartmann, Boris Jacob und Nadin Weiß, »RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement«, 2019, doi:10.5281/zenodo.2549344

Datenerzeuger aufzubewahren. Infrastrukturen für einen sicheren, geschützten Zugriff auf die Daten werden ebenso benötigt wie Lösungen für eine öffentliche Bereitstellung und die dafür notwendige Dokumentation, eindeutige Kennzeichnung und Lizenzierung. Die Ansprüche sind dabei nicht nur von Forschungsbereich zu Forschungsbereich sehr unterschiedlich, sondern auch für die jeweiligen Rollen: Forschende haben oft, und auch phasenabhängig, andere Bedürfnisse und Vorstellungen als ein Fördergeber oder ein Kooperationspartner aus der Wirtschaft.

Bei der Entwicklung einer geeigneten FDM-Strategie können Leitfäden und Checklisten unterstützen, zum Beispiel der »Leitfaden zum Forschungsdatenmanagement«⁷ des deutschen Projekts WissGrid oder die »Recommendations on Managing Research Data Addressed to Researchers«⁸ des spanischen Projekts Maredata.

1.1 Datenmanagementpläne

Idealerweise wird vor oder zu Beginn eines Forschungsvorhabens in einem Datenmanagementplan (DMP) erfasst, welche Forschungsdaten wie erzeugt werden und was mit den Daten während und auch nach Abschluss des Projekts geschieht. Hierzu zählen nicht nur Angaben über die Herkunft der Daten und Kontextinformationen rund um den Erhebungsprozess, sondern auch Informationen über allfällige Beschränkungen beim Zugriff auf die Daten, rechtliche und ethische Konformität sowie über eine spätere Zitierbarkeit und Langzeitverfügbarkeit oder gegebenenfalls auch Löschung. Ein DMP hilft bei der Planung eines Forschungsprozesses und bei der Festlegung von Verantwortlichkeiten in einem Forschungsprojekt, in dem mehrere Forschende bzw. Institutionen beteiligt sind. Er trägt somit auch dazu bei, einen erhöhten Ressourcenaufwand frühzeitig zu erkennen.

Da die Praktiken in Bezug auf das Managen von Daten in den einzelnen Disziplinen sehr unterschiedlich sind, verwenden manche Forschungsförderer keine spezifischen DMP-Vorlagen. Für solche Fälle liefert Science Europe

7 Jens Ludwig und Harry Enke, Hg., *Leitfaden zum Forschungsdaten-Management: Handreichungen aus dem WissGrid-Projekt* (Glückstadt: Hülsbusch, 2013).

8 Remedios Melero Melero et al., »Recommendations on Managing Research Data Addressed to Researchers« (2018), <http://hdl.handle.net/10609/98347>

mit den »Core Requirements for Data Management Plans«⁹ Anhaltspunkte, welche Mindestanforderungen ein DMP erfüllen sollte.

Die Mindestanforderungen werden wie folgt strukturiert:

1. Datenbeschreibung und Sammlung oder Wiederverwendung bestehender Daten
2. Dokumentation und Datenqualität
3. Speicherung und Sicherung während des Rechercheprozesses
4. Gesetzliche und ethische Anforderungen, Verhaltenskodizes
5. Teilen und Langzeitarchivierung von Daten
6. Verantwortlichkeiten und Ressourcen des Datenmanagements

Darüber hinaus gibt es mehrere Online-Tools zur Erstellung von DMPs. Die bekanntesten sind das DMPTool¹⁰ vom University of California Curation Center, DMPOnline¹¹ vom Digital Curation Centre, der Research Data Management Organiser (RDMO),¹² der in einem von der DFG geförderten deutschen Projekt entwickelt wurde, und der Data Stewardship Wizard.¹³ Dieses letzte Tool besteht aus einem intelligenten Fragebogen, der die Erstellung eines DMPs erleichtert.

Trotz der vorhandenen Tools und des großen Nutzens eines DMPs wird die DMP-Erstellung von Forschenden häufig als ein zusätzlicher bürokratischer und zeitaufwändiger Prozess und somit als Last angesehen. Um den manuellen Aufwand bei der Erstellung zu minimieren und gleichzeitig die Qualität und Nutzbarkeit der bereitgestellten Informationen zu optimieren, ist das Konzept von machine-actionable DMPs entstanden. Machine-actionable DMPs ermöglichen den Austausch von Informationen zwischen Systemen im Interesse der am Forschungsprozess beteiligten Stakeholder (Forschende, Forschungsförderer, Repositorienmanager und -managerinnen, IT-Provider, Bibliotheksmitarbeitende und andere).

9 Science Europe, Hg., »Practical Guide to the International Alignment of Research Data Management« (2019), zuletzt geprüft am 12.03.2020, <https://www.scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/>

10 <https://dmptool.org/>

11 <https://dmponline.dcc.ac.uk/>

12 <https://rdmorganiser.github.io/>

13 <https://ds-wizard.org/>

Derzeit werden einige machine-actionable DMP-Tools entwickelt. Ihre Ziele lassen sich wie folgt zusammenfassen:

- Steigerung der Qualität von DMPs, z. B. Automatisierung der DMP-Validierung
- Erleichterung des Prozesses für Forschende durch direkten Zugriff auf die Informationen beteiligter Akteure (Forschungssupport, IT etc.)
- Vermeidung von Doppelarbeit, z. B. bei der Eingabe von Publikationen
- Nutzung der institutionellen Infrastruktur wird gefördert (z. B. Storing-Angebote und Helpdesk)

Im Artikel »Ten principles for machine-actionable data management plans«¹⁴ sind die Grundsätze eines machine-actionable DMPs und die Vorteile, die er für die verschiedenen Stakeholder des Forschungsprozesses bietet, übersichtlich beschrieben. In Abb. 1 sind die zehn wichtigsten Grundsätze zusammengefasst.

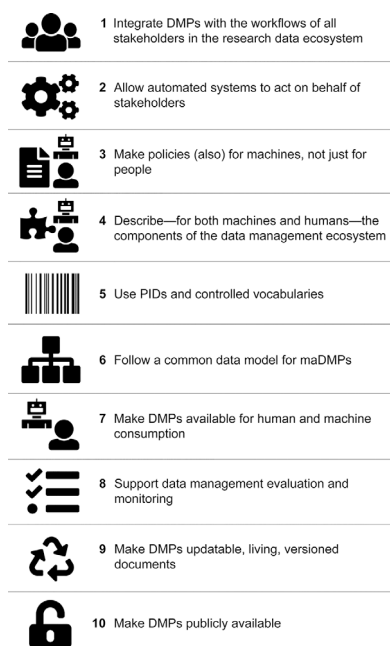


Abb. 1: Ten Principles for maDMPs at a Glance, doi:10.1371/journal.pcbi.1006750.g002. Aus: Miksa, Tomasz, Stephanie Simms, Daniel Mietchen und Sarah Jones. »Ten Principles for Machine-Actionable Data Management Plans.« PLoS computational biology 15, Nr. 3 (2019): e1006750. doi:10.1371/journal.pcbi.1006750.

14 Tomasz Miksa et al., »Ten Principles for Machine-Actionable Data Management Plans«, PLoS computational biology 15, Nr. 3 (2019), doi:10.1371/journal.pcbi.1006750

Beispiel des Workflows eines maDMP-Tools auf institutioneller Ebene

1. Die Mitarbeiterin bzw. der Mitarbeiter der Institution meldet sich mit ihren bzw. seinen institutionellen Zugangsdaten am Tool an.
2. Das Tool bietet einen Fragebogen, der an die Vorlagen der unterschiedlichen Fördergeber angepasst ist, und vermeidet so weit wie möglich Freitext-Eingaben.
3. Das Tool verbindet sich mit weiteren Systemen der Institution (z. B. CRIS-System, Projekt- und Publikationsdatenbank etc.) und füllt den DMP mit bestimmten Informationen aus diesen Quellen.
4. Je nach den Informationen über das Projekt und die Datensätze werden geeignete Repositorien vorgeschlagen, wenn das institutionelle Repositorium nicht ausreicht.
5. Das Tool verbindet sich mit dem institutionellen Repositorium, um den DMP und seine Updates zu speichern.
6. Falls es für den Fördergeber erforderlich ist, werden die Informationen in ein PDF-Format exportiert.

1.2 FAIR-Prinzipien und Open Data

Das FDM-Konzept koexistiert mit zwei weiteren, grundsätzlich voneinander unabhängigen Konzepten: den FAIR-Prinzipien und Open Data. Das Mini-Glossar (siehe nächste Seite) fasst die Kernpunkte der einzelnen Konzepte zusammen.

Obwohl alle drei Konzepte Daten (und Forschungsdaten) als Gegenstand haben, beziehen sie sich auf verschiedene Aspekte der Datenverwaltung. Jedes hat einen etwas anderen Fokus und eine andere Stärke, und gemeinsam angewendet bieten sie viel Potenzial für die Nachnutzung von Forschungsoutput. FDM ist dabei das Fundament: Wenn Daten nicht bereits in der Anfangsphase der Forschung richtig erstellt und gemanagt werden, ist es schwierig, sie später FAIR oder offen zu machen.

Ein gutes Datenmanagement ist notwendig – unabhängig davon, ob die Daten letztendlich publiziert oder mit gewissen Einschränkungen aufbewahrt werden. Voraussetzung für eine langfristige Nachnutzbarkeit der Forschungsdaten ist, dass sie mit einer detaillierten Dokumentation versehen und in einem

langlebigen, interoperablen Format in einem vertrauenswürdigen Datenrepositorium abgelegt werden. Auf diese Weise können sowohl die Arbeitsgruppen, die die Daten erhoben haben, als auch externe Forschende und Interessierte zukünftig auf die Daten zugreifen und sie wiederverwenden.¹⁵

Mini-Glossar*

Forschungsdatenmanagement: Reihe von Praktiken zum Umgang mit Daten, die während eines Forschungsprojektes wiederverwendet, gesammelt und erstellt werden.

*FAIR-Prinzipien:*** 15 Grundsätze, die sich für eine erhöhte Auffindbarkeit (F – *Findable*), Zugänglichkeit (A – *Accessible*), Interoperabilität (I – *Interoperable*) und Wiederverwendbarkeit (R – *Reusable*) von Forschungsdaten, Metadaten und weiteren Forschungsergebnissen einsetzen. Darüber hinaus legen sie Wert auf Menschen- und Maschinenlesbarkeit. → Daten, die die FAIR-Prinzipien erfüllen, können – müssen aber nicht – »offen« und somit für jedermann verfügbar sein. Die FAIR-Prinzipien erlauben auch eine Einschränkung des Datenzugangs, die in gewissen Fällen sinnvoll oder sogar erforderlich ist.

Open Data: Daten, die ohne Einschränkung oder mit minimalen Einschränkungen verfügbar sind. → Wenn ein Open-Datensatz gut dokumentiert und maschinenlesbar ist, eine offene Lizenz (z.B. CCO oder CC BY)*** hat, herstellerunabhängige Formate und offene Standards verwendet, entspricht er zusätzlich dem FAIR-Konzept.

Es sei darauf hingewiesen, dass die FAIR-Prinzipien und Open Data auf jeden Datensatz angewendet werden können, sei es auf Forschungsdaten, Daten des öffentlichen Sektors, Regierungsdaten oder andere Datentypen.

15 Einige Situationen, die mit einem guten FDM vermieden werden können, werden in diesem Video auf unterhaltsame Weise vorgestellt: Karen Hanson, Alisa Surkis und Karen Yacobucci, »Data Sharing and Management Snafu in 3 Short Acts« (NYU Health Sciences Libraries, 2012), doi:10.5446/31036

* Eine umfassendere Analyse der Konzepte: Rosie Higman, Daniel Bangert und Sarah Jones, »Three Camps, one Destination: The Intersections of Research Data Management, FAIR and Open«, *Insights: the UKSG journal* 32, Nr. 2016 (2019), doi:10.1629/uksg.468

** FAIR-Prinzipien – GoFAIR Initiative: <https://www.go-fair.org/fair-principles/>

*** Siehe z. B. Guido Kucsco und Adolf Zemann, »CCO 1.0 Universal – Beurteilung der Verzichtserklärung und der Lizenzerteilung im Rahmen der Fallback-Klausel nach österreichischem Recht«, <https://hdl.handle.net/11353/10.528411>

1.3 Beratungsleistungen

Die Ansprüche an ein effizientes FDM hängen stark von den unterschiedlichen Rollen, Blickwinkeln und Projektphasen ab. Die Bedürfnisse reichen von sicherer Verwaltung und einem geschützten Zugriff während des Forschungsprozesses bis zu Infrastrukturen, die es erlauben, Daten einem bestimmten Nutzerkreis zur Verfügung zu stellen oder, gemäß Förderrichtlinien, ausgewählte Datensätze Open Access im Internet zu publizieren. Ziel eines FDM-Beratungsservices ist es, die Forschenden entlang des gesamten Forschungsdaten-Lebenszyklus zu begleiten. Wie in der folgenden Übersicht zu erkennen ist, setzen viele Beratungsleistungen idealerweise schon im Vorfeld eines Projekts, d. h. in der Antragsphase, an.

Unterstützungsmöglichkeiten entlang des Forschungsdaten-Lebenszyklus

In der Pre-Projektphase:

- *Projektantragstellung*: Förderorganisationen verlangen die Beschreibung des geplanten Datenmanagements und des Umgangs mit ethischen Themen. Häufig ist eine Unterstützung bei der richtigen Formulierung in Förderanträgen sinnvoll. Für den Bereich »Ethik« kann ggf. die Kontaktperson für Ethikfragen an der eigenen Institution miteinbezogen werden.
- *Kosten für Datenmanagement* können bei vielen Fördergebern als verrechenbare Posten angegeben werden. Hilfestellung kann zum Beispiel durch eine Checklist mit möglichen Kostenfaktoren sowie durch Awareness-Workshops gemeinsam mit der Förderberatung der Institution erfolgen.

- *Datenmanagementpläne (DMPs)*: Hilfestellung bei der Suche nach Templates und beim Ausfüllen, DMP-Checks sowie Auskunft darüber, welche Information zu welchem Zeitpunkt bei den jeweiligen Fördergebern notwendig ist.
- *Rechtliche Konformität beim Datenmanagement*: Auskunft zu grundsätzlichen Fragen der Rechteinhaberschaft und zum Umgang mit personenbezogenen Daten. Für konkrete datenschutzrechtliche Themen sind die Rechts- und Datenschutzabteilungen einzubeziehen.
- *Lizenzen*: Für Daten und Software gibt es unterschiedliche Lizenzierungsmodelle.* Bei Primärdaten ist abzuklären, ob überhaupt ein urheberrechtlicher Schutz besteht. Nur in diesem Fall kann eine Lizenz vergeben werden. Erstorientierung bei Lizenzfragen kann angeboten werden, in manchen Fällen bedarf es jedoch einer juristischen Einzelfallprüfung bzw. einer Beratung durch die Expertinnen und Experten aus den Abteilungen Forschungssupport und Datenschutz.

In der Projektphase:

- *Infrastrukturen*: Beratung zu technischen Infrastrukturen und Tools zur sicheren Speicherung, zur Organisation und zum einfachen Austausch von Daten, wie Cloud-Services, Electronic Lab Notebooks etc.
- *Datendokumentation*, damit Nachnutzende Entstehungsprozess, Verarbeitung, Analyse und Interpretation der Daten nachvollziehen und die Ergebnisse reproduzieren können. Unterstützung und Bewusstseinsbildung können zum Beispiel im Rahmen von Workshops erfolgen.
- *Forschungssoftware*, die anderen zur Verfügung gestellt wird, muss ebenfalls nachvollziehbar dokumentiert sein. Hier ist eine Unterstützung bei Dokumentations- und Sharing-Fragen sinnvoll.

In der Post-Projektphase:

- *Publikation von Daten*: Für viele Fördergeber gelten Forschungsdaten als Ergebnisse, die ebenfalls verpflichtend Open Access zur Verfügung gestellt werden sollen. In Workshops und individuellen Beratungsterminen können die möglichen Wege vorgestellt werden.
- *Daten- und Coderepositorien*: Je nach Gegebenheiten vor Ort gilt es, die Forschenden im Umgang mit den hausinternen Systemen zu infor-

mieren bzw. bei der Suche nach geeigneten externen Systemen zu unterstützen.

- *Persistente Identifikatoren* sind unverzichtbar bei der Publikation von Daten. PIDs machen Daten, Datenerzeugerinnen und Datenerzeuger unverwechselbar und sichtbar, indem sie dauerhaft aufeinander referenzieren. Sie verhindern außerdem, dass ein Link ins Leere geht (Fehlermeldung »404 Not Found«). Hilfestellung kann bei der Auswahl geeigneter PIDs (z. B. Digital Object Identifier [DOI] für Daten und ORCID iD für Personen) gegeben werden.**

* Für urheberrechtlich geschützte Daten bieten sich die Creative Commons-Lizenzen (CC-Lizenzen) an: <https://creativecommons.org/licenses/>, siehe dazu auch den Beitrag »Urheberrecht und offene Lizenzen« in diesem Band. Bei der Auswahl einer passenden Lizenz für Software hilft dieses Tool: <https://choosealicense.com/>

** Zum Thema Sichtbarkeit und persistente Identifikatoren siehe auch den Beitrag »Sichtbarkeit und Auffindbarkeit« in diesem Band.

2 Veröffentlichung von Forschungsdaten

Unter der Prämisse »so offen wie möglich, so geschlossen wie nötig« fordern immer mehr Forschungsförderer und Forschungseinrichtungen die Veröffentlichung von Forschungsdaten. Zumindest die sogenannten zugrundeliegenden Daten, d. h. die Daten, die zu den Publikationsergebnissen führen, sollen frei zugänglich gemacht werden.

Wenn es sich bei den Forschungsdaten um personenbezogene, sensible oder vertrauliche Daten handelt, d. h. rechtliche und ethische Aspekte zu berücksichtigen sind,¹⁶ oder eine Embargofrist besteht, z. B. in Hochschul-Industriepartnerschaften, können die Daten nicht offen publiziert werden. Im ersten Fall ist es manchmal möglich, die Daten zu anonymisieren¹⁷ und die anonymisierten Daten zu veröffentlichen. Im zweiten Fall können die

16 Die Plattform forschungsdaten.info hat einen Entscheidungsbaum veröffentlicht, der die wesentlichen rechtlichen Aspekte der Datenveröffentlichung nach der aktuellen Verordnung (DSGVO) abdeckt. <https://bit.ly/2Oay9iY> Die genannten Paragraphen beziehen sich auf die entsprechenden deutschen Gesetze.

17 Zwei Werkzeuge, die für die Datenanonymisierung frei verwendet werden können, sind Anonymization ToolBox www.cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php?go=home und Amnesia <https://amnesia.openaire.eu/>

Daten nach Ablauf der Embargofrist veröffentlicht werden. Für den Fall, dass die Daten aus verschiedenen Gründen nicht publiziert werden können, ist es interessant, zumindest die entsprechenden deskriptiven Metadaten zu veröffentlichen.

Der Offenheitsgrad der Daten bestimmt auch die Art des PID, der verwendet werden kann. DOIs sind nicht geeignet, wenn weder Daten noch Metadaten veröffentlicht werden können. Für geschlossene Daten und Metadaten wird empfohlen, Handle¹⁸ oder Archival Resource Key (ARK)¹⁹ zu verwenden. Bei der Veröffentlichung von Daten können verschiedene Wege eingeschlagen werden. Drei dieser Wege werden im Folgenden beschrieben: (Daten-) Repositorien, Data Papers und Enhanced Publications.

2.1 (Daten-)Repositorium

Ein Repositorium ermöglicht die langfristige Aufbewahrung, die Publikation und das Auffinden digitaler Daten. Forschende und Forschungsgruppen können ihre Daten gemeinsam mit Kontextinformation hochladen und ihre Nutzungsbedingungen definieren. Metadaten und persistente Identifikatoren gewährleisten die Zitierbarkeit von Datensets und erhöhen die Sichtbarkeit von Forschenden und ihren zugehörigen Forschungsinstitutionen. Einige Repositorien akzeptieren alle Arten von Dateiformaten (z. B. Zenodo),²⁰ andere konzentrieren sich auf bestimmte Formate (z. B. HEPData).²¹

Immer mehr Institutionen entscheiden sich für den Aufbau eines internen Datenrepositoriums oder die Erweiterung ihres Dokumentenservers zur Aufbewahrung von Daten. Die Kapazität eines solchen Speichers, die akzeptierten Datenformate sowie die Möglichkeit der Aufbewahrung von Software oder Datenbanken hängen von den Bedürfnissen und dem akademischen Profil der Institution ab.

Neben den institutionellen Repositorien gibt es Fachrepositorien in verschiedenen Wissensgebieten. Einer der Vorteile eines Fachrepositoriums ist die erhöhte Sichtbarkeit der Daten in einem bestimmten Bereich und die

18 <https://www.handle.net/>

19 https://nzt.net/e/ark_ids.html

20 <https://zenodo.org/>

21 <https://www.hepdata.net/> (Format: YAML-Format zu einer Datei .zip, .tar, .tar.gz oder .tgz komprimiert).

Vernetzung innerhalb von Fachcommunities. Die Zeitschrift *Scientific Data* (Nature) bietet eine Liste²² von empfohlenen Fachrepositorien, in denen Autorinnen und Autoren Daten zu Publikationen hinterlegen können. Ein Beispiel für ein Fachrepositorium in Österreich ist AUSSDA²³ – The Austrian Social Science Data Archive.

Für die Auswahl eines geeigneten Repositoriums können folgende Kriterien in Betracht gezogen werden:

1. Das Repositorium muss den individuellen Datenanforderungen (z. B. akzeptierte Dateiformate, angebotene Metadatenfelder, Zugriff, Sicherung) entsprechen
2. Vergabe von persistenten Identifikatoren (PIDs)
3. Kostenmodelle – der Preis kann z. B. von der Datenmenge und der Speicherdauer abhängen
4. Klare Bedingungen, die den gesetzlichen Anforderungen und der DSGVO entsprechen (z. B. Datenschutz)
5. Lizenzmodelle (z. B. Creative Commons, spezielle Lizenzen für Software)
6. Zugangsmodelle des Repositoriums (z. B. öffentlich, nur für einen eingeschränkten Nutzerkreis oder geschlossen)
7. Möglichkeit, den Zugriff auf die Daten zeitlich zu regulieren (Embargo-fristen)
8. Funktionalitäten wie Zugriffs- und Downloadstatistiken, Zitationsvorschläge etc.
9. Langfristiges Bestehen des Dienstleisters
10. Rechtsraum des Repositoriums (EU oder außerhalb)

Fördergeber erwähnen häufig »vertrauenswürdige Datenarchive« (*trustworthy repositories*) in ihren Anforderungen. Dazu zählen in erster Linie Repositorien mit einer Zertifizierung wie z. B. dem CoreTrustSeal²⁴, dem Nestor Seal²⁵ oder nach ISO 16363.²⁶ Da derzeit jedoch viele Repositorien erst im

22 <https://www.nature.com/sdata/policies/repositories>

23 www.aussda.at

24 <https://www.coretrustseal.org/why-certification/certified-repositories/>

25 https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html

26 www.iso16363.org/iso-certification/certified-clients/

Aufbau sind und die Zertifizierungsprozesse einige Zeit in Anspruch nehmen, verweisen Förderorganisationen wie der FWF vorerst nur darauf, dass das verwendete Repository auf re3data²⁷ gelistet sein soll.

2.2 Data Paper

Ein Data Paper, auch *data descriptor* und *article of data* genannt, ist eine Veröffentlichung von dokumentierten Datensätzen in einem Data Journal. Es beschreibt einen Datensatz oder eine Sammlung von Datensätzen, einen Prozess oder eine Software. Neue Methoden oder innovative Schlussfolgerungen sind in einem Data Paper nicht erforderlich. Das Data Paper enthält Informationen darüber, wann, wie und warum die Daten erhoben wurden, woraus das Ergebnis besteht und gegebenenfalls einen Link zu ergänzenden Materialien, wie zum Beispiel den Rohdaten.²⁸

Data Papers werden in Data Journals veröffentlicht. Eine Liste von Data Journals ist auf der Website forschungsdaten.org²⁹ zu finden.

In einigen Fällen erfolgt die Veröffentlichung von Daten in einem Data Paper mit Hilfe eines Datenrepositoriums. In anderen Fällen bietet das Data Journal selbst eine Möglichkeit der Datenarchivierung.

Beim Schreiben eines Data Papers gibt es keine definierte Struktur, auch weil es sich um eine neue Art der Veröffentlichung handelt. Dennoch wird empfohlen, dass ein Data Paper zumindest aus den folgenden Abschnitten besteht:

- *Einführung*: Kontextinformation und eine Erklärung des Potenzials des Datensatzes.
- *Methoden*: Art der Datensammlung, Probenahmemethoden, benutzte Materialien und Codes, Nutzungslizenzen.
- *Beschreibung des Datensatzes*: Deskriptoren und Identifikatoren der Daten.

²⁷ <https://www.re3data.org/>

²⁸ Beispiel eines Data Papers, das im Data Journal »Data« veröffentlicht wurde: <https://doi.org/10.3390/data3040047>; Für die Veröffentlichung der Rohdaten wurde das allgemeine Repository Zenodo verwendet: <http://doi.org/10.5281/zenodo.1470515>

²⁹ Liste von Data Journals: https://www.forschungsdaten.org/index.php/Data_Journals

2.3 Enhanced Publications

In groben Zügen ist eine *Enhanced Publication* die digitale Publikation eines wissenschaftlichen Artikels, der mit Zusatzinformationen verknüpft ist (siehe dazu auch den Beitrag »Publizieren in wissenschaftlichen Zeitschriften« in diesem Band). Die zusätzlichen Informationen können Forschungsdaten (Evidenz der Forschung – *supporting information*), ergänzende Materialien (zur Veranschaulichung oder Verdeutlichung) oder Daten nach der Veröffentlichung (Kommentare, Rankings) sein.³⁰ Im Idealfall haben sowohl der Artikel als auch die Zusatzinformationen (auch *Subparts* genannt) einen persistenten Identifikator und deskriptive Metadaten.

Diese Verknüpfungen zwischen Artikeln und ihren Zusatzinformationen bilden eine Struktur ähnlich eines Graphen, wobei die Knoten die Informationen sind und die Kanten die semantischen Verbindungen (z. B. »is part of«, »has version«, »generates«).

Für die »Erweiterung« der Publikationen kann sowohl die Plattform der Zeitschrift als auch eine externe Infrastruktur (z. B. ein Repositorium oder eine virtuelle Forschungsumgebung) genutzt werden. Ein Beispiel für Zeitschriften, die die eigene Plattform nutzen, sind die PLOS-Zeitschriften. Hier erhält jeder Unterteil (Tabellen, Abbildungen etc.) einen eigenen DOI.³¹

3 Institutionelle Services zum FDM an der TU Wien

Die TU Wien baut derzeit Services zum FDM, zu Publikationen und dem Nachweis von Forschungoutput aus. Wichtig ist dabei, dass alle Systeme miteinander »kommunizieren« und möglichst viel Information automatisiert austauschen können. Im Pilotprojekt »Data Management Infrastructure for the TU Wien« sollen Konzepte für automatisierte Datenmanagement-Workflows und die Implementierung eines machine-actionable DMP Tools realisiert werden.

30 Weitere Informationen: Saskia Woutersen-Windhouwer und Marjan Vernooij-Gerritsen, *Enhanced Publications: Linking Publications and Research Data in Digital Repositories*, Trends in Research Information Management (Amsterdam: Amsterdam University Press, 2009), <https://hdl.handle.net/11245/1.311760>

31 Beispiel in PLOS One: <https://doi.org/10.1371/journal.pone.0218168>

Die TU Wien Bibliothek hat das breite Aufgabenfeld von FDM thematisch vorbereitet, indem sie den Lead für die Erstellung einer Policy für Forschungsdatenmanagement³² übernommen hat, die im Juli 2018 durch das Rektorat verabschiedet wurde.³³ Die Bibliothek hat auch eine koordinierende Einheit mit zwei Stellen für FDM-Fragen konzipiert.

3.1 Das Zentrum für Forschungsdatenmanagement

Zeitgleich mit der FDM-Policy konnte somit an der TU Wien ein bereichsübergreifendes Zentrum für Forschungsdatenmanagement eingeführt werden. Um den breiten Bogen von organisatorischen und technischen Aufgabenfeldern abzudecken und größtmöglichen Outreach innerhalb der Institution zu erreichen, sind drei der insgesamt fünf Zentrums-Mitarbeiterinnen und Mitarbeiter zusätzlich weiteren Bereichen innerhalb der TU Wien zugeordnet: dem Fachbereich Europäischer und internationaler Forschungssupport (EIFS), IT Solutions sowie dem Forschungsbereich Information und Software Engineering. Eine enge thematische Zusammenarbeit mit der Bibliothek ergibt sich, abgesehen von der räumlichen Ansiedlung, im Bereich PID-Services und mit den Fachreferentinnen und Fachreferenten für disziplinspezifische Fragen. Die inhaltliche Arbeit des Zentrums wird durch einen regelmäßigen Lenkungsausschuss mit Einbezug des Vizerektorats Forschung und Innovation und des Vizerektorats Digitalisierung und Infrastruktur gesteuert.

Das Zentrum für Forschungsdatenmanagement hat es sich zur Kernaufgabe gemacht, gemeinsam mit anderen Abteilungen an der Universität zur Unterstützung von Forschungsdatenmanagement und zum koordinierten Ausbau von Services beizutragen. Für Forschende und Projekttreibende ist das Zentrum eine zentrale Anlaufstelle zur Unterstützung beim Umgang mit Forschungsdaten.

32 TU Wien Policy für Forschungsdatenmanagement: https://www.tuwien.at/fileadmin/Assets/dienstleister/Datenschutz_und_Dokumentenmanagement/Policy__Forschungsdatenmanagement.pdf

33 Zur Erstellung von Policies siehe auch den Beitrag »Entwicklung von Policies und Strategien« in diesem Band.

3.2 Forschenden-Befragungen

In den ersten Monaten nach Gründung des Zentrums lag der Fokus auf der Identifizierung von existierenden FDM-Services, z. B. Data Storing-, Sharing- und Archiving-Services, sowie auf der Bedarfserhebung. Daher führte das Zentrums-Team eine Befragung zum Thema FDM durch. Diese erfolgte in Form von semistrukturierten Leitfadeninterviews mit ausgewählten Forschenden aus allen Fakultäten der Universität und mit drei Fachbereichen. Vorrangiges Ziel der Befragung war es, einen Einblick in die FDM-Praxis der unterschiedlichen Forschungsbereiche zu erlangen und etwaige Defizite und Bedürfnisse aus Sicht der Forschenden zu erfahren.

Die Ergebnisse zeichnen ein heterogenes Bild von Forschungsdaten an der TU Wien. Code und Software machen einen beträchtlichen Anteil aus. Im Hinblick auf das FDM gibt es deutliche Unterschiede zwischen den einzelnen Organisationsbereichen. Ein gutes Bewusstsein demonstrieren die meisten bei der sicheren Speicherung von Forschungsdaten und dem Datenaustausch mit anderen (z. B. über die TUproCloud³⁴) während des Forschungsprozesses. Defizite gibt es jedoch in den Bereichen, die über die eigentliche Projektphase hinauswirken, d. h. wenn es um die langfristige Aufbewahrung oder um die Publikation der Daten geht. Diesbezügliches FDM erfolgt nur in wenigen Fällen geordnet, oft fehlen sowohl die Dokumentation der Daten als auch klare Zuständigkeitsregelungen. Eine Herausforderung für die Nachnutzung der Daten ist neben der Sensibilität und Eigentümerschaft auch die Verwendung von kommerzieller (Mess-)Software mit eigenen Datenformaten.

Ein Kulturwandel zugunsten einer Nachnutzung von Daten und Software, auch im Sinne von Open Data, wird von einigen Forschenden erwartet, und viele sind sich bewusst, dass das Thema FDM nicht zuletzt aufgrund neuer Anforderungen von Fördergebern und Verlagen schon bald merklich an Bedeutung gewinnen wird. Darüber hinaus wird vermutet, dass die

34 Die TUproCloud ist ein TU-Wien-eigener Cloud-Dienst zum einfachen Hochladen, Speichern, Teilen und kollaborativen Bearbeiten von Daten und Dokumenten, der auch für externe Projektpartner freigegeben werden kann: <https://www.it.tuwien.ac.at/services/kooperation-und-kommunikation/collaboration/tuprocloud-sync-und-share-fuer-projekte/>

Interdisziplinarität in der Forschung und die interne und externe Vernetzung weiter zunehmen werden.

Zusätzlich zu den Leitfadeninterviews wertete das Zentrums-Team eine quantitative Online-Befragung von Forschenden aus. Die Daten entstammen einer Umfrage zu Forschungsdaten, die im Zuge des Projektes e-Infrastructures Austria³⁵ im Jahr 2015 in Österreich durchgeführt wurde. Die Antworten für die TU Wien (Anm. N=328) wurden noch einmal separat ausgewertet und in einem Bericht³⁶ zusammengefasst. Der Auswertung können wesentliche Informationen entnommen werden, wie z. B. der Bedarf der Umfrageteilnehmer an unterstützenden Angeboten im Umgang mit Forschungsdaten (siehe Abb. 2).

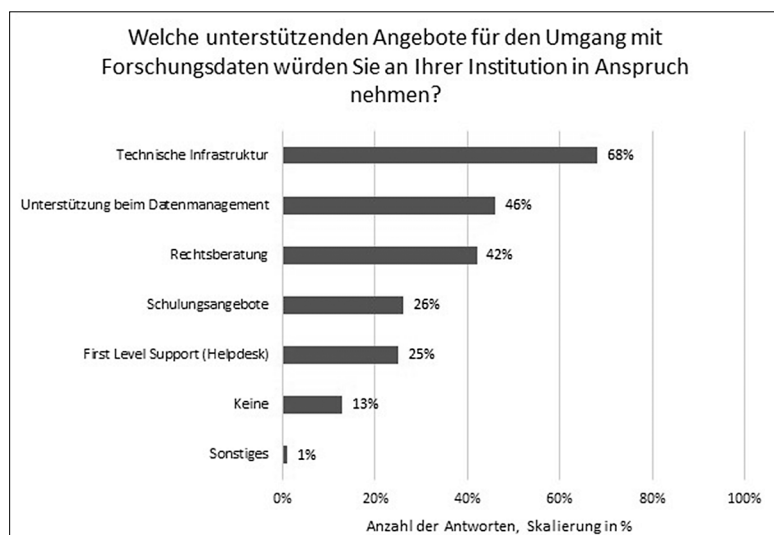


Abb. 2: Bedürfnisse der Forschenden der TU Wien im Zusammenhang mit Forschungsdaten. Quelle: Forschende der TU Wien und ihre Daten, <http://doi.org/10.5281/zenodo.2668768>.

³⁵ <https://e-infrastructures.univie.ac.at/>

³⁶ Malgorzata Z. Goraczek, Barbara Sánchez Solís und Christiane Stork, »Forschende der TU Wien und ihre Daten. Darstellung der Ergebnisse der quantitativen Online-Befragung an der TU Wien« (2019), doi:10.5281/zenodo.2668768

3.3 Zusammenarbeit Publikationsservices und Zentrum für FDM

Eine gute Zusammenarbeit zwischen Publikationsservices und dem Zentrum für Forschungsdatenmanagement ist besonders in der Beratung wichtig. Nachfolgend wird ein Use Case vorgestellt, der eine solche Kollaboration illustriert.

Use-Case

Ausgangslage:

Eine Forschungsgruppe der Universität nimmt an einem H2O2O-Projekt teil. In ihrem Vertrag gibt es eine Klausel über Open Data und Open Access, die die Ergebnisse müssen vor Ende des Projekts in einem Repositorium abgelegt werden.

Fragen: Wir haben einen Artikel veröffentlicht und der Herausgeber ermöglicht es uns, den Postprint in einem Repositorium zu archivieren. Darüber hinaus müssen wir die zugrundeliegenden Daten des Artikels veröffentlichen. Wie machen wir das? Welches Repositorium ist geeignet?

Lösung:

Für den Artikel: Den Forschenden wird vorgeschlagen, ihr Manuskript in ein Repositorium mit OpenAIRE-Anbindung hochzuladen. Wir empfehlen ihnen Repositorien wie *repositUM* (institutionell), *Zenodo* (übergreifend) oder ein fachspezifisches; das Manuskript erhält dort einen PID und ist Open Access verfügbar. Wir erklären die OA-Vorgaben für Publikationen in H2O2O (Version, Embargo, Ablage) sowie das Vorgehen, wenn diese dem unterzeichneten Vertrag mit dem Verlag widersprechen (Verhandlungsnotwendigkeit). Wir bieten an, bei längeren Literaturlisten die Vorgaben der Verlage zur Selbstarchivierung zu ermitteln.

Für die Daten: Für die Daten wird ein fachspezifisches Datenrepositorium gesucht. Das Repositorium muss für die Datensätze PIDs vergeben, offene Lizenzen zulassen und über ausreichende deskriptive Metadaten verfügen. Wir wählen ein Repositorium und erklären, wie der Upload-

Prozess abläuft. Besonderes Augenmerk wird auf die korrekte Dokumentation der Daten zur Nachnutzung gelegt.

Akteure: Zentrum für Forschungsdatenmanagement und Publikationsservices

3.4 Perspektiven

Wie aus den Interviews und den Ergebnissen der Onlinebefragung hervorgeht, gibt es an der TU Wien neben dem Bedürfnis nach allgemeiner und individueller Unterstützung beim FDM einen starken Wunsch nach Automatisierung und technischer Unterstützung.

Derzeit entwickelt die TU Wien einen Piloten (Proof of Concept) für ein automatisiertes, machine-actionable DMP-Tool. Ziel ist es, den manuellen Aufwand für die Forschenden zu minimieren und gleichzeitig die Qualität der bereitgestellten Informationen zu erhalten. Die Lösung basiert auf maschinell verarbeitbaren DMPs, den DMPs der zukünftigen Generation. Dieses Konzept inkludiert TU-Wien-interne Systeme wie z. B. Personendatenbank, Projektdatenbank, Forschungsinformationssystem etc. und erfordert daher beträchtliche Entwicklerarbeit. Das Modell lässt sich dennoch auch extern anpassen.

Im Hinblick auf ein Datenrepositorium bietet die TU Wien derzeit noch kein eigenes Service an, die Zentrums-Mitarbeiterinnen und -Mitarbeiter verweisen bei Nachfrage auf disziplinübergreifende Repositorien wie Zenodo oder fachspezifische Lösungen. Der Bedarf an TU-eigenen Repositorien und deren schrittweise Implementierung werden derzeit untersucht. Ziel ist der Aufbau von ineinandergreifenden Systemen (*integrated research data ecosystem*), in denen Forschung stattfindet und in denen Daten und Code von Forschenden und Partnern der TU Wien geteilt, archiviert und veröffentlicht werden können.

Der Ausbau soll schrittweise erfolgen. In der ersten Phase gilt es, den Grundstein für Weiterentwicklungen und Systemintegrationen zu legen. In dieser Phase soll ein Repositorium für Daten eingerichtet werden, das Funktionalitäten wie z. B. Self-Upload, ein abgestuftes Rollen- und Rechte-management, verschiedene Zugriffsarten von *closed* bis *open*, DOI-Vergabe, gute externe Sichtbarkeit der veröffentlichten Ressourcen etc. aufweist. Bestehende Systeme (z. B. Authentifizierungsserver) müssen verknüpft und

externe (z. B. PID-Provider) angebunden werden. Das gewählte Framework muss auch Programmierschnittstellen bereitstellen, die eine zukünftige Integration des Repositoriums mit Systemen ermöglichen, in denen Daten erstellt und verarbeitet werden. Um dem Bedarf der Forschenden an der TU Wien gerecht zu werden, soll zusammen mit dem Datenrepositorium auch ein zentrales, Git-basiertes Coderepositorium aufgebaut werden. In einem weiteren Schritt soll ein Datenbankenrepositorium, das auch komplexe, dynamische Daten wie z. B. 3D-Daten, Visualisierungen, Simulationen und andere Modelle annehmen kann, das Gesamtsystem vervollständigen.

Sowohl die Bibliothek als auch das Zentrum für Forschungsdatenmanagement teilen sich wichtige Aufgaben bei der Unterstützung der wissenschaftlichen Tätigkeit und der Veröffentlichung von Forschungsbildung. Um eine gute Zusammenarbeit zwischen beiden Einheiten zu gewährleisten und mögliche Überschneidungen von Aufgaben zu vermeiden, ist es daher notwendig, entsprechende Workflows zu etablieren und stetig anzupassen. Diese Arbeitsabläufe betreffen nicht nur die interne Arbeit, sondern auch die Kommunikation von Dienstleistungen für Forschende.

Weiterführende Literatur

- Clare, Connie, Maria Cruz, Elli Papadopoulou, James Savage, Marta Teperek, Yan Wang, Iza Witkowska und Joanne Yeomans. *Engaging Researchers With Data Management*. Open Book Publishers, 2019. doi:10.11647/OBP.0185.
- Mons, Barend. *Data Stewardship for Open Science*. Chapman and Hall/CRC, 2018. doi:10.1201/9781315380711.

Hilfreiche Weblinks

- Australian National Data Service. 23 (Research Data) Things: <https://www.ands.org.au/working-with-data/skills/23-research-data-things>
- Digital Curation Centre. Using RISE, the Research Infrastructure Self Evaluation Framework: www.dcc.ac.uk/resources/how-guides/RISE
- LEARN Toolkit of Best Practice for Research Data Management: <http://learn-rdm.eu/wp-content/uploads/RDMToolkit.pdf>
- MANTRA. Research Data Management Training: <https://mantra.edina.ac.uk/>

Kurzbiografien

Mag. Barbara Sánchez Solís leitet das Zentrum für Forschungsdatenmanagement an der TU Wien. Zuvor war sie für das Projektmanagement bei AUSSDA – The Austrian Social Science Data Archive (www.aussda.at) zuständig. An der Universität Wien koordinierte sie die erste Phase des österreichweiten, vom Bundesministerium für Bildung, Wissenschaft und Forschung geförderten Projekts e-Infrastructures Austria, das sich mit der Entwicklung von e-Infrastrukturen und Themen des Forschungsdatenmanagements befasste. ORCID iD: <https://orcid.org/0000-0002-3574-2755>

Dr. Paloma Marín-Arraiza absolvierte an der Universität Granada (Spanien) ihr Diplom in der Physik sowie einen Master in Information und wissenschaftlicher Kommunikation. Ihr Doktorat über Enhanced Publications, Semiotik und Forschungsdatenmanagement machte sie an der Universität Sao Paulo (Brasilien). Seit September 2018 ist sie Informationsmanagerin für PIDs an der Bibliothek der TU Wien. ORCID iD: <https://orcid.org/0000-0001-7460-7794>

DI. Christiane Stork ist Bauingenieurin mit einer Zusatzausbildung in der Umweltverwaltung und internationaler Erfahrung im Bereich Siedlungswasserwirtschaft. Sie hat den Zertifikatskurs Data Librarian an der Universität Wien absolviert und ist seit August 2018 Mitarbeiterin im Zentrum für Forschungsdatenmanagement der TU Wien. ORCID iD: <https://orcid.org/0000-0003-3176-4305>

MMag. Magdalena Andrae schloss neben den Studien Geschichte und Vergleichende Literaturwissenschaft 2013 den Lehrgang »Library and Information Studies« an der Universität Wien ab und leitete ab 2014 die Zeitschriftenabteilung an der Universitätsbibliothek in Salzburg. Seit 2016 ist sie an der TU Wien Bibliothek für Open Access zuständig. ORCID iD: <https://orcid.org/0000-0001-9174-7449>

Materialiensammlung zum Thema Publikationsberatung:
<https://doi.org/10.25364/publikationsberatung-materialien>

