Forschungsdaten
Sammeln, sichern, strukturieren

WissKom 2019

8. Konferenz der Zentralbibliothek, Forschungszentrum Jülich, 4. – 6. Juni 2019

Bernhard Mittermaier (Hrsg.)

Proceedingsband

Bibliothek / Library
Band / Volume 23
ISBN 978-3-95806-405-8

Mitglied der Helmholtz-Gemeinschaft

JÜLICH Forschungszentrum

Schriften des Forschungszentrums Jülich
Reihe Bibliothek / Library Band / Volume 23

Forschungszentrum Jülich GmbH
Zentralbibliothek, Verlag

Forschungsdaten
Sammeln, sichern, strukturieren

8. Konferenz der Zentralbibliothek,
Forschungszentrum Jülich

Bernhard Mittermaier (Hrsg.)

4. – 6. Juni 2019
Proceedingsband

Schriften des Forschungszentrums Jülich
Reihe Bibliothek / Library — Band / Volume 23

ISSN 1433-5557 — ISBN 978-3-95806-405-8

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über http://dnb.d-nb.de abrufbar.

Herausgeber und Vertrieb:	Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag 52425 Jülich Tel.: +49 2461 61-5368 Fax: +49 2461 61-6103 zb-publikation@fz-juelich.de www.fz-juelich.de/zb
Umschlaggestaltung:	Grafische Medien, Forschungszentrum Jülich GmbH
Druck:	Grafische Medien, Forschungszentrum Jülich GmbH
Copyright:	Forschungszentrum Jülich 2019

Schriften des Forschungszentrums Jülich
Reihe Bibliothek / Library, Band / Volume 23

ISSN 1433-5557
ISBN 978-3-95806-405-8

Vollständig frei verfügbar über das Publikationsportal des Forschungszentrums Jülich (JuSER)
unter www.fz-juelich.de/zb/openaccess.

This is an Open Access publication distributed under the terms of the Creative Commons Attribution License 4.0,
which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Inhaltsverzeichnis

Konferenzkomitees

Programmkomitee

Dr. Rafael Ball	ETH Zürich, ETH-Bibliothek
Dr. Bernhard Mittermaier	Forschungszentrum Jülich, Zentralbibliothek
Prof. Dr. Matthias S. Müller	RWTH Aachen, IT Center
Prof. Dr. Achim Oßwald	TH Köln, Institut für Informationswissenschaften
Dagmar Sitek	Deutsches Krebsforschungszentrum (DKFZ), Zentralbibliothek

Organisationskomitee

Thomas Arndt	Forschungszentrum Jülich, Zentralbibliothek
Yasmin Fattah	Forschungszentrum Jülich, Zentralbibliothek
Kai Pelzer	Forschungszentrum Jülich, Zentralbibliothek
Dr. Christoph Holzke	Forschungszentrum Jülich, Zentralbibliothek
Hilde Dobbelstein	Forschungszentrum Jülich, Zentralbibliothek
Dr. Bernhard Mittermaier	Forschungszentrum Jülich, Zentralbibliothek

WissKom2019 – Forschungsdaten
Sammeln, sichern, strukturieren

Die Wissenschaft erlebt durch die Informationstechnologie einen ihrer größten Umbrüche. Doch obwohl Computer und Netzwerke seit Jahrzehnten zum Forschungsalltag gehören, ist der Grad der Digitalisierung in den meisten Fachdisziplinen noch unzureichend. Der überwiegende Teil der Daten ist kaum dokumentiert und wird mit Ad-hoc-Verfahren verarbeitet und analysiert. Gleichzeitig steht die Wissenschaft vor neuen globalen Herausforderungen: Die Reproduzierbarkeitskrise wird durch nicht oder nicht vollständig publizierte Daten verschärft, die Verdopplung von Experimenten verschwendet Ressourcen und gefährdet Versuchstiere, und mit der Datenwissenschaft ist ein neuer Wissenschaftszweig auf dem Vormarsch, dessen Erfolg entscheidend von gutem, globalem Datenmanagement abhängt.

Dabei ist die größte Hürde zu besserem Datenmanagement nicht technischer Natur. Es müssen Aufmerksamkeit geschaffen und Berührungsängste abgebaut sowie alte Workflows angepasst und neue Anreize gefunden werden. Primär kommen hierfür Schulungen und Beratungsangebote zum Einsatz. Zudem bieten Datenmanagementpläne einen guten Ansatz, Forschern einen niederschwelligen Zugang zu dem Thema zu öffnen.

In der WissKom2019 beschäftigen wir uns mit diesen und anderen Aspekten des noch vergleichsweise jungen Themas Forschungsdatenmanagement, das die meisten Institutionen in den Bibliotheken ansiedeln. Dabei koexistieren unterschiedliche Reifegrade. Wir bilden dies in den Sessions ab, welche von „Strategie“ bis „Lessons learned“ reichen. Die achte Konferenz der Zentralbibliothek will Ansätze aufzeigen, wie Bibliotheken dieses Zukunftsfeld entscheidend mitgestalten können.

Ich danke allen herzlich, die zum Gelingen der Konferenz beigetragen haben: Den Vortragenden und Moderatoren, den Ausstellern und Sponsoren, den Mitgliedern des Programmkomitees, den Organisatoren und last but not least allen Teilnehmerinnen und Teilnehmern.

Dr. Bernhard Mittermaier
Leiter der Zentralbibliothek, Forschungszentrum Jülich

Festvortrag

Die FDM-Utopie und der Weg dorthin

Ania López

UB Uni Duisburg-Essen

Abstract

Digitalization has opened doors for science not only to discuss the results of their research, but also to access the data underlying the results in order to make the results directly traceable and reusable. Good research data management (RDM) is the essential core of this. While research data management has been part of good scientific practice in some scientific disciplines for decades, other disciplines have only begun to discuss what research data actually is in the course of science policy discussions over the past five years. In addition, there is an equally large discrepancy on the part of the infrastructure facilities. While some may look back on decades of experience in developing and operating digital research services, others are trying to approach new fields of activity through the topic of research data management. While research data management fundamentally reflects how scientific work functions digitally, it also challenges best practices and methods. Thus, it shakes the basic understanding of the scientific system.

In an ideal world, research data management represents the foundation of a utopia of the digital science landscape. In the course of the highly dynamic development of the topic, infrastructure facilities seek their place in this ideal scenario, just as scientific actors face the high suffering pressure of an exploding number of research data and methods.

But what does an ideal scenario research data management look like and how do current projects and activities sort themselves into it? Will the FDM utopia emerge from the meaningful integration of the German National Research Data Infrastructure (NFDI) and the European Open Science Cloud?

By presenting selected RDM actors and projects on the German and European level, an attempt is made to draw a picture of the current and future RDM landscape.

Im Zuge der Digitalisierung haben sich für die Wissenschaft die Türen geöffnet, nicht nur über die Ergebnisse ihrer Forschung zu diskutieren, sondern auch auf die den Ergebnissen zugrundeliegenden Daten zuzugreifen, um die Ergebnisse direkt nachvollziehbar und nachnutzbar zu machen. Essentieller Kern davon ist gutes Forschungsdatenmanagement (FDM). Während Forschungsdatenmanagement in

einigen wissenschaftlichen Disziplinen schon seit Jahrzehnten zur guten wissenschaftlichen Praxis gehört, beginnen andere Disziplinen erst im Zuge der wissenschaftspolitischen Diskussionen seit rund fünf Jahren darüber zu diskutieren, was Forschungsdaten überhaupt sind. Daneben findet sich eine ebenso große Diskrepanz seitens der Infrastruktureinrichtungen wieder. Während die einen auf jahrzehntelange Erfahrung mit dem Aufbau und Betrieb von Diensten für digitale Forschung zurückschauen können, versuchen sich andere über das Thema Forschungsdatenmanagement neuen Aufgabenfeldern zu nähern. Während Forschungsdatenmanagement grundlegend widerspiegelt wie wissenschaftliches Arbeiten digital funktioniert, stellt es gleichzeitig bewährte Praktiken und Methoden in Frage. Somit rüttelt es an dem Grundverständnis des wissenschaftlichen Systems.
In einer idealen Welt stellt Forschungsdatenmanagement die Grundlagen einer Utopie der digitalen Wissenschaftslandschaft dar. Im Zuge der hochdynamischen Entwicklung der Thematik, suchen Infrastruktureinrichtungen ihren Platz in diesem Idealszenario, ebenso wie wissenschaftliche Akteure dem hohen Leidensdruck einer explodierenden Anzahl von Forschungsdaten und -methoden begegnen.
Doch wie sieht ein Idealszenario Forschungsdatenmanagement aus und wie sortieren sich aktuelle Projekte und Aktivitäten darin ein? Wird die FDM-Utopie aus der sinnvollen Verzahnung der Nationalen Forschungsdateninfrastruktur und der European Open Science Cloud entstehen?
Anhand der Vorstellung ausgewählter FDM-Akteure und -Projekte auf deutscher und europäischer Ebene wird versucht ein Bild der aktuellen und zukünftigen FDM-Landschaft zu zeichnen.

1. Einführung

In vielen Diskussionen wird schnell von "dem einen" Forschungsdatenmanagement (FDM) gesprochen. Als wenn "das" Forschungsdatenmanagement eine Software oder Hardware wäre, die man kauft, aufsetzt, ausrollt und (ein bisschen) pflegt. Aus Infrastruktursicht kennt man solche "Dinge". Sei es ein gut aufgesetzter Publikationsserver, das Einführen oder die Verbesserung der Funktionalitäten einer Campus-App, sei es ein Linksolver um zum Volltext von Journalartikeln zu gelangen, Semesterapparate, die online in die eLearning-Plattform der eigenen Hochschule integriert sind, Verlinkung mehrerer Daten und Funktionen in einer einzigen Hochschul-Card, etc. Man kann es als neuen, schicken Service verkaufen um sich

damit bei der Kundschaft unabdingbar machen. Aus Sicht der potentiellen Kundschaft (im Kontext FDM die Forschenden) wird zum einen -und im günstigsten Fall- im FDM "die Lösung all meiner Probleme" gesehen und zum anderen -meist häufigerem Fall- eher das neue administrative Add-On, dem man sich unterwerfen muss, was man vielleicht nicht verhindern, womit man aber seiner Arbeit auch ein Label geben kann ("ich nutze ein Tool und erfülle damit Vorgaben").

Dieses "eine" Forschungsdatenmanagement existiert nicht, was hinsichtlich der Komplexität des Themas nicht verwunderlich ist. Während sich die einen der vereinfachten Vorstellung der digitalen Speicherung und Erschließung von Information hingeben, bedeutet FDM für andere Entdeckung, Umsetzung und Hinterfragung neuer Methoden und Arbeitsarten in der Forschung selbst. Das Ganze am liebsten noch standardisiert und über Disziplingrenzen hinweg. Dass es hierfür nicht "eine" Lösung gibt, liegt auf der Hand, macht das Thema aber auch so komplex und reizvoll.

Dieses "eine" FDM beinhaltet also eine Vielzahl an Vorstellungen und Erwartungen, verbunden mit einer ebenso großen Zahl an gewünschten Lösungen (Tools und Dienste einerseits, Arbeits- und Forschungsmethoden andererseits), die ich im Folgenden gerne unter dem Begriff FDM-Utopie bezeichnen möchte.

2. Die digitale Wissenschaftslandschaft

Bereits 2009 wurde der wohlbekannte Ausdruck "Daten sind das neue Öl" von der EU-Politikerin Kuneva geprägt (Kuneva 2009). Der damalige Fokus des Ausdrucks bezog sich auf wirtschaftliche und datenschutzrechtliche Aspekte für die Gesellschaft. Dennoch ist die Formulierung treffend für die Transformation, die wir momentan durch die Digitalisierung in vielen Lebensbereichen erfahren.

Auch die Wissenschaftslandschaft verändert sich massiv durch die Digitalisierung, aber nicht nur durch das bloße Vorhandensein von digitalen (Forschungs-)Daten. Es gehört weit mehr dazu, um von einer digitalen Wissenschaftslandschaft zu sprechen, die sich entsprechend von der Wissenschaftslandschaft von vor 20 Jahren unterscheidet.

Zum einen müssen für die digitale Wissenschaftslandschaft die Kommunikationsformen bzw. den Austausch von Informationen innerhalb der Wissenschaft selbst betrachtet werden, und zum anderen die genuinen Forschungsmethoden. Diese Zweiteilung entspricht der groben Unterscheidung

zwischen Aufgaben und Themen der Infrastruktur für die Wissenschaft einerseits und Aufgaben und Themen der Fachdisziplinen andererseits.

In Bezug auf den Austausch von Informationen unterscheidet sich die Wissenschaft heute kaum von der Wissenschaft von vor 20 Jahren. Zwar nutzen Forschende aller Fachgebiete digitale Methoden zur Kommunikation, dennoch spielt in der Wissenschaftskommunikation die Veröffentlichung von Artikeln in Fachzeitschriften weiterhin eine essentielle Rolle. Die Medien sind anders geworden (digital), die Kommunikationsmethoden bleiben aber unverändert. Während das Auftauchen wissenschaftlicher Fachzeitschriften 1665 eine regelrechte Revolution für die Wissenschaftslandschaft darstellte, ist der zunehmende digitale Zugang zu Fachartikeln höchstens bequem. Er ermöglicht zwar einen schnelleren Informationsaustausch, die Explosion in der Anzahl neuer Zeitschriften und Artikeln erschwert zugleich aber den Überblick. Dennoch bleiben die wiss. Fachzeitschriften ein Basisinstrument des etablierten Wissenschaftssystems, da das eigene wissenschaftliche Renommee von ihnen abhängt. Egal wie sehr sich die Geister um quantitative und qualitative Messungen von Zitierungen scheiden und trotz des Auftauchens von Altmetrics und der wissenschaftlichen Kommunikation in sozialen Medien, bleibt nach wie vor ein Standard wissenschaftlicher Kommunikation das häufige Publizieren in viel zitierten Fachzeitschriften. Wenn für die Forschenden die Publikation in Fachzeitschriften Standard bleibt, bleibt in Bezug auf die digitalen Möglichkeiten aber zu hinterfragen, warum sich die Publikationsstandards der Verlage trotz der digitalen Möglichkeiten nicht wesentlich oder nur sehr langsam verändern. Nur einzelne Zeitschriften vernetzen die Texte mit den zugrundeliegenden Daten, dies aber oft in einer sehr rudimentären Form. Technisch möglich und sehr wünschenswert wäre es doch z.B., dass beim Lesen eines Artikels auf ein Bild geklickt werden könnte, was direkt eine (bearbeitbare) Datei wiedergäbe, in der die Rohdaten enthalten sind, die die Grundlage des Bildes sind. Der interessierte Leser könnte damit direkt das postulierte Ergebnis nachvollziehen bzw. weitere Aspekte erforschen und erarbeiten[1]. Aktuell sind Daten nur wenig verknüpft und wenn sie es sind, ist der Aufwand des Lesers enorm, um die Daten direkt nutzen zu können. Forschungsdaten werden von Verlagen wie Text behandelt und die entsprechenden etablierten Standards hierfür

[1] Hierzu gab es einen Vorstoß vom Verlag Elsevier, Informationen dazu finden sich auf https://www.elsevier.com/connect/the-article-of-the-future (geprüft am 27.5.2019), weiterführende Links scheint es aber nicht zu geben.

genutzt. Dabei gibt es bereits Initiativen, wie die Data Citation Working Group der Research Data Alliance[2], die neue und insbesondere nicht-statische Zitiermöglichkeiten aufzeigt.

Die digitale Wissenschaftslandschaft ist also in Bezug auf den Austausch von Informationen noch lernfähig bzw. unterscheidet sich noch nicht viel von der Wissenschaftslandschaft von vor der Digitalisierung. Anders sieht es in den Forschungsmethoden aus.

Was die Forschungsmethoden angeht, gibt es nur noch wenig Forschende die ausschließlich mit "Papier und Bleistift" forschen. Dieses Privileg bleibt einigen wenigen Forschungsgebieten vorbehalten, die nicht mehr als Ihren Kopf und eine strukturierte Ablageform benötigen. So ist und bleibt es beispielsweise in der reinen Mathematik, in Bereichen der Philosophie und in Bereichen der Rechtswissenschaft. In den meisten Wissenschaftsgebieten haben Rechnerkapazitäten sowie das Vorhandensein und die Nutzungsmöglichkeit von Rohdaten zu neuen Arbeits- aber insbesondere auch Forschungsmethoden geführt. Durch den Einsatz von Machine Learning können Muster in Datenmengen erkannt werden, die mit herkömmlichen wissenschaftlichen Methoden nicht identifizierbar sind. Zum Beispiel in den Sozialwissenschaften hat das Vorhandensein von massenhaften personenbezogenen Daten aus sozialen Medien und aus der Medizin oder Biologie dazu geführt, dass neue gesellschaftliche Muster erkannt und gänzlich neue gesellschaftswissenschaftliche Fragen überhaupt gestellt und erforscht werden konnten. So knüpft King an das Bild der Daten als Öl der Zukunft an und malt eine datenreiche Zukunft für die (gesellschaftswissenschaftliche) Forschung aus, die sie sich aber erst erarbeiten muss (King 2011). Das sog. 4. Paradigma der Wissenschaft wurde bereits 2009 formuliert (Hey 2009), die Wissenschaft muss lernen, mit datengetriebenen Methoden umzugehen insbesondere in Bezug auf wissenschaftliche Anerkennung aber auch im Umgang und Definition von Vertrauenswürdigkeit. Das sind Aspekte die weit weg von reinen Infrastrukturaspekten sind wie z.B. dem bloßen Ablegen und Speichern von Daten.

[2] https://rd-alliance.org/groups/data-citation-wg.html (geprüft am 25.5.2019)

3. Die Nationale Forschungsdateninfrastruktur (NFDI)

Die digitale Wissenschaftslandschaft verändert sich also kontinuierlich im Rhythmus der fortschreitenden Digitalisierung. Auf verschiedenen Ebenen spielen für die Veränderung das Management der Forschungsprimärdaten eine wesentliche Rolle. Die FDM-Utopie ist somit eine Voraussetzung für die Entwicklung der Wissenschaftslandschaft.

Nicht umsonst hat sich der Rat für Informationsinfrastrukturen in seiner ersten Publikation zur Zukunft der digitalen Informationsinfrastrukturen für das Wissenschaftssystem dem Thema Forschungsdaten gewidmet (RfII 2016). Der 2014 von der Gemeinsamen Wissenschaftskonferenz einberufene Rat agiert auf systemischer Ebene und hat den Begriff der Nationalen Forschungsdateninfrastruktur (NFDI) geprägt.

Mit der Nationalen Forschungsdateninfrastruktur wagt Deutschland ein großes wissenschaftspolitisches Experiment. Dafür werden bis 2028 bis zu 90 Millionen Euro jährlich für den Aufbau und Betrieb der NFDI von Bund und Ländern bereitgestellt (GWK 2018). Die NFDI soll ein Gesamtkonstrukt sein, aufbauend auf mehreren Konsortien, die unterschiedliche fachliche Domänen repräsentieren (RfII 2018). Anders als bekannte wissenschaftliche Förderinstrumente bisher funktionieren, soll der Aufbau der NFDI nicht kompetitiv erfolgen. Dies erfordert neue Denkweisen. Nicht nur ist die Durchführung eines nicht-kompetitiven Verfahrens formal gesehen eine große Herausforderung, auch sind die Antragsteller seit Jahrzehnten systematisch darauf trainiert, sich im Wettbewerb um Fördermittel gegen andere zu positionieren. Nun sollen sich Antragsteller als Teil eines Ganzen verstehen, die Begutachtung soll international erfolgen. Um die Grundlagen für ein solches Vorgehen zu ermöglichen, wird in jeder der drei Förderrunden eine ausführliche Initialisierungsphase stattfinden, die u.a. mit der sog. NFDI-Konferenz den Antragstellenden die Möglichkeit gibt, sich mit anderen Konsortien abzustimmen und sich zu einem Teil des Ganzen zusammenzufinden. Die erste NFDI-Konferenz im Mai 2019 hat ein sehr heterogenes Bild von 57 eingereichten Interessensbekundungen für die erste Ausschreibungsrunde gezeigt (Eickhoff 2019). Darunter befinden sich allein 10 Aktivitäten, die sich selbst laut DFG-Fachsystematik der Medizin zuordnen, aber auch insgesamt 23 Einreichungen die entweder Multidisziplinär sind bzw. sich auf einen übergreifendes (crosscutting) Thema richten. Im Idealfall sollen pro Antragsrunde maximal 10 Konsortien gefördert

werden, so dass am Ende die NFDI aus maximal 30 Konsortien besteht. Auf der stattgefundenen ersten NFDI-Konferenz waren fast alle Fachgruppen laut DFG-Systematik vertreten. Allein das zeigt die Relevanz des Vorhabens bzw. die gute Vorarbeit in Bezug auf die Bekanntmachung des Prozesses. Dies alles vor dem Hintergrund, dass es bis zum jetzigen Zeitpunkt noch keine Ausschreibung gibt. Die Energie, die im Vorfeld besteht ist immens und scheint nicht abzuebben.
Die NFDI soll als Prozess verstanden werden. Es soll über Jahre hinweg ein komplexes System aufgebaut werden, in dem aus unterschiedlichen Konsortien Synergien entstehen, so dass der Zugriff auf und die Nutzung von Daten unterschiedlicher Domänen ermöglicht werden soll. Dafür bedarf es Standards, Infrastrukturen, Services und Kompetenzen (bzw. deren Aufbau). Insbesondere wird es dafür sehr viel Kommunikation brauchen.

4. Die European Open Science Cloud

In Deutschland wird nicht von einer Open NFDI gesprochen. Das Wort Open wird hier sehr bewusst vermieden bzw. nicht als Fahnenträger genutzt. Anders verhält es sich im europäischen Kontext.
Die European Open Science Cloud (EOSC) ist ein weitaus komplexerer Prozess als es die NFDI ist. Das Thema Forschungsdatenmanagement ist innerhalb der EU in einem größeren Kontext zu verstehen. Das EU Framework Programme for Research and Innovation "Horizon 2020" enthält ein Arbeitsprogramm "Europäische Forschungsinfrastrukturen (einschließlich e-Infrastrukturen)". Mit diesem Arbeitsprogramm sollen zum einen bereits bestehende Forschungsinfrastrukturen vernetzt werden und zum anderen neue Forschungsinfrastrukturen von gesamteuropäischem Interesse im Bereich Wissenschaft und Technik geschaffen werden (BMBF 2019).
In diesem Kontext spielt die ESFRI Roadmap eine wegweisende Rolle. Das ESFRI (European Strategy Forum on Research Infrastructures) hat 2018 in seiner Roadmap 18 sog. ESFRI-Projekte benannt, sowie eine Liste von 37 Landmarks. ESFRI Projekte sind ausgewählte Forschungsinfrastrukturen, die Exzellenz in ihrer wissenschaftlichen Domäne und einen hohen Reifegrad aufzeigen und von Bedeutung für die europäische Weiterentwicklung der Forschungsinfrastrukturen sind. Die ESFRI Roadmap enthält Forschungsinfrastrukturen aller Fachdomänen und ist gegliedert in die 5 Bereiche

Energie, Umwelt, Gesundheit und Ernährung, Physik und Ingenieurwissenschaften, sowie Soziale und Kulturelle Innovation. In letzterem Bereich werden zusätzlich zwei Themen mit hohem strategischem Potential für die zukünftige Entwicklung neuer Forschungsinfrastrukturen genannt: Religionswissenschaft sowie Digitale Dienste für offene (Gesellschafts-)Wissenschaften. ESFRI-Landmarks sind ESFRI-Projekte, sie sich bereits in der Implementierungsphase befinden (ESFRI 2018).
Eine der Fördermaßnahmen innerhalb des H2020-Rahmenprogramms "Europäische Forschungsinfrastrukturen (einschließlich e-Infrastrukturen)" ist "e-Infrastrukturen und EOSC" wo es konkret um die Themen High-Performance-Computing (HPC), offene Forschungsdaten und schnelle Datenverbindungen geht. Hierbei geht es um die Neuentwicklung aber insbesondere die Zusammenführung fragmentierter Ansätze.

Die European Open Science Cloud beschreibt sich selbst als:
" The EOSC will offer 1.7 million European researchers and 70 million professionals in science, technology, the humanities and social sciences a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines by federating existing scientific data infrastructures, currently dispersed across disciplines and the EU Member States" (EOSC 2019).

Durch H2020 in der entsprechenden Fördermaßnahme "e-Infrastrukturen und EOSC" werden dafür aktuell insgesamt 14 Projekte gefördert, die die Basis für den Aufbau der EOSC bilden. Diese Projekte sind unterschiedlich ausgerichtet. Zum Teil sind sie disziplinär, dann eher mit dem Fokus entsprechende ESFRI-Projekte/Landmarks mit EOSC-Diensten zu verbinden (ENVRI-FAIR, EOSC-Life, ESCAPE, PANOSC, SSHOC), zum Teil generisch bzw. mit Fokus auf Infrastruktur und generischen Services und/oder administrativ/unterstützend (eInfra-Central, EOSC Secretariat.eu, EOSC-hub, EOSCpilot, FAIR is FAIR, FREYA, OCRE, Open-AIRE Advance, RDA Europe 4.0) (s. Tab. 1).

Tabelle 1 Übersicht der EOSC-Projekte, Quelle (CORDIS 2019) und (EOSC 2019), Stand 24.05.2019.

Projekt	Projektwebseite	Förderzeitraum	Fördersumme
eInfra Central	http://einfracentral.eu/	01.2017 - 06.2019	1.499.037,50 €
ENVRI-FAIR	http://envri.eu/envri-fair/	01.2019 - 12.2022	18.997.878,75 €
EOSC Secretariat.eu	http://www.eoscsecretariat.eu	01.2019 - 06.2021	9.996.500,00 €
EOSC-hub	https://eosc-hub.eu/	01.2018 - 12.2020	30.000.000,00 €
EOSC-Life		03.2019 - 02.2023	23.745.996,25 €
EOSCpilot	https://eoscpilot.eu/	01.2017 - 04.2019	9.953.067,50 €
ESCAPE	https://www.projectescape.eu/	02.2019 - 07.2022	15.983.301,25 €
FAIR is FAIR	https://www.fairsfair.eu/	01.2019 - 12.2021	k.A.
FREYA	http://www.project-freya.eu	01.2017 - 12.2020	4.998.650,00 €
OCRE		01.2019 - 11.2021	12.000.000,00 €
OpenAIRE-Advance	https://www.openaire.eu/	01.2018 - 12.2020	9.999.997,50 €
PANOSC	http://panosc.eu	12.2018 - 11.2022	11.953.516,99 €
RDA Europe 4.0	https://www.rd-alliance.org/	03.2018 - 05.2020	3.500.000,50 €
SSHOC	https://sshopencloud.eu/	01.2019 - 04.2022	14.455.594,08 €

Interessant an dem Prozess zum Aufbau der EOSC und der NFDI wird sein, wie sich beide Prozesse integrieren werden. Der Prozess der EOSC ist keine 1-zu-1 Blaupause für die NFDI. Die NFDI selbst wird auch nicht die Basis für die EOSC sein. Vielmehr werden sich beide Prozesse über unterschiedliche Ebenen divers, quer, diagonal und sehr stark vernetzen und gegeneinander befruchten. Erfolgreiche NFDI-Konsortien werden entsprechend nachweisen, dass sie in Prozesse im EU-Kontext involviert sind (ESFRI-Landmarks oder -Projekte, EOSC-Projekte, etc.) oder mit diesen Kooperationen entwickeln.

In Bezug auf die Größenordnung und die Kosten für das Forschungsdatenmanagement hat bereits 2016 die EOSC High Level Expert Group der Europäischen Kommission empfohlen, dass 5% der Forschungsausgaben in "properly managing and stewarding data" ausgegeben werden sollte (EU; EOSC-HLEG 2016).

Wenn man sich für Deutschland die Ausgaben für Forschung und Entwicklung beschränkt auf Hochschulen anschaut, ergeben sich 15,3 Milliarden Euro Gesamtausgaben im Jahr 2015, davon 7,9 Mrd. für Grundmittelforschung und 7,4 Milliarden Euro für Drittmittelforschung (Destatis 2018). Würde man davon ausgehen, dass in einer FDM-Utopie der Zukunft 5% der Ausgaben für reines FDM eingesetzt werden müssten, wären das (in Relation zu den Forschungsausgaben der Hochschulen in 2015) 756 Millionen Euro. Im Vergleich dazu sind die geplanten Ausgaben für die NFDI (90 Millionen Euro pro Jahr) gerade mal 0,58%, also etwas mehr als ein Zehntel des Vorgeschlagenen (beschränkt auf Hochschulen). Natürlich wird die NFDI bzw. das Thema FDM nicht auf Hochschulen zu reduzieren sein.

Die FDM-Utopie -will man der HLEG glauben- wird also (für Deutschland) noch mehr als nur von der NFDI "unterfüttert" werden müssen.

5. Bestandteile der FDM-Utopie

Während die NFDI und die EOSC zwei wissenschaftspolitische Prozesse sind, die die Zukunft der digitalen Wissenschaftslandschaft massiv prägen, gibt es darüber hinaus eine Vielzahl von Aktivitäten im Bereich der Entwicklung von Forschungsdatenmanagement. Da das Thema FDM und die zugehörigen Aktivitäten aber so divers sind, ist eine reine Auflistung, Sortierung und Wertung konkreter Aktivitäten nicht möglich, da diese untereinander auch nicht vergleichbar sind. Stattdessen soll nachfolgend eine grobe Orientierung gegeben werden, in welchen Zweigen der Wissenschaftslandschaft welche Aspekte aktuell erarbeitet werden.

Für die Skizzierung und Einsortierung von FDM-Aktivitäten können vier Aktionsfelder identifiziert werden, die abseits der Prozesse rund um die EOSC und NFDI bestehen:

1. Politische Förderung durch explizite Förderlinien zu FDM

Über das Bundesministerium für Bildung und Forschung (BMBF) wurden ab 2016 zwei Förderlinien mit dem expliziten Fokus auf Forschungsdatenmanagement veröffentlicht (BMBF 2016; BMBF 2018). Noch konkreter eröffnet die Deutsche Forschungsgemeinschaft (DFG) die Möglichkeit über Sonderforschungsbereiche (SFB) Teilprojekte Informationsinfrastruktur (sog. INF-Projekte) zu fördern. Mit den INF-Projekten können Aspekte des FDM innerhalb eines SFBs aufgegriffen und erarbeitet werden. Aktuell werden 27 INF-Projekte auf der Wiki-Seite von forschungsdaten.org gelistet[3]. Von den insgesamt 277 aktuell geförderten SFBs in Deutschland waren zuletzt 32 auf einem Workshop in Göttingen mit Fokus INF-Projekte vertreten (eRA 2018). Über die nationalen Förderinstrumente hinaus ist das bereits erwähnte H2020-Förderprogramm für die Förderung expliziter FDM-Vorhaben relevant.

[3] https://www.forschungsdaten.org/index.php/INF_Projekte#SFB_1194 (geprüft am 23.05.2019)

2. Disziplinspezifische und methodische Vorhaben

Fachcommunities, die schon vor Jahren für sich einen Mehrwert im Management von Forschungsdaten erkannt haben, zeichnen sich durch eine Vielzahl an Angeboten und Projekten ab, die darauf zielen, den Austausch und die Sicherung von Daten zu ermöglichen, sowie Tools und Dienste für die Community anzubieten. Bekannte und internationale Projekte treten in der Regel aus den wissenschaftlichen Communities selbst hervor. Oft entwickeln sich aus ursprünglich drittmittelgeförderten Projekten entsprechende Angebote. Auch das BMBF und die DFG fördern seit vielen Jahren Initiativen mit dem Ziel Forschungsinfrastrukturen aufzubauen. Ein Beispiel in den Geisteswissenschaften ist dabei Dariah[4] und in der Biologie GFBio[5]. Darüber hinaus gibt es im nationalen, europäischen und internationalen Kontext eine Vielzahl an Vorhaben. Eine konsistente Aufzählung ist aber nicht möglich, da die Vorhaben oft Entwicklungen im Forschungsdatenmanagement zwar vorantreiben, und damit zur FDM-Utopie beitragen, aber in einem übergeordneten und disziplinspezifischen Zusammenhang stehen.

3. Generische Ansätze, Unterstützung durch Infrastruktureinrichtungen

Abseits politisch geförderter und disziplinspezifischer Vorhaben, gibt es eine Reihe von Aktivitäten, die sich dadurch auszeichnen, dass sie sich auf generische Aspekte des Themas FDM fokussieren. Dabei geht es um die Entwicklung von allgemeinen Tools, Software und Diensten. Gerne wird hier das Label "Entwicklung von Diensten für Long-Tail-Daten" benutzt. Es geht dabei um Themen wie die Ablage von Daten (Sync+Share, Speicherung), Repositorien, Tools zum Datenhandling, Archivierung von Daten, etc. Die Vielzahl der Akteure und Aktivitäten ist hierbei sehr groß. Sie reichen von einzelnen Infrastruktureinrichtungen (Bibliotheken, Rechenzentren an Hochschulen, außeruniversitären Forschungseinrichtungen aber auch innerhalb mittelgroßer Forschungsgruppen), bis hin zu größeren Konsortien aus verschiedenen Partnern. Zum Teil werden die Aktivitäten von den Infrastrukturakteuren selbst getragen, zum Teil handelt es sich auch um drittmittelgeförderte Projekte, die aber keinen disziplinspezifischen Fokus haben.

4. Flankierende Initiativen (Bottom-up-Aktivitäten, ggf. mit politischer Förderung)

[4] https://de.dariah.eu/ (geprüft am 25.5.2019)
[5] https://www.gfbio.org/ (geprüft am 25.5.2019)

Über die bereits genannten hinaus gibt es noch eine Reihe an Aktivitäten, bei denen es sich weder um Förderlinien, noch um den fachspezifischen oder generischen Aufbau von Forschungsinfrastrukturen oder Services handelt. Gemeint sind Initiativen die den ganzen FDM-Prozess begleiten und koordinierend und kommunikativ unterstützen. Hierzu zählen Initiativen wie RDA (Deutschland)[6], DINI/nestor-AG Forschungsdaten[7], GOFAIR[8] und einzelne Länderaktivitäten (s.a. Grasse 2018). Natürlich spielen sich innerhalb von RDA sowohl methodische als auch generische Themen ab, ebenso verfolgen einige Landesinitiativen das Ziel generische (für ein Land zugängliche) Services aufzubauen. Gemeinsam haben diese Initiativen aber, dass sie originär weder aus der Wissenschaft, noch aus der Infrastruktur stammen, sondern im besten Fall unterschiedliche Gruppierungen durch Vernetzung und gemeinsames Erarbeiten von Themen zusammenbringen. Damit stellen sie aus systemischer Sicht eine unterstützende Rolle des Gesamtprozesses dar.

6. Auf dem Weg zur FDM-Utopie

Die vier vorgestellten Aktionsfelder im Kontext FDM werden über die nächsten Jahre bestehen bleiben. Selbst wenn in zwei bis drei Jahren erste NFDI-Konsortien gefördert werden, erste EOSC-Projekte abgeschlossen und ihre Dienste über das EOSC-Portal verfügbar sind, wird die Wissenschaftslandschaft weiterhin weit weg von der FDM-Utopie sein. In drei Jahren werden viele Forschende weiterhin Daten auf USB-Sticks speichern und Datenmanagementpläne bei Projektbeantragung kurz und knapp formulieren und sie danach vergessen. Wir befinden uns weiterhin mitten in einer Trial&Error-Phase. Diese Phase wird -weil das Thema so vielschichtig, komplex, heterogen und an den Grundfesten des Systems rüttelnd ist- noch mindestens fünf Jahre andauern. In dieser Zeit werden sich Rollen ausprobieren, eine Standardisierung von Kompetenzen klären, Methoden, Bedarfe und sinnvolle Hilfsmittel herausstellen. Wir müssen diesem Prozess Zeit lassen und in Innovationen investieren und hierbei auf einen möglichst hohen Vernetzungsgrad achten.

Für einzelne Akteure (WissenschaftlerInnen oder Infrastrukturakteure) besteht die größte Schwierigkeit aktuell darin, einen Überblick zu behalten. Für die strategischen

[6] https://www.rda-deutschland.de/ (geprüft am 25.5.2019)
[7] https://dini.de/ag/dininestor-ag-forschungsdaten/ (geprüft am 25.5.2019)
[8] https://www.zbw.eu/de/ueber-uns/arbeitsschwerpunkte/forschungsdatenmanagement/go-fair/ (geprüft am 25.5.2019)

Entscheidungsgremien wiederum besteht die größte Schwierigkeit darin abzuschätzen, wieviel sie aktuell selbst in Innovationen investieren und wieviel Innovationen sie durch andere Akteure zulassen. Der Prozess ist insgesamt größer und länger als bisher angenommen.
Für Bibliotheken besteht die größte Herausforderung darin, sich vor allem von der Vorstellung von Silos und Repositorien als Lösungen für FDM zu verabschieden. Infrastrukturanbieter sind gefragt, um im Kontext des wissenschaftlichen Austausches Standards zu entwickeln: Von der Zitierung, zum Austausch bis hin zur Verlinkung von Informationen. Diese Standards können aber nicht wie bisher heißen, Daten "in Stein zu meißeln" und -wie Text- rein zitierbar zu machen. Die Ablage und Speicherung von Daten ist keine Herausforderung. Das kann schon jetzt in bereits vorhandenen Repositorien passieren, Neuentwicklungen müssen nicht mehr in der Breite stattfinden. Woran Infrastrukturanbieter arbeiten müssen, ist an der Flexibilität der Methoden, die Daten "lebendig" zu halten. Hier wird es in den kommenden Jahren viel Entwicklung geben. Diese wird aber ganz im Kontext von Trial&Error singulär passieren und in enger Verzahnung mit wissenschaftlichen Communities.

Eine weitere Herausforderung für die Infrastrukturanbieter ist die Klärung der Rollenbilder. Welche Aufgaben wird Personal an Infrastruktureinrichtungen in der FDM-Utopie übernehmen? Wo wird dieses Personal angesiedelt sein? In einer Übergangsphase braucht es flexible Lösungen, in denen Personen losgelöst von alten zum Teil starren Strukturen agieren können. Mittelfristig gilt es einen Platz in der digitalen Wissenschaftslandschaft für solche Data-Stewards zu finden. Die aktuellen Strukturen sind dafür zu wenig flexibel. In diesem Zusammenhang ist ein Modellprojekt an der TU Delft erwähnenswert, in der Data-Stewards für jede Fakultät eingesetzt werden um insbesondere Rollen, Aufgaben und Bedarfe im Zusammenspiel zwischen zentraler Infrastruktur (Bibliothek) und Fachcommunities (Fakultäten) modellhaft zu klären (Teperek 2018).

Insgesamt wird sich die FDM-Utopie vermehrt mit dem Aspekt der Interaktion mit der Privatwirtschaft auseinandersetzen. Es wird interessant sein, in wie fern Aktivitäten aus der Privatwirtschaft in den Aufbau der digitalen Wissenschaftslandschaft integrierbar sein werden - oder nicht. Bereits jetzt werben Produkte von Innoplexus,

Palantir oder Watson[9] damit, das Datenchaos mit Hilfe von künstlicher Intelligenz zu bändigen. Jegliche Arten von strukturierten und unstrukturierten Daten sollen zusammengebracht werden, ein Datenmodell entwickelt und damit den Nutzern die Möglichkeit eröffnet werden, neue Informationen aus ihren Daten zu erhalten. Der Kern liegt auch hier in der Entwicklung eines Datenmodells. Ein (einziges) Datenmodell für die Wissenschaft scheint an der Stelle ähnlich schwierig, wie der damalige Vorstoß der Universalbibliothek von Borges, die in der Lage sei, die Informationen der gesamten Welt zu speichern. Die Idee eines einzigen Datenmodells für die Wissenschaft ist daher stark mit dem Begriff einer FDM-Utopie verbunden. Ob wir jemals dahin kommen, ist fraglich. Wenn die Wissenschaft es schafft, möglichst übergreifende Datenmodelle zu etablieren, wäre sicherlich schon viel erreicht. In der Entwicklung von Datenmodellen und Etablierung von Standards liegt die Zukunft des Forschungsdatenmanagements.

Referenzen

Bundesministerium für Bildung und Forschung (BMBF); EU-Büro des BMBF (2019): ESFRI - Europäisches Strategieforum für Forschungsinfrastrukturen. https://www.eubuero.de/infra-esfri.htm, geprüft am 24.05.2019.

Bundesministerium für Bildung und Forschung (BMBF) (2016): Förderrichtlinie zur Erforschung des Managements von Forschungsdaten in ihrem Lebenszyklus an Hochschulen und außeruniversitären Forschungseinrichtungen. Bundesanzeiger vom 19.08.2016. https://www.bmbf.de/foerderungen/bekanntmachung-1233.html, geprüft am 25.05.2019.

Bundesministerium für Bildung und Forschung (BMBF) (2018): Richtlinie zur Förderung von Forschungsvorhaben zur Entwicklung und Erprobung von Kurationskriterien und Qualitätsstandards von Forschungsdaten im Zuge des digitalen Wandels im deutschen Wissenschaftssystem. Bundesanzeiger vom 13.06.2018. https://www.bmbf.de/foerderungen/bekanntmachung-1791.html, geprüft am 25.05.2019.

Eickhoff, Ulrike (2019): Overview of Extended Abstracts. NFDI-Conference, Bonn, 13-14 May 2019. https://www.dfg.de/download/pdf/foerderung/programme/nfdi/welcome_speech_dfg_head_office_eickhoff.pdf, geprüft am 23.05.2019.

European Open Science Cloud (EOSC) (2019): EOSC Portal. https://www.eosc-portal.eu/, geprüft am 23.05.2019.

[9] S.a. https://www.innoplexus.com/, https://www.palantir.com, https://dataplatform.cloud.ibm.com/ (geprüft am 24.5.2019)

eResearch Alliance (eRA) (2018): Workshop: Forschungsdatenmanagement und -infrastruktur in DFG-Sonderforschungsbereichen. Göttingen. http://www.eresearch.uni-goettingen.de/de/content/workshop-forschungsdatenmanagement-und-infrastruktur-dfg-sonderforschungsbereichen, geprüft am 25.05.2019.

European Commission; CORDIS (2019): Community Research and Development Information Service (CORDIS). EU Research Results. https://cordis.europa.eu/, geprüft am 24.05.2019.

European Strategy Forum on Research Infrastructures (ESFRI) (2018): Roadmap 2018. Strategy Report on Research Infrastructures. http://roadmap2018.esfri.eu/media/1066/esfri-roadmap-2018.pdf, geprüft am 24.05.2019.

European Union (EU); EOSC High Level Expert Group (EOSC-HLEG) (2016): Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud. https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf#view=fit&pagemode=none, geprüft am 24.05.2019.

Gemeinsame Wissenschaftskonferenz (GWK) (2018): Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018. https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf, geprüft am 25.05.2019.

Grasse, M.; López, A.; Winter, N. (2018): Landesinitiative NFDI – a Central Point of Contact for RDM for Higher Education Institutions in the German State of North Rhine-Westphalia. In: *Data Science Journal* (17), S. 25. DOI: 10.5334/dsj-2018-025.

Hey, Tony; Tansley, Stewart; Tolle, Kristin (2009): The Fourth Paradigm: Data-Intensive Scientific Discovery: Microsoft Research. https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/, geprüft am 24.05.2019.

King, Gary (2011): Ensuring the Data-Rich Future of the Social Sciences. In: *Science* 331 (6018), S. 719. DOI: 10.1126/science.1197872.

Kuneva, Meglena (2009): Keynote Speech. Roundtable on Online Data Collection, Targeting and Profiling, Brüssel, 31.03.2009. http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm, geprüft am 24.05.2019.

Rat für Informationsinfrastrukturen (RfII) (2016): Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Göttingen. https://d-nb.info/1104292440/34, geprüft am 25.05.2019.

Rat für Informationsinfrastrukturen (RfII) (2018): In der Breite und forschungsnah: Handlungsfähige Konsortien. Dritter Diskussionsimpuls zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) für die Wissenschaft in Deutschland. Göttingen. https://d-nb.info/1172854858/34, geprüft am 24.05.2019.

Statistisches Bundesamt (Destatis) (2018): Bildungsfinanzbericht 2018. Im Auftrag des Bundesministeriums für Bildung und Forschung und der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Wiesbaden. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsfinanzen-Ausbildungsfoerderung/Publikationen/Downloads-

Bildungsfinanzen/bildungsfinanzbericht-1023206187004.pdf?__blob=publicationFile&v=4, geprüft am 24.05.2019.

Teperek, Marta; Cruz, Maria J.; Verbakel, Ellen; Böhmer, Jasmin; Dunning, Alastair (2018): Data Stewardship: Adressing Disciplinary Data Management Needs. In: International Journal of Digital Curation 13 (1), S. 141–149. DOI: 10.2218/ijdc.v13i1.604.

Lessons Learned

Die Kuratierung sozialwissenschaftlicher Forschungsdaten – Praxisfragen und Beispiellösungen

Patrick Droß, Julian Naujoks

Wissenschaftszentrum Berlin für Sozialforschung (WZB)

Zusammenfassung

Der Beitrag bietet einen Blick in die Praxis der Veröffentlichung sozialwissenschaftlicher Forschungsdaten. Dabei beschreibt er den Arbeitsschritt der Datenkuratierung als wesentliche Komponente, um Forschungsdaten nachhaltig verfügbar zu machen. Anhand der Erfahrungen des institutionellen Forschungsdatenmanagements am Wissenschaftszentrum Berlin für Sozialforschung (WZB) werden einzelne Problemfelder und Praxislösungen dargestellt, die im Laufe des Kuratierungsprozesses auftreten. Dabei sich zeigt, dass die Datenkuratierung ein komplexer Vorgang ist, bei dem sich fachspezifische, aber auch (arbeits-)organisatorische Herausforderungen stellen. Diese werden anhand von drei exemplarischen Themenkomplexen skizziert: Datenschutzfragen müssen im Forschungsprozess frühzeitig mitgedacht werden. Dazu gehört, dass in informierten Einwilligungen die Rechte der Befragten geschützt, gleichzeitig aber spätere Nachnutzungsmöglichkeiten nicht durch zu restriktive Formulierungen ausgeschlossen werden. In Bezug auf die Verwertungs- und Nutzungsrechte muss gerade in Projektkonstellationen mit mehreren Beteiligten frühzeitig eine Abstimmung über die Datenveröffentlichung erfolgen. Schließlich verdeutlicht die Verknüpfung von Daten- und Textpublikation, wie sich Veröffentlichungsroutinen und Suchgewohnheiten der Wissenschaftler*innen auf die Auffindbarkeit der Forschungsdaten auswirken.

Abstract

This article offers an insight into the practice of publishing social science research data. It describes the step of data curation as an essential component to make research data sustainably available. Based on the experiences of institutional research data management at the WZB Berlin Social Science Center, particular problem areas and practical solutions that arise during the curation process are presented. This shows that data curation is a complex process in which both subject-specific and (work) organisational challenges arise. These challenges are outlined using three exemplary

topics: Data protection issues must be considered at an early stage in the research process. This includes protecting the rights of the respondents in informed consents, but at the same time not excluding the possibility of subsequent use by overly restrictive wording. With regard to the rights of exploitation and use, especially in project constellations with several participants, the publication of data must be coordinated at an early stage. Finally, the linking of data and text publication illustrates how publication routines and search habits of scientists affect the findability of research data.

1. *Einleitung: Datenqualität, Kuratierung und sozialwissenschaftliche Forschungsdaten*

Der Rat für Informationsinfrastrukturen (RfII) fordert in einem jüngst veröffentlichten Strategiepapier zum Thema Open Data, den Fokus vermehrt auf die Qualität der veröffentlichten Forschungsdaten zu legen. So soll die Entwicklung „anstelle eines quantitativen Wachstums vor allem das qualitative Ziel hochwertiger Datenbestände und Datendienste verfolgen“ (RfII 2019: 2). „Open“, so die Autor*innen, werde noch zu oft mit „online“ gleichgesetzt, ohne dass von einer langfristigen Nachnutzbarkeit auszugehen sei (ebd.: 5). Der Mehrwert, den die Nachnutzbarkeit der Forschungsdaten bieten könne, wäre jedoch gerade dann gewährleistet, wenn ihre Verfügbarmachung mit einer Qualitätssicherung entlang fachlicher Standards verbunden werde.

In der Open-Data-Praxis gewinnt aufgrund dieser vermuteten Abhängigkeit zwischen der tatsächlichen Nachnutzbarkeit von Forschungsdaten sowie Art und Umfang der Qualitätssicherung das Aufgabengebiet der Datenkuratierung an Bedeutung. Doch was meint die Tätigkeit der „Kuratierung“ in Bezug auf Forschungsdaten eigentlich genau? Büttner et al. rekurrieren auf die alltagsgebräuchliche Verwendung des Begriffs der Kuratierung (z.B. in der Konservierung und Ausstellung von Museumsobjekten), um damit das Vorhaben zu charakterisieren, Wissen für zukünftige Forschung zu erhalten und gleichzeitig zu schaffen (Büttner et al. 2011: 17). Ulrich Herb folgend liegt die Aufgabe der Datenkuratierung darin, die „inhaltlich-logische oder wissenschaftliche Nutzbarkeit“ (Herb 2015: 134) der Forschungsdaten zu sichern. Der Vorgang der Kuratierung von Forschungsdaten wird demnach als

Prozess verstanden, der Forschungsdaten aufbereitet, sie verfügbar und vor allem qualitätsgesichert nachnutzbar macht.

Der vorliegende Beitrag nimmt die fachspezifischen Anforderungen der Kuratierung von sozialwissenschaftlichen Forschungsdaten in den Blick (vgl. Droß et al 2017). Disziplinär sind die Sozialwissenschaften durch eine erhebliche Vielfalt in den Wegen der Datengenerierung zur Beantwortung der Forschungsfragen gekennzeichnet. So werden in klassischen quantitativen Surveys unter Verwendung standardisierter Erhebungsinstrumente Mikrodaten erhoben. In experimentellen Laborsettings wird der Einfluss gezielter Stimuli auf Proband*innen untersucht oder in Feldexperimenten das Verhalten ganzer Untersuchungsgruppen in den Blick genommen. Vielfach werden jedoch auch prozessproduzierte bzw. Sekundärdaten aufbereitet (z.B. Text und Data Mining), weiterverarbeitet und mit neuen Informationen angereichert. Schließlich umfasst die qualitative Forschung ein ganzes Spektrum unterschiedlicher Erhebungsmethoden: Dies reicht von klassischen Expert*inneninterviews über ethnografische Forschungsansätze (teilnehmende Beobachtung, Feldnotizen, Fotografie), der Codierung von Sekundärtexten (Zeitungsartikel, Parteiprogramme) bis hin zu Videoaufzeichnungen.

2. *Die Kuratierung von Forschungsdaten am Wissenschaftszentrum Berlin für Sozialforschung (WZB)*

Das WZB ist Mitglied der Leibniz-Gemeinschaft und betreibt sozialwissenschaftliche Grundlagenforschung. Das empirisch und interdisziplinär ausgerichtete Forschungsprofil des WZB steht exemplarisch für die erwähnte methodische Vielfalt der Datenerhebung und -generierung in den Sozialwissenschaften. Den sowohl quantitativ als auch qualitativ forschenden Wissenschaftler*innen am Institut steht dabei eine forschungsnahe und bedarfsgerechte Unterstützung durch die Infrastrukturabteilung „Wissenschaftliche Information“ zur Seite, in der die Bereiche Bibliothek, Forschungsdatenmanagement, Institutsarchiv sowie Open Access zusammengefasst sind.

Seit 2017 existiert am WZB eine Leitlinie zum Umgang mit Forschungsdaten.[1] Das WZB fordert darin seine Forscher*innen nicht nur zu einem nachhaltigen Umgang mit den von ihnen produzierten und verwendeten Forschungsdaten auf. Es empfiehlt ausdrücklich die Veröffentlichung von am WZB erstellten Forschungsdaten auf einem geeigneten Repositorium. Hier weist die Leitlinie dem institutionellen Forschungsdatenmanagement eine besondere Rolle zu: Es soll die Forscher*innen bei der Veröffentlichung aktiv unterstützen und die Qualitätssicherung nach fachlichen Standards maßgeblich gewährleisten. Die Intention ist hierbei, dass ein zentral und institutionell angebotener Service der Kuratierung die Verfügbarkeit von WZB-Forschungsdaten als Ganzes fördert und diese darüber hinaus den oben genannten Qualitätsansprüchen genügen.

Die Kuratierung von sozialwissenschaftlichen Daten ist ein äußerst komplexer Vorgang, bei dem diverse praktische, organisatorische und rechtliche Aspekte zu beachten sind (u.a. Datenschutz der Proband*innen, die Rechte von Dritten, etc.). Ein zentraler Kuratierungsservice ist also aus institutioneller Perspektive sinnvoll: In Einzelfällen erlangte Expertise wird gebündelt und kann in verschiedenen Beratungsangeboten für weitere Nutzer*innen zur Verfügung gestellt werden. Zudem muss immer auch zwischen Aufwand und Nutzen der Datenaufbereitung abgewogen werden, bspw. wenn durch Anonymisierungsmaßnahmen Kontextinformationen entfernt werden (vgl. DGS 2019: 5f.). Zwar sind Forscher*innen die Expert*innen ihrer Daten, dennoch fehlt es ihnen häufig an zeitlichen und personellen Ressourcen, die Daten für eine Veröffentlichung vorzubereiten. Der institutionelle Service der Datenkuratierung entlastet die Forscher*innen somit auch erheblich in ihrer Arbeit.

Fachlich und personell ist das Forschungsdatenmanagement am WZB für die Kuratierung der Daten gut aufgestellt: Zu dem Bereich gehören ehemalige Wissenschaftler*innen, der Datenschutzbeauftragte des WZB sowie Fachangestellte für Markt- und Sozialforschung (FAMS), die die Abteilung auch ausbildet. Gleichzeitig war das WZB gemeinsam mit dem GESIS-Leibniz-Institut für Sozialwissenschaften, dem ZBW-Leibniz-Informationszentrum Wirtschaft und dem Deutschen Institut für Wirtschaftsforschung (DIW) Teil des SowiDataNet-Projektkonsortiums, welches ein Repositorium für Forschungsdaten aus den Sozial- und Wirtschaftswissenschaften

[1] Vgl. „Leitlinie zum Umgang mit Forschungsdaten am Wissenschaftszentrum Berlin für Sozialforschung". Online verfügbar unter: www.wzb.eu/system/files/docs/sv/win/forschungsdaten_leitlinien.de_.pdf (zuletzt geprüft am 11.04.2019).

entwickelt hat (vgl. Droß/Linne 2016). Zurzeit nutzt das WZB für die Veröffentlichung seiner Forschungsdaten SowiDataNet in der Betaphase.

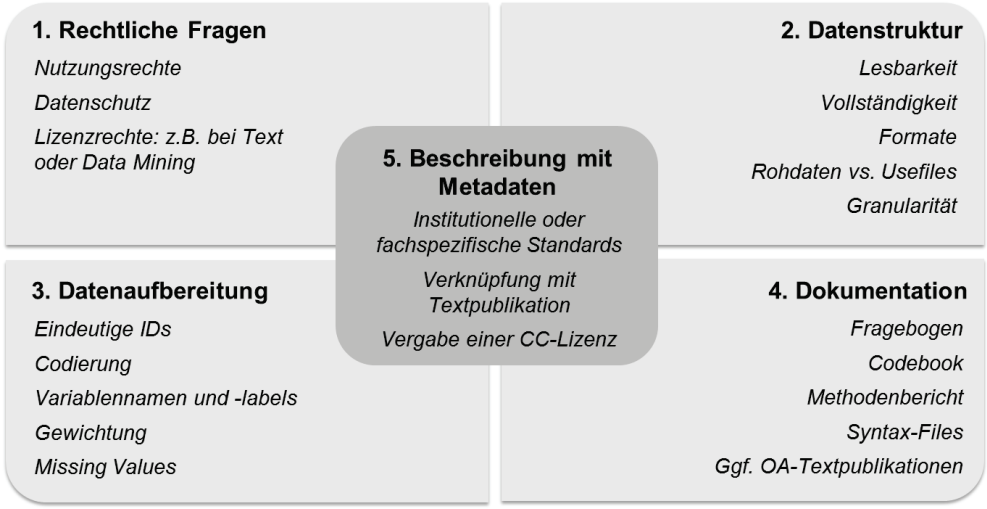


Abbildung 1: Praktische Schritte der Datenkuratierung am WZB

Der gegenwärtige Ablauf des zentralen Kuratierungsservices am WZB lässt sich grob in drei Schritten beschreiben: (1) Zu Beginn steht die Kontaktaufnahme, die beidseitig sowohl durch Forscher*innen aber auch proaktiv durch Mitarbeiter*innen des Forschungsdatenmanagements erfolgt. In einem Beratungsgespräch werden zunächst der Stand der Datenaufbereitung, rechtliche Fragen und selbstverständlich auch die prinzipielle Eignung und Nachnutzbarkeit der Forschungsdaten geklärt. (2) Anschließend stellen die Wissenschaftler*innen dem Forschungsdatenmanagement ihre Daten zur Verfügung. Hier beginnt der eigentliche Prozess der Kuratierung: Die Mitarbeiter*innen des Forschungsdatenmanagements prüfen Datenqualität, -struktur und -aufbereitung (siehe Abbildung 1), sichten die Begleitdokumentation und übernehmen die Metadaten-Beschreibung. Änderungen an den Daten erfolgen möglichst durch die Forscher*innen selbst oder nach Rücksprache mit den Forscher*innen durch das Forschungsdatenmanagement und werden nachvollziehbar dokumentiert. (3) Nach einer finalen Zustimmung durch die Forscher*innen werden die Forschungsdaten schließlich inkl. DOI-Registrierung über da|ra im SowiDataNet-Repositorium veröffentlicht.

3. Praxisfragen und Beispiellösungen

Am WZB ist – wie vermutlich in der gesamten Forschungsdatenlandschaft – die Implementierung von Best Practices des Forschungsdatenmanagements noch nicht abgeschlossen. Im Gegenteil: Im Rahmen der Kuratierung der Forschungsdaten stoßen wir immer wieder auf praktische Probleme, Fragen und komplexe Einzelfälle, für die noch keine Standards der Bearbeitung vorliegen, für die aber gleichwohl ein Umgang gefunden werden muss. Drei solcher „Problembereiche" mitsamt den praktischen Lösungsansätzen aus unserer alltäglichen Arbeit wollen wir im Folgenden vorstellen. Die Darstellungen haben dabei nicht den Anspruch abschließender Empfehlungen, sondern eher den Charakter eines Werkstattberichts. Keinesfalls ersetzten sie die Rechtsberatung im Einzelfall. Wir wollen mit ihnen vielmehr zu einer Diskussion beitragen und anregen, in der jenseits der allgemeinen Empfehlungen, inzwischen zahlreich vorliegenden Guidelines und Forschungsdaten-Richtlinien verstärkt die Praxisfragen der Datenkuratierung thematisiert werden, die uns Praktiker*innen in zunehmendem Maße beschäftigen.

3.1 Datenschutz: Die Bedeutung der informierten Einwilligung

In einer internen Befragung am WZB gaben 2015 rund 66 Prozent der Forscher*innen an, dass in ihren aktuellen Projekten neue Daten generiert werden. Da es sich im sozialwissenschaftlichen Kontext zumeist um Daten mit Personenbezug handelt, unterliegt der gesamte Verarbeitungsprozess dieser Forschungsdaten datenschutzrechtlichen Vorschriften. Diese wurden bis Mai 2018 durch das alte Bundesdatenschutzgesetz (BDSG) definiert und werden seither durch die Europäische Datenschutz-Grundverordnung (DSGVO) und das BDSG-neu (2018), ggf. auch durch die jeweiligen Landesdatenschutzgesetze geregelt.[2]

Wurden Datenschutz-Vorschriften in der Vergangenheit im Einzelfall nicht immer ganz exakt umgesetzt, stellte dies zumeist kein gravierendes Problem dar. Die Daten wurden in erster Linie ohnehin nur in der Forschung genutzt und, wenn überhaupt, nur

[2] Die Komplexität des Themas „Forschungsdatenmanagement und Datenschutz" kann im Rahmen dieses Beitrags nicht im Detail behandelt werden. Einen hervorragenden und v.a. aktuellen Überblick bieten Watteler/Ebel 2019.

in einem engen Forschungskontext unter Kolleg*innen geteilt. Ein Zugriff durch unbefugte Dritte war damit weitgehend ausgeschlossen. Mit der weltweiten Verfügbarkeit von in Repositorien veröffentlichten Datensätzen wird der datenschutzkonforme Umgang mit Forschungsdaten für Sozialwissenschaftler*innen jedoch wichtiger denn je. Der Aufbau von Datenschutz-Expertise im Forschungsdatenmanagement ist daher unerlässlich. Doch auch für die Forschung ist die Sensibilisierung notwendig, denn wie im Forschungsdatenmanagement insgesamt gilt: Je früher im Forschungsprozess über den Datenschutz nachgedacht wird, je einfacher ist die Datenveröffentlichung.

Die datenschutzrechtlichen Vorgaben der DSGVO sind im Grundsatz nicht neu. Maßgeblich bleibt, dass für die Verarbeitung personenbezogener Befragungsdaten in sozialwissenschaftlichen Projekten in den allermeisten Fällen eine informierte Einwilligung (informed consent) der Untersuchungspersonen eingeholt werden muss. In dieser stimmen die Befragten auf Basis ausreichender Angaben zum Forschungsvorhaben, der Erhebung und Verarbeitung ihrer Daten zu. Dabei sind nun jedoch auch die neuen Regelungen der DSGVO hinsichtlich der Informationspflichten zu beachten. So muss bspw. über die Betroffenenrechte, die Rechtsgrundlage und eine Kontaktmöglichkeit zum Datenschutzbeauftragten aufgeklärt werden. Insbesondere für die Betroffenenrechte enthalten die neuen gesetzlichen Vorschriften vielfältige Ausnahmeregelungen für die Forschung, die bislang jedoch nicht systematisch aufbereitet wurden. Als genereller Trend lässt sich zudem beobachten, dass im Zuge der Einführung der DSGVO das Thema Datenschutz eine deutlich größere Aufmerksamkeit seitens der Betroffenen, aber auch der Aufsichtsbehörden erfahren hat. Für Wissenschaftseinrichtungen ergibt sich daher die Notwendigkeit, sich verstärkt mit dem Thema Datenschutz zu beschäftigten – nicht zuletzt, weil das Fehlverhalten einzelner Forscher*innen in Form von Bußgeldern oder Reputationsschäden auf die gesamte Einrichtung zurückfallen kann.

Bei der Einholung des datenschutzrechtlichen informed consent bestehen in der Praxis nach unseren Erfahrungen erhebliche Unsicherheiten. Bei der Formulierung kommt es häufig zu zwei typischen Fehlern, die später zu unnötigen Barrieren für eine Nachnutzbarkeit der Forschungsdaten werden können: 1) Erklärungen sind zu knapp gehalten und genügen nicht den formalen gesetzlichen Anforderungen.

Hier besteht die Gefahr, dass die Gültigkeit der Einwilligung im Nachhinein angezweifelt werden kann. Eine Datenveröffentlichung wäre in diesen Fällen stets mit gewissen Unsicherheiten verbunden. 2) Erklärungen sind formal korrekt, es werden aber überzogene Einschränkungen aufgenommen: So ist bspw. die Zusage, dass „*sämtliche Daten* zum Ende des Projektes gelöscht werden“ oder, dass „*die Daten* keinesfalls an Dritte weitergeben werden“ für eine geplante Datenveröffentlichung problematisch, da sie eine Nachnutzung pauschal ausschließen. Hier steckt der Teufel quasi im Formulierungsdetail und es ist stets eine Differenzierung zwischen „den Forschungsdaten“ und „den personenbezogenen Merkmalen“ notwendig. Letztere sind zu löschen, wenn der Forschungszweck erreicht ist und dies sollte den Untersuchungsteilnehmer*innen auch zugesichert werden. Die durch die Löschung personenbezogener Identifikatoren anonymisierten Daten können jedoch zur Nachnutzung veröffentlicht werden. Idealerweise wird daher im Rahmen der Formulierung eines informed consent ein dreistufiges Vorgehen beschrieben, welches sowohl den Proband*innen Datenschutzkonformität garantiert, gleichzeitig aber auch die Veröffentlichung der Daten ermöglicht:

1) Die Trennung der Forschungsdaten von den personenbezogenen Daten während des Forschungsprozesses (Pseudonymisierung).
2) Die Löschung personenbezogener Merkmale spätestens mit Abschluss des Forschungsprozesses (also eine Anonymisierung der Daten) und
3) der Hinweis, dass anonymisierte Forschungsdaten archiviert und auch nach Abschluss des Forschungsvorhabens zur Nachnutzung bereitgestellt werden können. Dies sorgt für umfängliche Transparenz und rechtliche Absicherung.

Wie eng informierte Einwilligung und die spätere Nachnutzbarkeit der Forschungsdaten in Zusammenhang stehen, verdeutlicht auch ein weiteres Praxisproblem: Häufig wird den Untersuchungsteilnehmer*innen zugesichert, dass die Ergebnisse der Studie nur für wissenschaftliche Zwecke verwendet werden. Diese Einschränkung der Nachnutzung ist dabei vielfach auch ein expliziter Wunsch der Forscher*innen selbst. Denn: Aus ihrer Sicht besteht die Gefahr, „dass die Teilnahmebereitschaft an Umfragen/wissenschaftlichen Surveys und das Vertrauen von Versuchspersonen sich verändern, wenn Daten nicht allein wissenschaftlich genutzt werden“ (RfII 2019: 3). Doch so verständlich der Wunsch nach einer ausschließlich wissenschaftlichen Verwendung der Forschungsergebnisse auch sein mag, so sehr bereitet sie für eine spätere Nachnutzung der Daten Probleme.

Repositorien erlauben nämlich vielfach als Default einen Download auch für nichtwissenschaftliche Zwecke, was dann im Widerspruch zu Formulierung der Einwilligung steht.

Eine Einschränkung der Nachnutzung auf rein wissenschaftliche Zwecke lässt sich in der Praxis nur schwer umsetzen: Technisch hängt dies zunächst freilich von den Möglichkeiten des Repositoriums ab, in welchem die Daten abgelegt werden sollen. Theoretisch könnten hier Identitätsmanagementsysteme wie Shibboleth Abhilfe schaffen, deren Einsatz ist aber bislang kein Standard in der sozialwissenschaftlichen Forschungslandschaft. Einzelprüfungen der Nachnutzungszwecke sind im Rahmen eingeschränkter Datenzugänge in Repositorien zwar vielfach möglich, jedoch stets mit erhöhtem Arbeits- und Organisationsaufwand verbunden. Hier müsste festgelegt werden, wer im Einzelfall den Nachnutzungszweck kontrollieren (z.B. die Forscher*innen, das institutionelle Forschungsdatenmanagement) und wie die Prüfung erfolgen soll (E-Mail-Adresse, Zugehörigkeit zu einer Forschungseinrichtung, usw.). Letztlich steht die Beschränkung auf eine wissenschaftliche Nachnutzung auch der Idee einer generellen Offenheit entgegen, die mit dem Open-Science-Paradigma für Forschungsergebnisse angestrebt wird. Hierzu zählt auch, dass spätestens mit der Vergabe von offenen Lizenzen die Einschränkung auf einen bestimmten Kreis von Nachnutzer*innen nicht mehr vereinbar wäre.

Zusammenfassend lässt sich festhalten, dass dem Thema Datenschutz eine große Bedeutung im Prozess der Kuratierung sozialwissenschaftlicher Forschungsdaten eingeräumt werden muss. Das Fallbeispiel der informierten Einwilligung zeigt, dass es trotzdem möglich ist, die Forschungsdaten unter Wahrung der rechtlichen Schutzvorgaben nachnutzbar zu machen, wenn die Anforderungen von Beginn an im Forschungsprozess mitgedacht werden. Dies lässt sich am besten durch eine frühzeitige Beratung durch das Forschungsdatenmanagement sicherstellen. Gleichzeitig hat das Beispiel der Beschränkung der Nachnutzung auf wissenschaftliche Zwecke gezeigt, dass bei der Erstellung der informierten Einwilligung darauf geachtet werden muss, dass keine zu restriktiven Formulierungen verwendet werden, die eine Veröffentlichung der Forschungsdaten ausschließen. Gleichwohl muss hier festgehalten werden, dass gerade die Sozialwissenschaften eine Disziplin sind, in der nicht alle Forschungsdaten veröffentlicht werden können und in der insbesondere der Umfang der datenschutzkonformen Kuratierung stets in einem Nutzen-Aufwand-Verhältnis gesehen werden sollte (DGS 2019: 5).

3.2 Die Beachtung von Verwertungs- und Nutzungsrechten bei der Datenveröffentlichung

Neben den datenschutzrechtlichen Vorgaben ist innerhalb des Kuratierungsprozesses außerdem die Frage zu klären, wem die Forschungsdaten rechtlich zuzuordnen sind, d.h. wer letztlich über die Veröffentlichung und die Modalitäten der Nachnutzung durch Dritte entscheiden kann. Auch im sozialwissenschaftlichen Kontext wird daher aktuell diskutiert, ob Forschungsdaten als urheberrechtlich geschützte Werke zu betrachten sind. Wäre dies der Fall, läge die Entscheidungsbefugnis über die Datenveröffentlichung klar bei den jeweiligen Primärforscher*innen. Für die Sozialwissenschaften zeichnet sich gemäß eines jüngst veröffentlichten Kurzgutachtens jedoch die Einschätzung ab, dass ein urheberrechtlicher Schutz für Forschungsdaten, die durch standardisierte fachtypische Erhebungsmethoden, also bspw. mittels schriftlicher, Online- oder Telefonbefragung generiert werden, in der Regel nicht anzunehmen ist (vgl. Lauber-Rönsberg et al. 2018). Für qualitative Forschungsdaten könnte ggf. ein Urheberrechtsschutz bestehen, die Hürden sind jedoch in Hinblick auf die notwendige „geistige Schöpfung“ sehr hoch. Für den Einsatz fachüblicher Methoden, wie etwa der Verwendung eines entlang der Forschungsfragen entwickelten Interviewleitfadens, gehen wir daher aktuell ebenfalls eher nicht von einem Urheberrechtsschutz aus. Auch wenn diese Einschätzungen als allgemeine Richtschnur gelten können, kommen die Fachjurist*innen in ihrem Gutachten letztlich zu dem für Forschungsdatenpraktiker*innen nicht ganz befriedigendem Ergebnis, dass „die Schutzfähigkeit einzelner Forschungsdaten in der Regel nur im Einzelfall und selbst dann nicht mit hinreichender Rechtssicherheit beurteilt werden kann“ (ebd.: 3). Selbstverständlich hat dies auch Auswirkungen für die Vergabe von offenen Lizenzen. Zwar wird für die Veröffentlichung von Forschungsdaten aktuell zu der Verwendung der Creative-Commons-Lizenzen „CC0“ oder „CC-BY 4.0“ geraten. Jedoch stellt sich hier das Problem, dass, wenn für sozialwissenschaftliche Forschungsdaten überwiegend kein Urheberrechtsschutz besteht, sich letztlich aus den vergebenen Lizenzen keine rechtlich durchsetzbaren Pflichten, bspw. zur Namensnennung, ableiten lassen. Für unsere Praxis halten wir die angestrebte Vergabe von CC-Lizenzen dennoch für sinnvoll, da sie einerseits einen symbolischen Wert besitzen und zum anderen eine Absicherung für den (vermutlich seltenen) Einzelfall darstellen, in welchem es sich doch um urheberrechtlich geschützte Daten handelt.

Auch unabhängig vom Urheberrechtsschutz stellt sich in der Praxis des Forschungsdatenmanagements jedoch die ganz grundsätzliche Frage, wer über die Veröffentlichung der Daten entscheiden kann. Auch für nicht urheberrechtlich als Werk geschützte Forschungsdaten empfiehlt das Gutachten auf Grundlage des allgemeinen Persönlichkeitsrechts der Wissenschaftler*innen sowie aufgrund der forschungsinternen Normen der guten wissenschaftlichen Praxis, dass die Entscheidungsbefugnis über die Veröffentlichung der Daten bei den Primärforscher*innen liegen sollte. Zwar können sich im Rahmen von arbeits- bzw. dienstrechtlichen Regelungen Forschungsinstitute Nutzungsrechte an den Arbeitsergebnissen der Forscher*innen als Arbeitgeber sichern und die Daten auch ohne Zustimmung der Forscher*innen veröffentlichen. Für die aktuelle Praxis halten wir es jedoch grundsätzlich für empfehlenswert, Daten nicht über die Köpfe der Primärforscher*innen hinweg zu veröffentlichen und die datenproduzierenden Wissenschaftler*innen stets selbst darüber entscheiden zu lassen, ob und unter welchen Bedingungen ihre Daten für Dritte zugänglich gemacht werden sollen. Statt quasi-verpflichtender Bestimmungen sollte u.E. eher auf Informationsangebote, gute Argumente über die Vorzüge einer Datenveröffentlichung oder den Verweis auf die Empfehlungen der institutsinternen Forschungsdatenleitlinie zurückgegriffen werden.
Veröffentlichen die Forscher*innen ihre Daten auf eigenen Wegen, stellt sich die Frage der Zustimmung nicht. Veröffentlicht aber – wie am WZB – ein institutionelles Forschungsdatenmanagement die Daten für oder in Kooperation mit den jeweiligen Wissenschaftler*innen, empfiehlt es sich in jedem Fall, deren Zustimmung frühzeitig einzuholen. Dies ist auch gerade dann wichtig, wenn die Daten noch am Institut vorliegen, die Forscher*innen jedoch bereits an einer neuen Einrichtung tätig sind. Denn Daten werden häufig als Kopie mitgenommen und teils auch in die Forschungsvorhaben am neuen Institut eingespeist – bzw. können sie auch bereits am neuen Institut veröffentlicht worden sein.

Ein weiteres komplexes Szenario liegt für uns in der Klärung der Entscheidungsbefugnisse über Datenveröffentlichung und -verwendung in Kooperationsprojekten. Die Forschung in Verbünden nimmt zu und ist seitens der Forschungsförderer auch vielfach explizit erwünscht. Dieses Problem betrifft deshalb sowohl kleinere, oft nur halb formalisierte aber vor allem auch große Projektkooperationen.

Würde ein urheberrechtlicher Schutz bestehen, wären die Rechte aller Miturheber*innen zu berücksichtigen und zu schützen. Selbst wenn jedoch „nur“ von einem Persönlichkeitsrecht der Forscher*innen und den Normen der guten wissenschaftlichen Praxis auszugehen ist, sollte die gemeinsame Entscheidung und Abstimmung der Projektpartner*innen stets Voraussetzung für die Veröffentlichung der Daten sein. Diese erfolgt im optimalen Fall bereits vor Projektbeginn – bspw. über das Instrument einer Kooperationsvereinbarung – und wird im Detail durch einen Datenmanagementplan ergänzt, woran bereits zum Projektstart zu denken ist. Darin können einer Partnereinrichtung die Aufgaben der Datenaufbereitung und -veröffentlichung innerhalb der Projektkonstellation übertragen und die jeweiligen Bedingungen festgehalten werden. Hierbei wäre zum Beispiel zu denken an: Zeitpunkt und Ort der Veröffentlichung, die Dauer einer Embargo-Frist, die Einhaltung der Datenschutzvorschriften, die Vergabe einer CC-Lizenz und die Art der Nennung aller beteiligten Primärforscher*innen und Institute.

Findet die Klärung der Frage der Datenveröffentlichung zum Ende oder erst nach Abschluss eines Projektes statt – was in den Sozialwissenschaften häufig (noch) der Fall ist –, sollte ebenfalls ein Einverständnis aller Projektpartner*innen eingeholt werden. Oft sind zu diesem Zeitpunkt eine Vereinbarung oder ein Datenmanagementplan nicht mehr realisierbar. Am WZB sind wir daher darum bemüht, eine niedrigschwellige Zustimmung einzuholen, für die wir eine Vorlage bereithalten. Auch wenn uns in der Praxis noch kein Fall bekannt ist, bei dem es zwischen beteiligten Kooperationspartner*innen Interessenskonflikte bei der Veröffentlichung von Forschungsdaten gegeben hat, bietet diese Zustimmung doch diverse Vorteile: Sie gewährleistet ein Mindestmaß an Rechtssicherheit und Transparenz und verspricht, unkoordinierte Mehrfachveröffentlichungen derselben Forschungsdaten zu vermeiden. Gewiss bleiben frühzeitig auf Projektebene erstellte Datenmanagementpläne ohne Frage das Instrument der Wahl, um solche Szenarien zu verhindern, in den Sozialwissenschaften sind sie jedoch bislang eher die Ausnahme als die Regel.

3.3 Die Verknüpfung von Text- und Datenveröffentlichung

Ein Meilenstein für die Anerkennung von Forschungsdaten als eigenständiger wissenschaftlicher Output liegt in ihrer Zitierbarkeit. Vergleichbar zu den im Forschungsprojekt entstandenen Textpublikationen wird mit der Zitation von Forschungsdaten die Arbeit der Datenproduzent*innen als Forschungsleistung sichtbar. Auch wenn sich bereits verschiedene Initiativen und wissenschaftliche Communities damit beschäftigt haben, Empfehlungen zur Zitation von Forschungsdaten und bibliografischen Regeln zu erstellen, gibt es (noch) keinen einheitlichen Standard (vgl. Hausstein 2019). Weitgehend Einigkeit besteht jedoch darüber, dass aufgrund des dynamischen und fragilen Charakters digitaler Forschungsdaten Speicherort, eindeutige Version und Metadaten mittels persistenter IDs (z.B. Digital Object Identifier (DOI)) identifizierbar und auffindbar sein müssen. Dadurch wird es prinzipiell möglich, mittels der Datenzitation in den sich anschließenden Textveröffentlichungen auf die Forschungsdaten eindeutig und transparent zu verweisen.
Spätestens hier stellt sich die Frage nach der wechselseitigen Verknüpfung von Daten- und Textpublikationen. In der Praxis stößt man schnell auf die Herausforderung, dass die Aufnahme der Datenzitation in die erste Textpublikation häufig an ein spezifisches Zeitfenster gebunden ist. Die Option, einen Verweis auf die veröffentlichten Forschungsdaten inkl. DOI in einem Zeitschriftenaufsatz unterzubringen, besteht eigentlich nur vor der Einreichung des Artikels beim Verlag und ggf. noch im Rahmen des Review-Prozesses. Ab der finalen Veröffentlichung ist die Aufnahme im Aufsatz selbst de facto nicht mehr möglich. Wir bemühen uns daher in der Praxis, Forscher*innen diesbezüglich rechtzeitig zu informieren. Als ein nützliches Instrument hat sich zudem erwiesen, für die Forschungsdaten im Repositorium einen DOI reservieren zu lassen. Dadurch wird es möglich, die Datenzitation und -verlinkung in die Textpublikation aufzunehmen, auch wenn die tatsächliche Veröffentlichung der Forschungsdaten noch nicht final abgeschlossen ist.

2016 hat eine Studie die Zitationsrate von Forschungsdaten anhand des „Data Citation Index“ sowie dem heterogenen Feld der Altmetrics untersucht. Sie kam zu dem Ergebnis, dass 85 % Prozent der Forschungsdaten unzitiert bleiben (Peters et al. 2016: 740). Gewiss müsste hier tiefer in die Analyse der Kontextbedingungen eingestiegen

werden. Fachspezifische Standards wären zu berücksichtigen und es ist davon auszugehen, dass mit dem zunehmenden Bedeutungsgewinn von Open Data in den letzten Jahren die Tendenz steigend sein dürfte. Dennoch lässt sich ohne weiteres festhalten, dass die Zitationsmechanismen bei Forschungsdaten noch nicht analog zu den Textpublikationen funktionieren. Technisch wird es in Zukunft mit Sicherheit möglich werden, mittels automatisierter Prozesse Zitationen von Forschungsdaten zu erfassen und sichtbar zu machen. Vollintegrierte Forschungsinformationssysteme, die Publikationen, Daten, Projektinformationen u.ä. verbinden, könnten eine optimale Lösung darstellen. Sie sind jedoch mit hohem Entwicklungs- und Ressourcenaufwand verbunden und daher nicht an jedem Institut realisierbar. Zwar ist davon auszugehen, dass die Suche nach und Auffindbarkeit von Forschungsdaten mittelfristig über Fachrepositorien oder Meta-Suchen stattfinden wird. Unsere Erfahrung für die Sozialwissenschaften zeigt jedoch, dass bislang wenig nach Daten in Repositorien recherchiert wird – und wenn dann vor allem durch Forschungsdatenmanager*innen, die Datenrecherchen als Service anbieten.

Es ist folglich davon auszugehen, dass die Forscher*innen einstweilen über die klassische Suche nach Textpublikationen indirekt auf die dazugehörigen Forschungsdaten stoßen. Am WZB haben wir uns daher zunächst für ein pragmatisches Vorgehen entschieden. Sofern Forscher*innen Textpublikationen auf Basis ihrer am WZB veröffentlichen Forschungsdaten erstellen, bitten wir sie, einen Hinweis zur Datenveröffentlichung inkl. DOI im Artikel zu ergänzen. Wird die Textpublikation nach Veröffentlichung im Bibliothekskatalog aufgenommen, ergänzen wir im Katalog zudem den Verweis ins Forschungsdaten-Repositorium. Im Falle von Zweitveröffentlichungen im Open-Access-Repositorium wird auch hier der Ablageort der dazugehörigen Forschungsdaten angeführt. Schließlich nehmen wir zugehörige Textpublikationen der Forscher*innen im Datenrepositorium auf. Denn auch für potentielle Nachnutzer*innen der Daten ist es sicher von großem Interesse, welche Primär- bzw. Sekundärpublikationen bereits auf Grundlage der Daten erstellt wurden. Doch hier ergibt sich ein weiteres zeitliches Problem: Werden die Daten am Ende eines Forschungsprojektes veröffentlicht, liegen die Textpublikationen noch nicht vor. Selbst bei kleineren Projekten kann sich in den Sozialwissenschaften die Publikationsphase über mehrere Jahre erstrecken. Hier wäre die Anforderung entsprechend hoch, die Literatur im Repositorium nachträglich zu ergänzen.

In der Praxis versuchen wir als institutionelles Forschungsdatenmanagement daher zumindest, dies für die Hauptpublikationen der Datenproduzent*innen zu realisieren und bitten die Forscher*innen nach der Datenveröffentlichung entsprechend, uns über aktuelle Publikationen zu informieren.

4. Fazit

Für das Ziel, Datenbestände langfristig verfügbar und nachnutzbar zu machen, ist eine Qualitätssicherung entlang fachlicher Standards unverzichtbar. Der Weg zur Entwicklung dieser Standards führt unseres Erachtens über die alltägliche Praxis der Datenkuratierung und den Austausch über die jeweils fachspezifischen Probleme und bereits vorliegenden Lösungsvorschläge. Es konnte mit den drei vorgestellten Praxisbeispielen gezeigt werden, dass eine frühzeitige Berücksichtigung von zentralen Aspekten der Kuratierung innerhalb des Forschungsprozesses die Nachnutzbarkeit von Forschungsdaten fördert. Sowohl die Thematik der informierten Einwilligung aus dem Bereich Datenschutz, die Verständigung über Verwertungs- und Nutzungsrechte bei der Veröffentlichung, aber auch die Verknüpfung von Text- und Datenpublikation haben zudem verdeutlicht, dass die forschungsnahe Infrastruktur hierbei eine zentrale Rolle einnimmt. Es muss sich noch zeigen, wie viel Standardisierung die Kuratierung letztlich zulassen wird. Angesichts der Vielfalt und Komplexität der hier aufgerufenen Themen steht bereits jetzt schon außer Frage, wie groß hierbei der Bedarf an Expertise und Unterstützungsangeboten durch ein professionelles institutionelles Forschungsdatenmanagement ist.

5. Literatur

Büttner, S., Hobohm, H.-C. & Müller, L. (2011): Research Data Management. In: dies. (Hg.): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen, S. 13–24.

Deutsche Gesellschaft für Soziologie (DGS) (2019): Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie. Stellungnahme des Vorstands und Konzils der DGS. Online verfügbar unter https://www.soziologie.de/uploads/media/DGS-Stellungnahme_zum_Forschungsdatenmanagement_08.01.2019_01.pdf, zuletzt geprüft am 26.04.2019.

Droß, P. J. & Linne, M. (2016): Sicheres und einfaches Data Sharing mit SowiDataNet. Dokumentieren – veröffentlichen – nachnutzen. In: Bibliotheksdienst, Vol. 50, H. 7, S. 649–660. Online verfügbar unter https://doi.org/10.1515/bd-2016-0079.

Droß, P., Fräßdorf, M., Kubaty, P. & Naujoks, J. (2017): Open Data in den Sozial- und Wirtschaftswissenschaften: das Forschungsdatenrepositorium SowiDataNet, In: Kratzke, J. & Heuveline, V. (Hg.): E-Science-Tage 2017: Forschungsdaten managen. Heidelberg: heiBOOKS, Heidelberg, S. 31-42. Online verfügbar unter http://hdl.handle.net/10419/172495.

Hausstein, B. (2019): Zitierbarmachung und Zitation von Forschungsdaten. In: U. Jensen, S. Netscher und K. Weller (Hg.): Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten. Opladen, Berlin, Toronto: Barbara Budrich, S. 177–192. Online verfügbar unter https://doi.org/10.3224/84742233.11.

Herb, U. (2015): Open Science in der Soziologie. Eine interdisziplinäre Bestandsaufnahme zur offenen Wissenschaft und eine Untersuchung ihrer Verbreitung in der Soziologie. Glückstadt: VWH. https://zenodo.org/record/31234#.WXSd4elLdPY.

Lauber-Rönsberg, A., Krahn, P. & Baumann, P. (2018): Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements. Kurzfassung. Im Rahmen des DataJus-Projektes. Online verfügbar unter https://tu-dresden.de/gsw/jura/igetem/jfbimd13/ressourcen/dateien/publikationen/DataJus_Zusammenfassung_Gutachten_12-07-18.pdf, zuletzt geprüft am 26.04.2019.

Peters, I., Kraker, P., Lex, E., Gumpenberger, C., Gorraiz, J. (2016): Research data explored: an extended analysis of citations and altmetrics. In: *Scientometrics* 107, S. 723–744. Online verfügbar unter https://doi.org/10.1007/s11192-016-1887-4.

Rat für Informationsinfrastrukturen (RfII) (2019): Stellungnahme des Rates für Informationsinfrastrukturen (RfII) zu den aktuellen Entwicklungen rund um Open Data und Open Access. Online verfügbar unter http://www.rfii.de/?p=3748, zuletzt geprüft am 26.04.2019.

Watteler, O., Ebel, T. (2019): Datenschutz im Forschungsdatenmanagement. In: U. Jensen, S. Netscher und K. Weller (Hg.): Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten. Opladen, Berlin, Toronto: Barbara Budrich, S. 57–80. Online verfügbar unter https://doi.org/10.3224/84742233.05.

Entwicklung und Betrieb eines Metadatenmanagement-systems für Forschungsdaten aus dem Bereich der Hochschul- und Wissenschaftsforschung – Lessons Learned

Karsten Stephan, René Reitmann

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Hannover

Abstract

Im Rahmen des Aufbauprojektes und der ersten Betriebsphase des Forschungsdatenzentrums für Hochschul- und Wissenschaftsforschung wurde in den Jahren 2015 bis 2019 mit einem interdisziplinären Team ein komplexes webbasiertes informationstechnisches System entwickelt. Das Metadatenmanagementsystem dient der Ausweisung der im Forschungsdatenzentrum verfügbaren Forschungsdaten, der strukturierten Erfassung und Editierung von Metadaten zu diesen Daten sowie der Suche im vorliegenden Metadatenbestand. Zentraler Erfolgsfaktor der Entwicklung war die Kombination aus softwaretechnischer Expertise, einer klaren Vision des zu entwickelnden Systems und der Einbringung von fachlichem Domänenwissen in die Entwicklung. Mit einem noch stärkeren Einbezug von Domänenwissen und einer deutlich stärkeren Berücksichtigung interdisziplinärer Dynamiken hätte der Entwicklungsprozess weiter optimiert werden können.

As part of the building phase and the first operating phase of the Research Data Centre for Higher Education Research and Science Studies, an interdisciplinary team developed a complex web-based information technology system between 2015 and 2019. The metadata management system serves to identify the research data available in the Research Data Centre, to record and edit the metadata for this data in a structured way and to search the available metadata. The central success factor of the development was the combination of software engineering expertise, a clear vision of the system to be developed and the integration of domain knowledge into the development. With an even stronger inclusion of domain knowledge and a significantly stronger consideration of interdisciplinary dynamics, the development process could have been further optimized.

Einleitung

Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) ist ein durch Bund und Länder gefördertes Forschungsinstitut mit Sitz in Hannover und Berlin. Als internationales Kompetenzzentrum der Hochschul- und Wissenschaftsforschung führt das DZHW Datenerhebungen und Analysen durch, erstellt forschungsbasierte Dienstleistungen für die Hochschul- und Wissenschaftspolitik und stellt der Scientific Community eine Forschungsinfrastruktur im Bereich der Hochschul- und Wissenschaftsforschung zur Verfügung.

Als Teil dieser Forschungsinfrastruktur wurde von März 2015 bis Mai 2017 in einem vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projekt ein Forschungsdatenzentrum (FDZ) für das Forschungsfeld der Hochschul- und Wissenschaftsforschung aufgebaut. Das FDZ nimmt anonymisierte Daten quantitativer Befragungen und qualitativer Interviews auf, dokumentiert diese und stellt sie der wissenschaftlichen Community für Forschungs- und Lehrzwecke zur Verfügung. Der Datenbestand des FDZ setzt sich aus DZHW-eigenen und externen Studien des Forschungsfeldes zusammen. Arbeitsschwerpunkte des FDZ-Aufbauprojektes waren eine nachträgliche Aufbereitung ausgewählter DZHW-Studien als Datenstartbestand des FDZ sowie der Aufbau der Infrastruktur für die Datenherausgabe des zukünftigen FDZ. Das entwickelte Metadatenmanagementsystem (MDM[1]) ist zentraler Bestandteil dieser Infrastruktur. Das MDM dient der Ausweisung der im FDZ vorhandenen Daten, der strukturierten Erfassung und Editierung von Metadaten zu den Daten sowie der Suche im vorliegenden Metadatenbestand.

Im Projekt waren insgesamt 13 Personen beschäftigt, für die drei informationstechnischen sowie einen sozialwissenschaftlichen Mitarbeiter bildete die Entwicklung des MDM den Schwerpunkt ihrer Arbeit. Zur ersten Ausbaustufe des MDM gehören ein zentraler Metadatenspeicher, Prozesse zur Erzeugung der zu importierenden Metadaten sowie eine Suchfunktionalität. Diese Komponenten gingen mit Eröffnung des FDZ im Juni 2017 in den

[1] Metadatenmanagementsystem: https://metadata.fdz.dzhw.eu,
Entwicklungsrepository: https://github.com/dzhw/metadatamanagement.

Produktivbetrieb. Von Juni 2017 bis Mai 2019 wurden in einer zweiten Ausbaustufe[2] verschiedene Komponenten zur Unterstützung der qualitätsgesicherten Datenaufnahme ergänzt.

In der Planungs- und Beantragungsphase des Projektes stand zur Debatte, bestehende technische Metadatensysteme anderer Forschungsdatenzentren bzw. einzelne Teile dieser Systeme einzusetzen anstatt ein eigenes Metadatenmanagementsystem zu entwickeln, insbesondere weil zu diesem Zeitpunkt noch nicht absehbar war, inwieweit es gelingen würde, erfahrenes informationstechnisches Personal zu rekrutieren. Nachdem ein erfahrener Softwarearchitekt gewonnen werden konnte, fiel die Entscheidung zugunsten eines eigenen Systems aus, zumal auch nur wenige vorhandene Metadatensysteme existierten, die die Anforderung erfüllten, Metadaten stark strukturiert zu speichern und gleichzeitig diese Systeme zu individuell auf die Belange der entwickelnden Einrichtungen zugeschnitten waren. Diese Entscheidung hat sich im Rückblick als richtig erwiesen. Dies ist auch dem Umstand zu verdanken, dass es durch die Kombination der Kenntnisse der Mitarbeiter und der Möglichkeit von Schulung und externer Beratung gelungen ist, die benötigte informationstechnische Expertise zu den verwendeten Techniken aufzubauen. Zudem konnte im weiteren Verlauf des Projektes an unterschiedlichen Punkten auf etablierte Open Source Frameworks und Komponenten zurückgegriffen werden.

Das Aufbauprojekt des FDZ und damit auch die Entwicklung der ersten Ausbaustufe des MDM standen unter bestimmten Rahmenbedingungen. Als zeitlicher Rahmen waren die schon erwähnten 27 Monate, startend ab März 2017 festgelegt. Die zwei Hauptaufgaben des Projektes – Aufbereitung und Dokumentation von Datensätzen zur Sekundärnutzung inklusive der Festlegung von Dokumentationsstruktur und -inhalten einerseits und Aufbau einer (technischen) Infrastruktur zum Metadatenmanagement und zur Aufnahme und Herausgabe dieser Datensätze andererseits – erforderten ein aus Sozialwissenschaftler*innen und Softwareentwickler*innen interdisziplinär zusammengesetztes Projektteam. Aufgrund der relativ kurzen Projektlaufzeit

[2] Vom BMBF im Rahmen des Projektes „Maßnahmen zur Effizienzsteigerung des FDZ-DZHW in den Bereichen Datenaufnahme, Nutzeranfragen und Kommunikation“ gefördert.

mussten beide Aufgabenbereiche parallel bearbeitet werden, waren aber gleichzeitig stark voneinander abhängig. Als besonders herausfordernd erwies sich, dass parallel zur Erstellung eines inhaltlichen Domänenmodells (welche Elemente beinhaltet eine Studie, in welchen Beziehungen stehen diese zueinander und welche Metadaten müssen für welches Element erfasst werden) schon mit der Arbeit an dem technischen System begonnen werden musste, dessen Basis eben dieses Domänenmodell bildete. Die Folge waren zahlreiche Anpassungen, aber auch Neukonzeptionen von Struktur und Programmcode des MDM. Eine längere Projektlaufzeit mit weniger Personalressourcen pro Zeiteinheit sowie die sukzessive anstelle der gleichzeitigen Rekrutierung der Mitarbeiter*innen erscheint im Rückblick sinnvoll.

Funktionalität des MDM – Lessons Learned

Die Vision hinter der Entwicklung des MDM ist die Verfügbarkeit eines Systems, das Transparenz zum Forschungsdatenbestand des Forschungsfeldes herstellen, die Nachvollziehbarkeit von Forschungsdaten und -prozessen gewährleisten und Forschungsprozesse von Wissenschaftler*innen unterstützen kann. Dazu müssen detaillierte, strukturierte Beschreibungen der Forschungsdaten und ihrer Entstehungsprozesse generiert und auffindbar sein.

Zielgruppe des MDM sind (potentielle) Datennutzer*innen auf nationaler und internationaler Ebene, die sich in der Regel im Bereich der Hochschul- und Wissenschaftsforschung bewegen. Das Auffinden von Daten und die Beurteilung der Eignung für das eigene Forschungsvorhaben soll aber auch Forscher*innen aus angrenzenden Disziplinen oder Studierenden möglich sein, die die konkreten Daten, die konkrete Studienreihe, das FDZ als deutsche Anlaufstelle für Forschungsdaten des Forschungsfeldes oder auch die im Forschungsfeld üblicherweise verwendeten Begrifflichkeiten und Ordnungskriterien (noch) nicht kennen.

Auf Grundlage von Vision und Zielgruppe leiteten sich Anforderungen an die Funktionalität des MDM ab, die teilweise vor Projektstart spezifiziert worden waren, häufig aber auch erst im weiteren Verlauf des Projektes entstanden.

Gemäß der zentralen Forderung einer möglichst weitreichenden Transparenz zu den Forschungsdaten und ihrem Entstehungsprozess, sollten alle Informationen zu den Daten, die nicht datenschutzrechtlich sensibel sind, über das zu entwickelnde System offen zugänglich und strukturiert durchsuchbar sein.

Forschungsdatensätze der Hochschul- und Wissenschaftsforschung enthalten oft anonymisierte, aber dennoch datenschutzrechtlich sensible Angaben. Im DZHW werden sie daher in einem besonders geschützten Datennetz verarbeitet. Die Metadaten hingegen, die umfangreiche Informationen zum Datengenerierungsprozess und zu den Eigenschaften der resultierenden Forschungsdatensätze enthalten, sind datenschutzrechtlich unbedenklich. Schon zu Beginn des Projektes wurde entschieden, dass das MDM ausschließlich Metadaten verarbeiten und nicht die Funktion der Bereitstellung der Forschungsdatensätze übernehmen sollte. Damit war es ohne aufwändige Datenschutzklärungen möglich, aktuelle Cloud-Technologien einzusetzen, die im Entwicklungsprozess schnelle Technologiewechsel und beim Betrieb des Systems flexible Skalierungsmöglichkeiten boten.

Die Metadaten sollten strukturiert in einem zentralen Metadatenrepositorium gehalten werden. Die strukturierte Speicherung der Metadaten bildet dabei die Basis sowohl für die Erhöhung des Automatisierungsgrades im Rahmen der FDZ-Prozesse als auch im Rahmen des hauseigenen Forschungsdatenmanagements. Sie ermöglicht die Implementierung einer elaborierten Suche, erleichtert Metaanalysen und gestattet die Entwicklung automatisierter Reports. Zudem erlaubt sie die Anbindung des MDM an andere Systeme. Dies betrifft sowohl hausinterne Systeme der Datenproduktion als auch externe Dienste, wie den Dienst da|ra[3] zur Registrierung von Digital Object Identifiern (DOI). Mit der strukturierten Speicherung ist auch die Voraussetzung gegeben, das MDM an übergreifende Dienste und Metaportale anbinden zu können, die im Kontext der aktuellen Vernetzungsbestrebungen

[3] Registrierungsagentur für Sozial- und Wirtschaftsdaten: http://www.da-ra.de/home/.

im Bereich der Forschungsdateninfrastruktur aufgebaut werden (wie z.B. NFDI[4] oder EOSC[5]).

Für die Registrierung einer DOI für eine in das FDZ aufgenommenen Studie muss ein Satz bestimmter Metadaten zu dieser Studie an den Registrierungsdienst da|ra übermittelt werden. Für das MDM wurde entschieden, den Prozess zur Registrierung von DOI zu automatisieren. Der Implementierungsaufwand für diese Funktionalität war nicht gering und angesichts der in der Projektlaufzeit im FDZ befindlichen Anzahl an Studien überstieg der Aufwand den einer manuellen Registrierung mit Hilfe der von da|ra angebotenen Webmaske deutlich. Die automatisierte Anbindung an den Registrierungsdienst war eine bewusst vorgesehene Investition in die Zukunft. Der wichtigste Grund für die automatisierte Anbindung an da|ra war, im MDM die grundlegende Programmlogik zu schaffen, um die Inhalte des MDM mit geringem Aufwand zukünftig auch mit anderen informationstechnischen Systemen vernetzen zu können.

Die Metadaten, die für das MDM erfasst, strukturiert und geprüft werden müssen, stammen aus unterschiedlichen Quellen. Ein Teil der Metadaten kann aus den Datendateien der Forschungsdaten extrahiert werden. Andere Metadaten müssen von Mitarbeiter*innen manuell erfasst und geprüft werden. Wieder andere Metadaten können direkt aus den Softwaresystemen, die im Haus zur Datenerhebung verwendet werden (z.B. Onlinebefragungssystem), extrahiert werden. Im MDM sollten die Prozesse der Metadatenerfassung durch Menschen möglichst einfach gestaltet werden. Zudem sollte eine möglichst umfassende Validierung beim Import erreicht werden. Das letztendliche Ziel bestand darin, einen möglichst hohen Automatisierungsgrad zu erreichen, und zur Qualitätssicherung verstärkt bewährte Verfahren aus der Softwareentwicklung (wie z.B. Versionskontrolle, automatisierte Tests) in die Prozesse einfließen zu lassen. Der Teil der Metadaten, der direkt den Forschungsdatensätzen entstammt, konnte aus datenschutzrechtlichen Gründen nur im geschützten Datennetz generiert werden. Da dieses wiederum nicht an das offen zugängliche MDM angebunden werden durfte, wurden die

[4] Nationale Forschungsdateninfrastruktur: https://www.dfg.de/foerderung/programme/nfdi/index.html
[5] European Open Science Cloud: https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud.

Arbeitsprozesse zur Erzeugung und Erfassung der Metadaten noch komplizierter und mussten im Laufe des Projektes auch mehrfach angepasst werden. Nichtsdestotrotz ist auch im Rückblick der hohe Aufwand gerechtfertigt, um – im Spannungsfeld zwischen datenschutzrechtlichen Vorgaben und adäquater Forschungsdatenbeschreibung – (potentiellen) Datennutzer*innen strukturierte Metadaten in möglichst weitreichendem Umfang zur Verfügung stellen zu können.

Um die Auffindbarkeit der Daten für Forschende verschiedener Disziplinen auch im internationalen Kontext zu erleichtern, wurde eine moderne, komfortabel nutzbare Suchfunktionalität benötigt, die Recherchen in deutscher und englischer Sprache ermöglicht. Für die Entwicklung des MDM wurde entschieden, Informationen detaillierter zu erfassen als dies in vielen anderen Metadatensystemen im Bereich der empirischen Sozialforschung üblich war, um so zu den relevanten Elementen einer Studie (Studie, Datenerhebungen, Fragebögen, Fragen, Datensätze, Variablen, datensatzbezogene Publikationen) eigene Detailseiten anbieten zu können. Eine Detailseite führt verschiedene Eigenschaften des entsprechenden Elementes auf. Auf diese Weise sind die einzelnen Elemente auch als Suchergebnisse auffindbar und diese über Filtermöglichkeiten eingrenzbar. Die Detailseiten sollten nicht nur über die Suche innerhalb des Systems sondern auch über Suchmaschinen auffindbar sein. Darüber hinaus sollten sie über unveränderliche, „bookmarkable“ URL aufgerufen werden können. Anders als in Metadatensystemen mit flacher Hierarchie der Seitenstruktur, die stark auf die Beschreibung einer Studie fokussieren, bestand im MDM durch die Konzentration auf die Detailseiten der Einzelelemente einer Studie die Notwendigkeit, in der Benutzungsoberfläche die Beziehungen der Elemente einer Studie zueinander sowie den Gesamtzusammenhang der Studie expliziter darzustellen. Dies wurde gelöst, indem auf jeder Detailseite Links zu den (bzw. zu Listen von) Detailseiten der verbundenen Elemente angezeigt werden. Auf der Detailseite einer Studie werden zudem zusätzlich die wichtigsten Informationen zu dieser Studie in aggregierter Form dargestellt, da dies der Erwartungshaltung vieler Nutzer*innen entspricht, die mit anderen Metadatensystemen vertraut sind. Die Fragen, inwieweit dieser

Erwartungshaltung im MDM entsprochen werden sollte und inwieweit die Komplexität der Domäne vor den Nutzer*innen verborgen werden sollte, waren und sind große Diskussionspunkte innerhalb des Projektteams. Für die nähere Zukunft sind Usability Tests geplant, die auch dazu führen könnten, dass bestimmte Entscheidungen zur Gestaltung der Benutzungsoberfläche revidiert werden.

Schon zu Beginn des Projektes stand fest, dass das MDM langfristig auch Archivprozesse unterstützen muss: Wenn die Daten einer Studie veröffentlicht werden, erhalten Daten und Metadaten eine Versionsnummer. Datensätze einer Version dürfen nicht mehr verändert werden. Falls relevante Änderungen an den Daten oder den verbundenen Metadaten vorgenommen werden, muss eine neue Version mit einer neuen Versionsnummer freigegeben werden. Zur Sicherstellung der Nachvollziehbarkeit von Publikationen, die auf der Grundlage älterer (Meta)Datenversionen entstanden sind, müssen ältere Versionen aber weiterhin im MDM verfügbar bleiben. Der Aufwand für die Implementierung dieser Funktionalität ist vergleichsweise hoch. Im Laufe des Projektes wurde daher entschieden, den Bereich der Archivierung zugunsten anderer Funktionalitäten zunächst zurückzustellen. Gegen Ende des Projektes war ersichtlich, dass Archivierung während der Projektlaufzeit nicht mehr zu realisieren war, so dass das MDM ohne eine Archivfunktionalität in den Produktivbetrieb ging. Als Workaround wurden (manuelle) Prozesse definiert, wie ältere (Meta)Datenversionen bis zur Verfügbarkeit einer Archivfunktion außerhalb des MDM verwahrt werden. Erst seit März 2019 steht die Funktionalität im MDM zur Verfügung. Damit musste der manuelle Workaround länger bestehen als ursprünglich gewünscht. Im Rückblick zeigt sich aber, dass die späte Umsetzung der Archivfunktionalität sinnvoll war, zumal sich ab Produktivsetzung dieser Funktionalität der Aufwand zur Implementierung von Erweiterungen, insbesondere der Benutzungsoberfläche, deutlich erhöht hat.

Ebenfalls für die zweite Ausbaustufe des MDM war die Implementierung der Prozesse zur qualitätsgesicherten Datenaufnahme vorgesehen. Unterstützt werden sollte dabei in erster Linie eine forschungsprozessbegleitende Dokumentation von Studien. Hierbei werden Elemente des Forschungsprozesses möglichst früh im Forschungsprozess von den

Primärforschenden selbst qualitätsgesichert dokumentiert. Die Dokumentationen zu einer Studie wachsen so mit fortschreitendem Projektstand langsam an. Der Ansatz der forschungsprozessbegleitenden Dokumentation steigert die Qualität der Dokumentation. Zudem können Forschungsdaten relativ schnell mit Abschluss des datenerhebenden Projektes herausgegeben werden. Die Erfassung und Editierung der Dokumentationen wird im MDM durch Eingabemasken unterstützt, die Validatoren zur formalen Prüfung der erfassten Metadaten enthalten. Zudem wird die komfortable Erfassung der Metadaten durch die Bereitstellung eines Zugriffs auf die Änderungshistorie und damit durch die Möglichkeit des Rückgängigmachens von vorgenommenen Änderungen unterstützt. Für die Prozesse der forschungsprozessbegleitenden Dokumentation wurde ein „Projektcockpit“ implementiert, welches den Dokumentationsstand des zu erfassenden Projektes darstellt, die Kommunikation zwischen Datengebenden und FDZ-Mitarbeiter*innen zum Stand und zur Qualität der erstellten Dokumentationen unterstützt sowie die komfortable Navigation zu den Detailsseiten der zu dokumentierenden Elemente ermöglicht. Wie auch bei der Archivierungsfunktionalität war der Aufwand zur Implementierung der forschungsprozessbegleitenden Dokumentation vergleichsweise hoch. Anders als bei der Archivierungsfunktionalität stand hier aber schon vor Projektbeginn fest, dass die Implementierung dieser Funktionalität erst in einer zweiten Aufbauphase des MDM erfolgen sollte. Zwar hätten Eingabemasken die beschriebenen komplexen Prozesse der Metadatenerfassung während der ersten Aufbauphase etwas vereinfachen können. Doch waren die Aufwände zur Umsetzung der forschungsprozessbegleitenden Dokumentation zu wenig einschätzbar und somit das Risiko, dann basalere Funktionalitäten des MDM innerhalb der begrenzten Projektlaufzeit nicht mehr umsetzen zu können, zu groß. Und obwohl Funktionalitäten wie Archivierung und forschungsprozessbegleitende Dokumentation sehr wertvoll und außerdem auch „einfach cool“ sind, folgte die Entwicklung des MDM auch hier dem Grundsatz, zum einen zunächst die wirklich basalen Funktionalitäten zu implementieren und zum anderen Funktionalitäten erst auf eine einfachere Art und Weise und erst im zweiten Schritt in einer komplexeren Variante umzusetzen.

Technologische Entscheidungen

Das MDM ist eine Single Page Application mit einem zustandslosen Server. Der Server nutzt mehrere externe Dienste u.a. für die Datenspeicherung, die Suche und die Kommunikation mit den Nutzer*innen. Der Client ist eine Single Page Application, die mit AngularJS realisiert wurde. Die GUI-Komponenten von Angular Material sind die Grundlage für das visuelle Design der Anwendung. Die Angular Ressourcen bilden eine Fassade für die Kommunikation mit dem HTTP-Server. Der Server ist eine 12-factor-app[6], welche Spring Boot (ein Java-Framework) und Cloudfoundry (eine Plattform für skalierbare Webanwendungen) nutzt, um die 12 Faktoren in einer Java Anwendung zu implementieren.[7]

Der Serverprozess ist ein Java-Prozess, der eine JSON-over-HTTP API (manchmal auch REST-API genannt) für die Kommunikation mit dem Client anbietet. Die REST-API wird aktuell mit Spring MVC realisiert. Zur Benutzerauthentifizierung wird OAuth2 verwendet, welches mit Spring Security implementiert wurde. Der Serverprozess ist zustandslos und kann dadurch sehr einfach skaliert werden. Um die Containerisierung der Java-Anwendung musste sich das Entwicklungsteam dank Cloudfoundries Buildpacks nicht selber kümmern. Die Containerinstanzen werden von Cloudfoundry automatisch neugestartet, falls sie nicht mehr „gesund“ sind, d.h., wenn sie auf spezielle HTTP Requests nicht in einer vorgegebenen Zeit mit einem vorgegebenen Wert antworten. Der Loadbalancer von Cloudfoundry übernimmt außerdem automatisch die Verteilung der HTTP-Requests auf die unterschiedlichen Containerinstanzen. Zusätzlich übernimmt er das SSL-Offloading. Für die Fehleranalyse sammelt Cloudfoundries Loggregator automatisch die Logausgaben aller Serverinstanzen ein. Metriken über CPU-Last, Memoryverbrauch, etc. sowie eigene Metriken werden von Pivotals Cloudfoundry Instanz (Pivotal Web Services) ebenfalls von den Containerinstanzen für die Fehleranalyse eingesammelt. Intern ist die Java Anwendung wie eine klassische Spring Anwendung strukturiert, d.h. es gibt fachliche Slices abgeleitet von den Aggregates der Anwendung (Studie,

[6] https://12factor.net/.
[7] AngularJS: https://angularjs.org/; Angular Material: https://material.angularjs.org/latest/; Spring Boot: https://spring.io/projects/spring-boot, Cloudfoundry: https://www.cloudfoundry.org/.

Erhebung, Instrument, Frage, Datensatz, Variable, Publikation) und technische Schichten (Web bzw. REST, Service, Repository).[8]

Die Containerinstanzen nutzen die folgenden externen Dienste: Als primärer Datenspeicher, welcher eine normalisierte Form der Aggregates persistiert, wird MongoDB verwendet. Elasticsearch dient als sekundärer Datenspeicher, welcher eine stark denormalisierte Form der Aggregates indiziert, um eine performante Suche auch über die Aggregatsgrenzen hinweg zu erlauben. Der Server nutzt SendGrid, um den Benutzer*innen HTML-E-Mails senden zu können. Diese Mails werden mit Thymeleaf (eine Template Engine für HTML Dokumente) erstellt. Der Server nutzt zudem RabbitMQ als Message Broker um (administrative) Nachrichten via WebSockets an alle Benutzer*innen zu senden. Weiterhin wird die XML-API von da|ra genutzt, um bei jedem Release Digital Object Identifier für die Studien zu registrieren. Um ein da|ra-konformes XML aus den Metadaten des MDM zu erzeugen, wird Freemarker verwendet.[9]

Aufgrund der fachlichen Unsicherheiten zu Beginn der Entwicklung, die erst im Verlauf des Projektes schrittweise aufgelöst wurden, konnte die hier beschriebene Architektur zu Beginn des Projektes noch nicht abschließend festgelegt werden. Vielmehr ist sie im Projektverlauf schrittweise in einem evolutionären Prozess entstanden.

Einige zentrale Entscheidungen wurden zu Beginn der Entwicklung getroffen und dann auch nicht mehr revidiert. Aufgrund der geringen Teamgröße und dem gewählten DevOps-Ansatz, wonach der Betrieb einer Anwendung vom entwickelnden Team verantwortet werden sollte, wurde entschieden, die Anwendung von Anfang an auf einer Cloud Plattform zu entwickeln und zu betreiben. Dadurch konnten externe Dienste genutzt werden, die flexibel skalierbar sind sowie vergleichsweise schnell gebucht und auch gewechselt werden können. Damit wurde die Flexibilität gewonnen, technologische Entscheidungen unkompliziert auch wieder ändern zu können. Eine Reihe anderer, durchaus auch zentraler Entscheidungen, wie zum Beispiel die Frage

[8] Spring MVC: https://docs.spring.io/spring/docs/current/spring-framework-reference/web.html, Spring Security: https://spring.io/projects/spring-security, Pivotal Web Services: https://run.pivotal.io/.

[9] MongoDB: https://docs.mongodb.com/, Elasticsearch: https://www.elastic.co/de/products/elasticsearch, SendGrid: https://sendgrid.com/, Thymeleaf: https://www.thymeleaf.org/, RabbitMQ: https://www.rabbitmq.com/, Freemarker: https://freemarker.apache.org/.

des verwendeten Datenspeichers (von Elasticsearch über die Kombination von PostgreSQL und Elasticsearch zur Kombination von MongoDB und Elasticsearch), konnte auf dieser Grundlage schnell angepasst werden. Eine weitere zentrale Änderung während der Entwicklung war die Umstellung von serverseitigem auf clientseitiges Rendern.

Im Rückblick lässt sich festhalten, dass die letztendlich ausgewählten Technologien gut geeignet waren, um die benötigten Funktionalitäten des MDM zu implementieren. Ohne die Wahl eines agilen Entwicklungsmodells, welches auch offen für tiefergehende Technologieveränderungen ist, hätte die gute Passung zwischen benötigter Funktionalität und technologischer Architektur allerdings nicht erreicht werden können.

Eine Technologieentscheidung, die aktuell noch diskutiert wird, ist die Auswahl von AngularJS in der Versionsreihe 1. Ende 2015 existierte der Nachfolger Angular in der Versionsreihe 2+, der dann kontinuierlich weiterentwickelt wurde (aktuell Version 7), bereits in einer Beta-Version. Die Unterschiede beider Versionsreihen sind relativ groß, so dass ein jetziger Wechsel von Versionsreihe 1 auf Versionsreihe 2+ sehr aufwändig wäre. Rückblickend wäre es sinnvoll gewesen, direkt auf Versionsreihe 2+ zu setzen; damals wurde das Risiko aufgrund der nicht absehbaren Zukunft von Versionsreihe 2+ allerdings als zu hoch bewertet.

Prozess der Entwicklung des MDM – Lessons Learned

Im Rahmen des FDZ-Aufbaus wurden verschiedene Expert*innen auf unterschiedlichen Ebenen konsultiert. Im wissenschaftlichen Beirat des Aufbauprojektes, welcher mit Übergang des FDZ in den Regelbetrieb in einen FDZ-Beirat überführt wurde, waren bzw. sind Expert*innen aus den Bereichen Hochschul- und Wissenschaftsforschung, Survey-Methodologie, qualitative Sozialforschung, Informatik, Forschungsdateninfrastruktur und FDZ-Prozesse vertreten. Weiterhin konnte das FDZ in der Aufbauphase in hohem Maße von der Beratung durch Mitarbeiter*innen und Leitungen anderer FDZ profitieren. Rückblickend lässt sich sagen, dass solch verschiedenartige fachliche Expertise nicht nur für das FDZ-Projekt im Ganzen sondern auch für die MDM-Entwicklung im Konkreten in stärkerem Ausmaß benötigt worden wäre.

Eingeplant war, dass für den Entwicklungsprozess des MDM eine tiefgreifende Expertise zu den im Forschungsfeld besonders relevanten Datentypen (quantitative Befragungsdaten und qualitative Interviewdaten) benötigt werden würde, das benötigte zeitliche Ausmaß dieser Expertise war in der Projektplanung allerdings unterschätzt worden. Mit dem Projektleiter, der im Rahmen des Scrum-Entwicklungsprozesses auch die Rolle des Product Owner übernahm, war eine Person an der Entwicklung beteiligt, die über ausgeprägte Erfahrungen im Bereich der quantitativen empirischen Sozialforschung verfügte. Die zeitlichen Ressourcen für die Übernahme der Rolle als Domänenexperte waren aufgrund der anderen Rollen allerdings sehr begrenzt. Daher füllte zusätzlich ein sozialwissenschaftliches Teammitglied die Rolle des Domänenexperten für die empirische Sozialforschung aus. Doch auch dieser Mitarbeiter konnte durch seine zweite Rolle als Softwaretester und seine dritte Rolle als Koordinator der sozialwissenschaftlichen Aspekte der Metadatenimportprozesse Domänenexpertise nicht im notwendigen zeitlichen Umfang einbringen. Im Laufe des Projektes wurden daher studentische und wissenschaftliche Hilfskräfte eingestellt, um beim Testen der Software und der Erstellung der Metadatenimportprozesse zu unterstützen. Auf diese Weise wurden Ressourcen für die Einbringung von Domänenexpertise frei. Ideal wäre es gewesen, auch zusätzliche Domänenexpertise aus anderen Bereichen, wie zum Beispiel der qualitativen Sozialforschung oder zu FDZ-Prozessen, eng und mit deutlich höheren Ressourcen in die Entwicklung des MDM einbinden zu können. Teammitglieder mit entsprechender Expertise waren aber durch andere zentrale Projektaufgaben gebunden. Ein Modell für zukünftige Projekte könnte sein, die benötigte Expertise phasenweise extern einzukaufen, ähnlich, wie es im informationstechnischen Bereich mit der Beauftragung von Beratern für spezifische Technologien praktiziert wird. Darüber hinaus wäre es hilfreich gewesen, auch zukünftige Datennutzer*innen als Tester des MDM in die Entwicklung mit einzubinden.

Think big, start small, iterate quickly gehören zu den Leitlinien agiler Softwareentwicklung und waren auch Leitlinien bei der Entwicklung des MDM, für die das agile Entwicklungsmodell Scrum eingesetzt wurde. Ausgehend von der Vision des MDM konnten zu Beginn des Projektes manche Anforderungen

an das zu entwickelnde System schon sehr klar, andere aber nur sehr grob spezifiziert werden. Die konkreten Prozesse, die durch die Software unterstützt werden sollten, waren zu Entwicklungsbeginn noch nicht definiert. Als Entwicklungsmodell kamen daher nur agile Modelle in Frage, die Änderungen oder Ergänzungen von Anforderungen sowie Wechsel von Technologien im laufenden Projekt als essentielle Bestandteile des Entwicklungsprozesses beinhalten. Scrum beinhaltet entsprechende Rollenverteilungen, Prozesse und Kommunikationsformate, die für die Entwicklung des MDM angepasst wurden. Insbesondere das Kommunikationsformat der Retrospektive, in dem explizit eine teambildende Metaebene adressiert wird, war sehr hilfreich, um die fachlichen und technologischen Unsicherheiten der Entwicklung nicht auf Arbeits- und Teamprozesse übergreifen zu lassen.

Die Entwicklungszyklen (Sprints) des MDM waren in der Regel zwei Wochen lang. Die Implementierung folgte dem Grundsatz, Funktionalität zunächst basal umzusetzen und sie, sofern nötig, in späteren Iterationen zu erweitern und auszudifferenzieren. Jeder Sprint erzeugte dabei eine neuere Version des MDM, von denen die meisten auch (mit entsprechend eingeschränkter Funktionalität) produktiv hätten eingesetzt werden können. Der Produktivbetrieb des MDM wurde trotzdem erst zum Ende der ersten Aufbauphase aufgenommen. Zum einen, da die Aufbereitung und Dokumentation der Forschungsdatensätze, deren Metadaten das MDM verwaltet, erst parallel zur Entwicklung des MDM stattfand. Zum anderen, da die Implementierung von Erweiterungen für ein produktives System deutlich aufwändiger ist, was in der zweiten Aufbauphase des MDM aktuell auch zu beobachten ist. Daher wurde schon zu Beginn der Projektlaufzeit trotz der früheren Verfügbarkeit lauffähiger MDM-Versionen entschieden, so spät wie möglich in den Produktivbetrieb zu gehen.

Für die Entwicklung des MDM wurde bei GitHub ein offenes Entwicklungsrepository angelegt, womit die notwendige Versionskontrolle für die Entwicklung gewährleistet war. Auch wird dort die technische Dokumentation des Systems zur Verfügung gestellt. Zudem wurde und wird die Issue-Funktionalität von GitHub für die Einrichtung des in Scrum vorgesehenen Product Backlog eingesetzt. Das Product Backlog umfasste die schon

bekannten Anforderungen an das zu entwickelnde System. Entstehen neue Anforderungen, so werden diese dem Product Backlog als einzelne Backlog Items hinzugefügt. Aus der Menge der Backlog Items wurde für jeden Sprint eine Untermenge an Items ausgewählt, die im jeweiligen Sprint implementiert werden sollten. Die Entscheidung, welche Items für den nächsten Sprint ausgewählt werden, traf das Entwicklungsteam auf Grundlage der vom Product Owner („Auftraggeber“) vorgenommenen Priorisierung der Backlog Items. Für die MDM-Entwicklung übernahm der Projektleiter des FDZ-Aufbaus die Rolle des Product Owner. Dies führte phasenweise zu einer zu geringen Verfügbarkeit der Product Owner-Rolle, wenn Aufgaben der Projektleiter-Rolle Priorität erhalten mussten.

Auch an anderer Stelle kam es aufgrund der gegebenen Personalressourcen zu Engpässen und auch zu Rollenkonflikten. So übernahm der Softwarearchitekt, der gleichzeitig der einzige Senior Entwickler des Projektes war und daher neben eigenen Implementierungsarbeiten auch Qualitätssicherungsmaßnahmen übernehmen musste, zusätzlich die Rolle des Scrum Masters. In letzterer Funktion war er für die organisatorische Durchführung des Scrum Prozesses zuständig. Rückblickend wäre es wünschenswert gewesen, die verschiedenen Rollen auf mehrere Personen verteilen zu können. Angesichts der allgemeinen Schwierigkeit für Projekte des Öffentlichen Dienstes, informationstechnisches Personal zu rekrutieren, war das Entwicklungsteam mit einem Senior und zwei Junior Entwicklern in der ersten Aufbauphase aber schon vergleichsweise gut ausgestattet. Seit Oktober 2018 unterstützen Freelancer das Entwicklerteam. Deren Rekrutierung war notwendig geworden, da beide Junior Entwickler das Projekt gegen Ende der ersten Aufbauphase verließen und es nicht gelang, neue Softwareentwickler*innen mit den notwendigen Kenntnissen zu gewinnen. Da die Beschäftigung von Freelancern im ursprünglichen Projektantrag nicht vorgesehen war, war der Prozess ihrer Rekrutierung organisatorisch schwierig und langwierig. Eine der wichtigsten Erkenntnisse aus der MDM-Entwicklung ist daher auch, für zukünftige softwaretechnische Entwicklungsprojekte von Anfang an die Freelancer-Option mit einzuplanen.

Am Ende jedes Sprints wurde das gesamte (interdisziplinäre) Projektteam des FDZ-Aufbaus in einem institutionalisierten Kommunikationsformat des Scrum-Prozesses (Review) über die neuen oder veränderten Funktionalitäten des MDM informiert. Hauptzweck der Reviews war das Einholen von fachlichem Feedback der nicht-informationstechnischen Projektmitglieder zu den vorgenommenen Erweiterungen bzw. Änderungen. Das funktionierte im Projekt am Anfang kaum, da die sozialwissenschaftlichen und die informationstechnischen Arbeitsbereiche des Projektes von den Teammitgliedern beider Bereiche als stark getrennt empfunden wurden. Die Erkenntnis, dass die im FDZ befindlichen Forschungsdatensätze und das MDM, das ihre Metadaten verwaltet, eine Einheit bilden, reifte bei den sozialwissenschaftlichen Teammitgliedern mit Fortschreiten der Benutzungsoberfläche des MDM und bei den informationstechnischen Teammitgliedern mit Verwendung der „echten“ Metadaten anstelle von Testdaten. Im Rückblick hätte der Entwicklungs-, aber auch der Teambildungsprozess hier beschleunigt werden können, indem von Seiten der Projektleitung den Teammitgliedern schon zu Beginn des Projektes viel expliziter die Vision des MDM und der Zusammenhang zwischen technischer Entwicklung des MDM und fachlicher Aufbereitung der Forschungsdaten verdeutlicht worden wäre. Mittlerweile existiert (auch für neue Teammitglieder) diese empfundene Trennlinie zwischen den Arbeitsbereichen nicht mehr, auch weil durch MDM-Funktionalitäten wie Warenkorb zur Bestellung der Forschungsdaten und Implementierung der forschungsprozessbegleitenden Dokumentation, die Verbindung von Forschungsdaten, Metadaten und technischem System zu ihrer Verwaltung deutlich sichtbar ist.

Fazit

Zwingend erforderlich für den erfolgreichen Abschluss eines solchen Projektes sind unserer Ansicht nach drei Faktoren: erstens, eine klare Vision des zu entwickelnden Systems; zweitens eine hohe softwaretechnische Expertise und drittens der kontinuierliche Einbezug fachlichen Domänenwissens in die technische Entwicklung. Unter den gegebenen Rahmenbedingungen wäre die Entwicklung des MDM zudem ohne Einsatz eines agilen Entwicklungsmodells nicht möglich gewesen. Wenn es gelungen wäre, noch kontinuierlicher und

auch noch heterogenere Domänenexpertise in den direkten Entwicklungsprozess einzubeziehen, hätte manche Funktionalität des MDM weniger oft überarbeitet werden müssen. Mit einem expliziteren Bewusstsein für die Interdisziplinarität des Teams und für die daraus folgenden unterschiedlichen Sprachen und Fachkulturen, wären manche Teambildungs- aber auch Arbeitsprozesse einfacher gewesen. Wie dargestellt, lassen sich außer diesen beiden noch weitere Bereiche des Projektablaufs finden, in denen sich im Rückblick Optimierungspotentiale zeigen. Insgesamt ist die Entwicklung des MDM aber sowohl im Ergebnis als auch im Entwicklungsprozess auch unter diesen nicht optimierten Gegebenheiten als sehr positiv zu bewerten.

Das Projekt FORDATIS – Aufbau einer Forschungsdateninfrastruktur in der Fraunhofer-Gesellschaft

Andrea Wuchner

Fraunhofer IRB Stuttgart

Einführung

Forschungsdaten und Forschungsdatenmanagement spielen in der Wissenschaft eine immer bedeutendere Rolle. 2010 verabschiedete die Allianz der Wissenschaftsorganisationen die Grundsätze zum Umgang mit Forschungsdaten. Diese erkennen Forschungsdaten als einen „Grundpfeiler wissenschaftlicher Erkenntnis“ an und legen die Grundsätze des Umgangs mit ihnen fest (Allianz der Wissenschaftsorganisationen, 2010). Zunächst war es ausreichend, die Regeln der guten wissenschaftlichen Praxis (DFG 2013) zu befolgen und Daten mindestens 10 Jahre aufzubewahren. Inzwischen wird gefordert, dass Daten nicht nur konserviert, sondern auch öffentlich zugänglich gemacht werden. So hat die Deutsche Forschungsgemeinschaft im Jahr 2015 Leitlinien zum Umgang mit Forschungsdaten (DFG 2015) veröffentlicht. Das europäische Förderprogramm Horizon 2020 (European Commission 2016) fordert zudem seit 2017 die Teilnahme aller geförderter Projekte am Open Research Data Pilot (European Commission 2016). In beiden Fällen gilt, dass Forschungsdaten aus geförderten Projekten veröffentlicht werden müssen. Auch auf politischer Ebene wird der freie Zugang zu Daten immer stärker eingefordert (RfII 2016). Mit der European Open Science Cloud (European Commission, 2018) und der Nationale Forschungsdateninfrastruktur NFDI (RfII 2016) sollen auf europäischer und nationaler Ebene übergreifende Infrastrukturen entstehen, deren Ausgestaltung noch offen ist.

Auch die wissenschaftlichen Verlage erkennen das Vermarktungspotenzial von Forschungsdaten zunehmend. Durch das Herausbringen eigener Datenjournale, die ausschließlich der Publikation von Forschungsdaten dienen, und durch die Verpflichtung der Datenabgabe bei klassischen Publikationen, versuchen sich die Verlage Rechte an wertvollen Forschungsdaten zu sichern.

Als Reaktion auf all diese Entwicklungen arbeiten viele außeruniversitäre Forschungseinrichtungen und Universitäten an der Entwicklung von

Infrastruktur-Angeboten und Services für das Forschungsdatenmanagement in ihrer Institution. So wurde beispielsweise in der Leibniz-Gemeinschaft ein Projekt zur Konzepterstellung eines institutionellen Forschungsdaten-managements für die Leibniz Universität Hannover ab 2014 durchgeführt (Meyer, Janna, & Soßna, 2017). Innerhalb der Max-Planck-Gesellschaft werden zentrale Services rund um das Forschungsdatenmanagement von Max Planck Digital Library (Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V., 2019) und der Max Planck Computing and Data Facility (Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V., 2013) angeboten. Auch für die Universitäten stellt Forschungsdatenmanagement ein wichtiges Thema dar: Vielfältige Services und Dienstleistungsangebote werden von Rechenzentren und Bibliotheken erbracht. So wurde beispielsweise am Karlsruher Institut für Technologie KIT das Serviceteam Research Data Management gegründet. Dieses bietet Beratung und Dienste, „um in allen Phasen des Forschungszyklus ein effektives Forschungsdatenmanagement gemäß den FAIR-Prinzipien (Findable – Accessible – Interoperable – Reusable) zu fördern. Weiterhin baut das Serviceteam im Rahmen von gemeinsamen Projekten mit Forscherinnen und Forschern auch fachspezifische Informationsinfrastrukturen auf.“ (Karlsruher Institut, kein Datum) Daneben betreiben diverse Universitäten in Deutschland institutionelle Forschungsdaten-Repositorien.

Begleitet werden diese Infrastruktur-Bemühungen von verschiedenen Initiativen zur Information über und zur Standardisierung von Forschungsdatenmanagement. Auf internationaler Ebene etwa durch die Research Data Alliance (Research Data Alliance, kein Datum), deren deutsche Mitglieder sich inzwischen im Verein RDA Deutschland e.V. (RDA Deutschland e.V., kein Datum) engagieren oder auch national durch die DINI-AG/Nestor AG Forschungsdaten (Deutsche Initiative für Netzwerkinformation e. V., kein Datum). Es gibt eine Vielzahl von Workshops und Konferenzen, die das Thema aufgreifen, beispielsweise die eScience-Tage Heidelberg (Universität Heidelberg, 2018), organisiert vom Projekt bwFDM-Info II (KIT Karlsruhe, 2018) unter Beteiligung des Karlsruher Instituts für Technologie, der Universität Konstanz und der Universität Heidelberg. Als Beratungsangebot im deutschen

Raum wird die Webseite forschungsdaten.info betrieben (Universität Konstanz, 2019), die sich an Wissenschaftler richtet. Hinzu kommt die Seite forschungsdaten.org (Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ, 2013), die sich an Infrastruktur-Dienstleister richtet.

Die Fraunhofer-Gesellschaft stellt mit 72 Instituten, 25.000 Wissenschaftlerinnen und Wissenschaftler und einem jährlichen Gesamtbudget von 2,3 Milliarden Euro Europas größte Forschungsorganisation für anwendungsorientierte Forschung dar. Im Rahmen der am Menschen ausgerichteten Forschungsfelder "Gesundheit und Umwelt", "Schutz und Sicherheit", "Mobilität und Transport", "Produktion und Dienstleistung", sowie "Kommunikation und "Wissen" entstehen an den Fraunhofer-Instituten große Mengen heterogener Forschungsdaten – Forschungsdaten, die sich hinsichtlich Erhebungsmethoden, verwendeter Standards und Geräten sowie Format stark unterscheiden. Daneben nimmt die Fraunhofer-Gesellschaft innerhalb der deutschen Wissenschaftslandschaft eine Sonderstellung ein: Sie ist lediglich zu 30% grundfinanziert, 70% des Budgets müssen durch öffentliche Forschungsprojekte sowie durch Kooperationsprojekten mit der Industrie eingeworben werden. Als Folge dessen, spielen Rechte am geistigen Eigentum[1] eine besondere Rolle für die Fraunhofer-Gesellschaft: Es ist also im Einzelfall zu prüfen, welche Daten für eine Veröffentlichung geeignet sind (Küsters & Klages, 2019).

Die qualitätsgesicherte Publikation von Forschungsdaten und die Sicherstellung der Erfüllung der internen und externen Anforderungen wurden von der Fraunhofer-Gesellschaft als wichtiges Themenfeld identifiziert. Es gab daher 2014 erste Überlegungen und Bedarfsermittlungen. Es wurde die Frage geklärt, welche Anforderungen an eine Forschungsdateninfrastruktur innerhalb der Fraunhofer-Gesellschaft bestehen und wie diese unter Einbeziehung aller relevanten Stakeholder aufgebaut werden kann. Ein wichtiger Punkt dabei war die Gestaltung des Zusammenspiels der IT in Fraunhofer-Zentrale und –Instituten, des Fachinformationsmanagements der Institute und der zentralen

[1] engl. IPR = Intellectual Property Rights

Dienste. Zu dem bestand an den Instituten ein erheblicher Informations- und Beratungsbedarf im Umgang mit Forschungsdaten.

Im Oktober 2015 verabschiedete der Vorstand der Fraunhofer-Gesellschaft die Fraunhofer-Open-Access-Strategie 2020. Der Aufbau eines institutionellen Forschungsdatenrepositoriums gehörte zu den grundlegenden Maßnahmen zur Umsetzung der Strategie: „Das bestehende zentrale Open-Access-Repositorium »Fraunhofer-ePrints« der Fraunhofer-Gesellschaft wird um ein Forschungsdaten-Repositorium »Fordatis« ergänzt und als integrierte Infrastruktur bereitgestellt“ (Fraunhofer-Gesellschaft, 2015).

Die Umsetzung war Teil des Responsible Research & Innovation-Ansatzes, den die Fraunhofer-Gesellschaft im Rahmen des EU-Projekt JERRI – Joining Efforts for Responsible Research and Innovation (JERRI - Joining Efforts for Responsible Research & Innovation, 2017) verfolgt. Hier soll durch die tiefgreifende Verankerung „Deep Institutionalization“ und Umsetzung verschiedener Dimensionen[2] in der Fraunhofer-Gesellschaft, darunter auch die Dimension Open Access zu einer verantwortungsbewussten Forschung und Innovation beigetragen werden. In der Dimension Open Access wurde dies durch das Infrastruktur-Projekt “FORDATIS – Forschungsdateninfrastruktur angestrebt, das von Juni 2016 bis Mai 2018 am Stuttgarter Fraunhofer Competence Center »Research Services & Open Science« durchgeführt wurde. Ziele des Projektes waren der Aufbau eines eigenen Forschungsdatenrepositoriums für veröffentlichte und zu veröffentlichende Forschungsdaten der Fraunhofer-Gesellschaft, die Etablierung eines Schulungsangebots für die Fraunhofer-Wissenschaftler und weitere Interessensgruppen zum Thema “Forschungsdatenmanagement” sowie die Einrichtung eines Support-Angebots für das Thema Forschungsdatenmanagement. Eingeleitet wurde das Projekt durch eine Fraunhofer-weite Umfrage zum Thema Forschungsdatenmanagement, die wichtige Fragen in Bezug auf das Thema klären sollte und Einschätzungen zum konkreten Bedarf lieferte.

[2] Die Dimensionen sind Ethik, Gender, Wissenschaft & Ausbildung, öffentliches Engagement und Open Access

Im folgenden Abschnitt wird die Umsetzung des Projekts beschrieben.

Darstellung der Methodik

Um festzustellen, welche Angebote und Bedarfe zum Thema Forschungsdatenmanagement an den Instituten bereits existieren und um ein Bild des Umgangs mit Forschungsdaten der Wissenschaftler innerhalb der Fraunhofer-Gesellschaft zu gewinnen, wurde im März 2017 eine Fraunhofer-weite Umfrage bei IT-Managern[3] , Fachinformationsmanagern[4] und Wissenschaftlern der Fraunhofer-Gesellschaft durchgeführt. Als Grundlage für die Konzeption der Umfrage wurden dafür bereits an anderen Forschungsorganisationen durchgeführte Umfragen, wie beispielsweise die Umfrage der HU-Berlin (Simucovic, Kindling, & Schirmbacher, 2013) und die Umfrage der Universität Marburg (Projekt Forschungsdatenmanagement und -Archivierung der Universität Marburg, 2015) als Grundlage genommen und angepasst. Die Umfrage wurde mit Hilfe eines zentralen Online-Tools durchgeführt mit je einem Abschnitt für Wissenschaftler, IT-Manager und Fachinformationsmanager. Die folgende Grafik gibt Auskunft über die Beteiligung an der Umfrage:

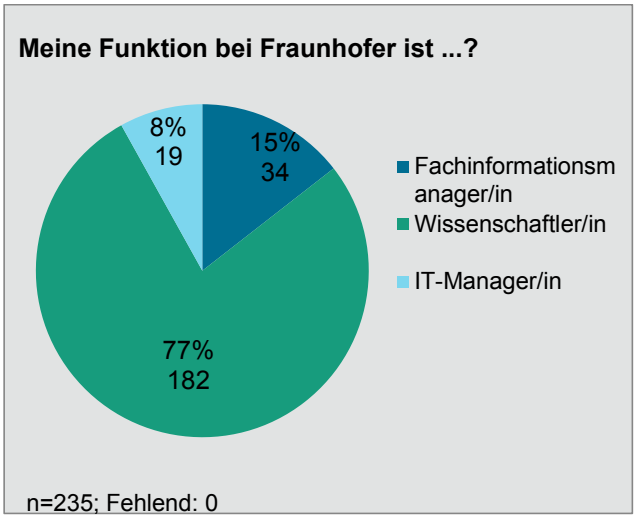


Abbildung 1: Rücklaufquote der Umfrage

[3] IT-Manager: Für die IT verantwortlichen Personen an den Fraunhofer-Instituten

[4] Fachinformationsmanager: Rolle an den Fraunhofer-Instituten, für Fachinformationsversorgung und weitere bibliothekarischen Dienstleistungen am Institut zuständige Person

Forschungsdatenmanagement wurde von allen drei befragten Gruppen als relevantes Thema wahrgenommen. Sowohl IT-Manager als auch Fachinformationsmanager hatten bereits begonnen, sich mit dem Thema auseinander zu setzen. Beide Gruppen wurden von Wissenschaftlern mit Anfragen zum Thema konfrontiert und gaben an, schon Services anbieten zu können.

Für den Umgang mit Forschungsdaten an den Instituten gab es nur in wenigen Fällen festgelegte Prozesse und Verantwortlichkeiten. In den meisten Fällen wurden die Forschungsdaten in Standard-Formaten, wie Microsoft-Formate, csv und jpeg generiert, selbstgeschriebene Software wurde für die Erstellung nur selten verwendet. 17% der befragten Wissenschaftler gaben an, bereits Forschungsdaten veröffentlicht zu haben, 53% wäre nach eigenen Angaben dazu grundsätzlich bereit. 30% der befragten Wissenschaftler, gaben an, Infrastruktur für Speicherung und Veröffentlichung zu benötigen. Ein Schulungs- und Beratungsangebot wurde von allen drei befragten Gruppen gewünscht. Als Antwort auf diese Umfrage, wurden verschiedene Maßnahmen umgesetzt.

Zunächst wurde ergänzend zu bestehenden Strukturen für den Publikationssupport ein zentraler Support für das Thema Forschungsdaten aufgebaut. Dazu wurde eine zentrale Mailing-Adresse eingerichtet, um vor allem die Fachinformationsmanager an den Instituten bei der Beantwortung von Fragen zu unterstützen. Auch die direkte Beratung der Wissenschaftler wurde damit ermöglicht. Bis heute (Stand März 2019) wurden zahlreiche Anfragen per E-Mail und telefonisch bearbeitet. Hierbei handelt es sich um 2nd-Level-Support-Fragen, deren Beantwortung komplexer ist und ausführliche Recherche erfordert. Nachgefragte Themen waren dabei Datenmanagementpläne, Anforderungen der Forschungsförderer, Fragen zur Nutzung von »Fordatis« und zunehmend auch politische Entwicklungen wie die Nationale Forschungsdateninfrastruktur. Der Support berät auch bei der für die Fraunhofer-Gesellschaft wichtigen Fragen, welche Daten veröffentlicht werden sollen.

Ergänzend zum Support wurde 2018 ein eintägiges Schulungs-Angebot für den Themenbereich Forschungsdatenmanagement etabliert, das von Fraunhofer-Instituten gebucht werden kann. Ziel ist es, diese Schulung vor Ort an den einzelnen Fraunhofer-Instituten durchzuführen. Inhaltlich werden dabei Grundlagen des Forschungsdatenmanagements, Anforderungen der Förderorganisationen, rechtliche Rahmenbedingungen, Erstellung von Datenmanagementplänen und Veröffentlichungen von Forschungsdaten behandelt. Bislang wurde die Veranstaltung 10-mal an verschiedenen Fraunhofer-Standorten durchgeführt und hat 140 Teilnehmer erreicht.

Ein weiteres Ziel des FORDATIS-Projekts war der Aufbau einer Fraunhofer-weiten Infrastruktur zur Veröffentlichung von Forschungsdaten. Dafür sollte die bibliografische Datenbank »Fraunhofer-Publica« mit ihrem Volltext-Repositorium »Fraunhofer-ePrints« um ein Forschungsdatenrepositorium »Fordatis« für veröffentlichte Forschungsdaten erweitert werden.

Aus Projektantrag und Umfrage ließen sich folgende Hauptanforderungen ableiten:

- Eindeutige Referenzierbarkeit aller publizierten Datensätze: Die Forschungsdaten werden über einen eindeutigen Identifier referenziert. Damit sind sie jederzeit eindeutig auffindbar, auch wenn sich der physikalische Speicherort der Daten ändert.
- Login: Alle Wissenschaftler bei Fraunhofer können sich am System anmelden und dieses benutzen. Ein manuelles Account-Management soll verhindert werden.
- Verknüpfung zwischen Forschungsdatenpublikationen und anderen Outputformaten und Ressourcen: »Fordatis« ermöglicht die Verknüpfung zwischen Forschungsdaten und anderen Ressourcen wie wissenschaftlicher Software und Publikationen.
- FAIR-Compliance: »Fordatis« ermöglicht die Veröffentlichung von Forschungsdaten gemäß den FAIR-Prinzipien (FORCE 11, 2016): Daten werden so veröffentlicht werden, dass sie auffindbar "findable", zugänglich "accessible", interoperabel "interoperable" und nachnutzbar "reusable" sind.
- Einheitliche, standardisierte Erfassung der Metadaten: »Fordatis« bietet allen Fraunhofer-Instituten eine einheitliche, standardisierte Erfassung der Metadaten zu ihren lokal erhobenen heterogenen

Forschungsdatenbeständen. Fraunhofer-spezifische Merkmale, wie zum Beispiel die Zughörigkeit zu Fraunhofer-Verbünden werden abgebildet.

- Sicherung der Datenhoheit: »Fordatis« wird auf Fraunhofer-eigenen Servern betrieben. So wird sichergestellt werden, dass die Hoheit über die Daten und Metadaten bei der Fraunhofer Gesellschaft selbst verbleibt.
- Google-Indexierung der Metadaten: Die in »Fordatis« gespeicherten Metadaten werden von Google indexiert und sind dort auffindbar. Eine Integration in die Google Dataset Search ist zu prüfen.
- Vernetzung mit anderen Repositorien und Portalen: Durch den Einsatz verbreiteter Metadaten-Standards wird die Vernetzung mit anderen Repositorien, Portalen und Aggregatoren wie zum Beispiel OpenAIRE oder Re3data angestrebt. Damit werden die Anforderungen von Förderorganisationen (zum Beispiel der europäischen Kommission durch die Umsetzung des Open Research Data Pilots in Horizon2020) erfüllt. Auch ist die Sichtbarmachung der Daten in den Fachcommunities durch Harvesting durch die entsprechenden Repositories möglich.
- Bindeglied für Forschungsdaten in den internen Fraunhofer-Datenraum: »Fordatis« stellt das Bindeglied für Forschungsdaten in den Fraunhofer-Datenraum dar. Beim Fraunhofer-Datenraum handelt es sich um einen geschlossenen Datenraum, in dem verschiedene Datenarten wie Publikationsdaten, Patentdaten aber auch Forschungsdaten übergreifend aufgefunden und abgefragt werden sollen. Die veröffentlichten Daten können so Bestandteil übergreifender Abfragen sein.
- Qualitätsgesicherte Erfassung: Analog zu den Publikationsdaten in der »Fraunhofer-Publica« wird eine qualitätsgesicherte Erfassung etabliert. Zu veröffentlichende Forschungsdaten durchlaufen einen Workflow, in dem die beschreibenden Metadaten mehrmals überprüft werden.

Zu Projektbeginn Mitte 2016 zeichnete sich bereits ab, dass die »Fraunhofer-Publica«, die bislang mit der proprietären Software »STAR« betrieben wurde, mittelfristig auf eine moderne Software-Architektur migriert werden soll. Vor diesem Hintergrund wurde entschieden, das Forschungsdaten-Repositorium nicht in die bestehende »Publica« zu integrieren, sondern zunächst als eigenständiges System zu realisieren, das mittelfristig mit der erneuerten »Publica« zusammengeführt wird.

Um die Entscheidung für die einzusetzende Software vorzubereiten, wurden die oben beschriebenen Anforderungen zunächst in ein Lastenheft formuliert.

Status der Items

Die Items sollen in verschiedene Status versetzt werden können:

a) Gespeichert: Item ist nur für die eingebende Person bzw. für die Nutzerrolle „Bearbeiter" sichtbar.
b) In Kontrolle: Rolle QM hat das Item zur Kontrolle „In Kontrolle" genommen.
c) Freigegeben: nach erfolgreicher Kontrolle wurde das Item freigegeben. Item ist über die Suche recherchierbar und Metadaten werden in das Set für das Harvesting aufgenommen.
d) In Bearbeitung: müssen noch Ergänzungen im Datensatz vorgenommen, wird das Item von der Rolle QM zum „Bearbeiter" zurückgeschickt. Das Item hat den Status „In Bearbeitung"
e) Gelöscht: Wird ein Freigegebenes Item gelöscht, erhält es den Status „gelöscht", der Metadatensatz wird zusammengedampft und mit dem Vermerk „Dieses Item wurde gelöscht" versehen. Es soll weiterhin über den Handle aufrufbar sein. Items, die nur gespeichert sind, sollen „spurlos" gelöscht werden können.

Bearbeitung von Items

Items sollen von der Rolle Moderator jederzeit bearbeitet werden können. Metadaten sollen verändert werden können, sowie Dateien hochgeladen bzw. entfernt werden. Freigegebene Items sollen nach der Bearbeitung nur in den Status „Freigegeben" versetzt werden können.

DOI-Vergabe

Die DOI soll von der Itemlanganzeige und von der Itemkurzanzeige aus möglich sein. Sie soll je Forschungsdatensatz (= Item) möglich sein. Die DOI soll anschließend automatisch in die Metadaten geschrieben werden. Die DOI-Vergabe soll für Moderatoren und Administratoren für Datensätze im Status „Freigegeben" möglich sein. Während der DOI-Vergabe soll ein Formular angehakt werden, in dem der Mitarbeiter versichert, dass der betreffende Datensatz noch keine DOI hat und dass es sich um ein wissenschaftliches Objekt handelt.

Löschen von Items

Freigegebene Items sollen von der Rolle Moderator gelöscht werden können. Dazu wird der Metadatensatz wird zusammengedampft und mit dem Vermerk „Dieses Item wurde gelöscht" versehen. Es soll weiterhin über den Handle aufrufbar sein.
Gespeicherte Items soll vom Depositor gelöscht werden können. Sie sollen nachhaltig aus dem Repository entfernt werden.

Abbildung 2: Ausschnitt aus dem Lastenheft von Fordatis

Diese wurde durch ein umfangreiches Use-Case-Dokument ergänzt.

Abschnitt	Inhalt
Bezeichner	F-1-5
Name	Einloggen eines Nutzers (Fraunhofer-Wissenschaftler, Data-Curator)
Autoren	wua
Priorität	5
Kritikalität	5
Quelle	wua
Verantwortlicher	wua
Kurzbeschreibung	Der Nutzer loggt sich in Fordatis ein.
Auslösendes Ereignis	Der Nutzer möchte entsprechend seiner Berechtigung in Fordatis arbeiten.
Akteure	Nutzer, Fordatis
Vorbedingung	Account des Nutzers ist bereits angelegt.
Nachbedingung	Der Nutzer ist in Fordatis eingeloggt.
Ergebnis	Login des Nutzers
Hauptszenario	1.) Nutzer ruft Fordatis auf. 2.) Nutzer klickt auf die Schaltfläche „Login“. Die Loginmaske öffnet sich. 3.) Der Nutzer gibt seine Login-Daten ein und klickt auf die Schaltfläche „Login“ 4.) Der Bereich „Meine Datensätze“ mit Schaltflächen für die weiteren den Rechten entsprechenden möglichen Aktionen erscheint. Der Nutzer ist eingeloggt.
Alternativszenarien	
Ausnahmeszenarien	Fordatis ist nicht erreichbar. Der Login funktioniert nicht.
Qualitäten	

Abbildung 3: Ausschnitt aus dem Use-Case-Dokument für Fordatis

Der nächste Schritt war die Auswahl einer geeigneten Software zur Realisierung von »Fordatis«. Folgende Kriterien waren dafür entscheidend:

- Kompatibilität mit der erneuerten »Fraunhofer-Publica« bzw. »Fraunhofer-ePrints« und weiteren Systemen wie dem »Fraunhofer-FIS« (Küsters & Klages, 2019): Zum Projektstart war es abzusehen, dass die bibliografische Nachweisdatenbank »Fraunhofer-Publica« und

das Volltext-Repositorium »Fraunhofer-ePrints« von der proprietären Software STAR abgelöst werden. Bei der einzusetzenden Software für »Fordatis« sollte es sich um eine Software-Lösung handeln, mit der beide Systeme betrieben werden können, damit Synergie-Effekte und Know-How für beide Systeme genutzt werden.

- Open-Source-Produkt: Bei der einzusetzenden Software sollte es sich um ein Open-Source-Produkt handeln, um die Nachteile einer proprietären Software nicht mehr in Kauf nehmen zu müssen. Darunter fallen Lizenz-Gebühren. Auch kann von Weiterentwicklungen einer großen Open-Source-Community profitiert werden. Die Fraunhofer-Gesellschaft als öffentliche Einrichtung kann eigene Entwicklungen ebenfalls weitergeben.
- Große Nutzercommunity: Um die Vorteile eines Open-Source-Produkts möglichst gut auszunutzen, sollte einen Open-Source-Software mit möglichst großer Community und einer großen Anzahl an erfolgreich betriebenen Installationen eingesetzt werden.

Die Entscheidung fiel nach Sichtung von Invenio und Fedora zu Gunsten der Software DSpace. Diese wird in der Version DSpace 6.3 JSPUI eingesetzt.

Um die Umsetzung des Repositoriums möglichst effizient und mit dem nötigen technischen Know-How durchzuführen, wurde das Vorhaben von einem externen Dienstleister unterstützt. Insgesamt fanden drei Workshops zur Spezifizierung von Funktionalitäten und Metadatenschema und zur Erarbeitung eines Layouts statt, das mit dem Fraunhofer-Corporate-Design konform ist. Die Implementierung wurde im September 2018 begonnen und dauert bis zum heutigen Tag an.

Die oben beschriebenen Anforderungen wurden folgendermaßen realisiert:

- Eindeutige Referenzierbarkeit der Forschungsdaten: Hierfür sollen Digital Object Identifier (DOIs) eingesetzt werden. Diese werden von der TIB-Hannover bereitgestellt.
- Login: Um für alle Wissenschaftler der Fraunhofer-Gesellschaft einen Account mit Rechten zur Dateneingabe bereit zu stellen, wird die Anbindung an den Fraunhofer-Verzeichnisdienst realisiert.
- Verknüpfung zwischen Forschungsdaten und anderen Ressourcen: Die Verknüpfung von Forschungsdaten mit anderen Ressourcen wie Publikationen oder Software erfolgt über ein Metadatenfeld, in das der

Unique Identifier der zu verknüpfenden Ressource eingetragen wird. Die Art der Beziehung kann über ein Auswahlmenü spezifiziert werden.

- FAIR-Compliance: Vom Fraunhofer FIT (Beyan, et al., 2018) wurde eine Evaluierung der FAIR-Compliance von »Fordatis« durchgeführt. Hierfür wurde die FAIR-Metrik (Wilkinson, 2018) der FAIRMetrics Group (Fair Metrics Group , 2018) eingesetzt. Die Kriterien konnten erfüllt werden. Zusätzlicher Handlungsbedarf besteht hinsichtlich eines Konzepts für die Langlebigkeit von Metadaten sowie die Einbindung Community-spezifischer Standards.
- Einheitliche, standardisierte Erfassung der Metadaten: Das Application Profile von »Fordatis« basiert auf dem Dublin-Core-Standard (Dublin Core Metadata Initiative, 2013). Es bietet eine standardisierte, einheitliche Erfassung aller Forschungsdaten. Gleichzeitig werden Fraunhofer-spezifische Merkmale, wie zum Beispiel die Zughörigkeit zu Fraunhofer-Verbünden abgebildet. Hierfür wurden eigene Felder unter dem Namespace fordatis hinzugefügt.
- Sicherung der Datenhoheit: »Fordatis« wird auf Fraunhofer-eigenen Servern, der »Fraunhofer-Cloud« betrieben. So soll sichergestellt werden, dass die Hoheit über die Daten und Metadaten bei der Fraunhofer Gesellschaft selbst verbleibt. Die Server stehen an mehreren Standorten in Deutschland. Zudem sind sie skalierbar.
- Google-Indexierung der Metadaten: Die in »Fordatis« gespeicherten Metadaten sollen von Google indexiert werden und dort auffindbar sein. DSpace bringt einige Werkzeuge mit, um die Inhalte indexieren zu lassen. Hierfür wird die URL von Fordatis zunächst manuell bei Google registriert. Des Weiteren bietet DSpace ein Sitemap-Feature an, das es Suchmaschinen-Crawler erleichtert, die Inhalte zu indexieren. Die Sitemap enthält eine Liste aller Items, Sammlungen und Communities und verhindert, dass Crawler jede Seite einzeln aufrufen müssen. Diese Sitemap ist sowohl im Footer als auch in der robot.txt verlinkt, damit Google optimal Zugriff darauf hat. Ein weiterer Schritt ist die Anpassung der robots.txt-Datei, welche Einstellungen zum Crawler-Zugriff enthält. Hier ist es wichtig, dass die Datei direkt auf höchster Ebene (Top Level) in DSpace eingebunden wird. Des Weiteren muss sichergestellt sein, dass verschiedene URLs indexiert werden können. Um die Indexierbarkeit auch bei komplexen Layout-Anpassungen zu gewährleisten, beinhaltet das <head>-Element der Item-HTML-Seite Item-Metadaten. Weiterhin soll die Weiterleitung von direkten Download-Zugriffen auf Dateien zu den Landing-Pages vermieden werden, ebenso die Generierung von PDF-Deckblättern, welche DSpace anbietet.

- Vernetzung mit anderen Repositorien und Portalen: DSpace enthält eine OAI-PMH-Schnittstelle und setzt intern das Dublin Core-Metadatenformat ein. Dadurch ist das Harvesten der Metadaten in andere Repositorien möglich. Für die Lieferung an OpenAIRE wird ein Mapping von Dublin Core zum Standard DataCite (DataCite, 2014), Version 3.1 eingesetzt.
- Bindeglied für Forschungsdaten in den internen »Fraunhofer-Datenraum«: »Fordatis« soll das Bindeglied für Forschungsdaten in den »Fraunhofer-Datenraum« darstellen. Beim »Fraunhofer-Datenraum« handelt es sich um einen geschlossenen Datenraum, der sich derzeit im Aufbau befindet und auf den alle Fraunhofer-Institute Zugriff haben sollen. Dort werden verschiedene Datenarten, die innerhalb der Fraunhofer-Gesellschaft entstehen, wie Publikationsdaten, Patentdaten aber auch Forschungsdaten, zusammengeführt, so dass sie übergreifend aufgefunden und abgefragt werden können. Die veröffentlichten Daten können so Bestandteil übergreifender Abfragen sein. Die Anbindung an den Datenraum soll über eine RESTAPI gewährleistet werden.
- Qualitätsgesicherte Erfassung: Analog zu den Publikationsdaten in der »Fraunhofer-Publica« soll eine qualitätsgesicherte Erfassung etabliert werden. Dafür durchlaufen die Forschungsdaten innerhalb von »Fordatis« einen Workflow, in dem die beschreibenden Metadaten mehrmals überprüft werden und abschließend freigegeben werden.

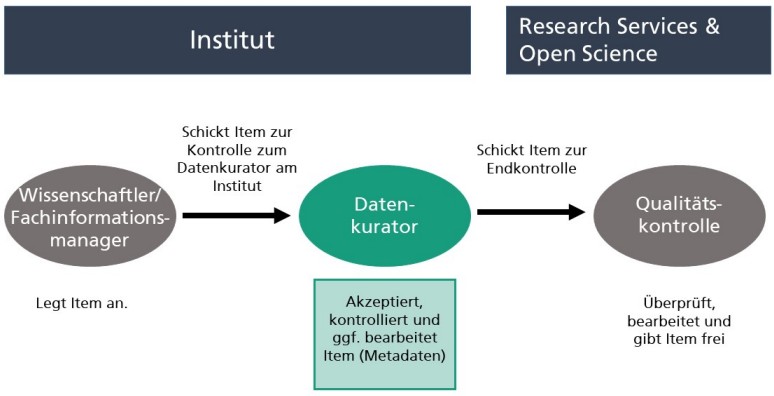


Abbildung 4: Workflow zur Qualitätssicherung

Die Daten werden vom Wissenschaftler bzw. vom Datenkurator in »Fordatis« eingegeben. Anschließend werden sie von einem zuvor bestimmten Datenkurator im Institut, der oftmals in der Institutsbibliothek angesiedelt ist, überprüft, gegebenenfalls bearbeitet und zur zentralen Qualitätskontrolle an die Abteilung »Research Services & Open Science« weitergeleitet. Hier werden die Datensätze nochmal überprüft und dann freigegeben. Erst dann sind sie für die Öffentlichkeit in »Fordatis« sichtbar. Die Kontrolle der Daten erstreckt sich nur auf die Metadaten, hierfür wurde ein eigenes Regelwerk erstellt, das festlegt, welche Daten in welcher Form in welches Feld einzutragen sind. Eine inhaltliche Kontrolle der Forschungsdaten kann nicht geleistet werden, hierfür sind die Wissenschaftlerinnen und Wissenschaftler selbst verantwortlich.

- Rechtliche Compliance: Im Zusammenhang mit Forschungsdaten und dem Betrieb eines Repositoriums gibt es verschiedene rechtliche Voraussetzungen zu beachten. Zunächst stellt sich die Frage, wem die Forschungsdaten „gehören“ und wer über ihre Veröffentlichung bestimmen darf. Hierbei ist zwischen urheberrechtlich gestützten und nicht urheberrechtlich geschützten Forschungsdaten zu unterscheiden. Um in dieser Frage Klarheit zu erhalten, hat die Fraunhofer-Gesellschaft ein rechtliches Gutachten in Auftrag gegeben. Des Weiteren sind für den Betrieb eines Forschungsdaten-Repositoriums die notwendigen, rechtlichen Voraussetzungen zu schaffen: Da über Shibboleth eine Verknüpfung zu personenbezogenen Daten besteht und auch Autorennamen erfasst werden, ist es notwendig ein Verarbeitungsverzeichnis gemäß DSGVO Artikel 30 zu erstellen, sowie eine Genehmigung des Gesamtbetriebsrats einzuholen. Ferner werden gut sichtbare Datenschutzhinweise sowie ein rechtssicheres Impressum benötigt. Um die Forschungsdaten öffentlich bereitzustellen und sie gegebenenfalls später für die Langzeitarchivierung in ein anderes Format umwandeln zu dürfen, ist es notwendig, von den Datenautoren eine Deposit-Lizenz einzuholen. In »Fordatis« wird hierfür ein PDF eingebunden, das elektronisch signiert werden muss und anschließend an die »Abteilung Research Services & Open Science« geschickt werden muss. Anschließend wird die Deposit-Lizenz dort als Bitstream zu den Forschungsdaten mithochgeladen.

Im November 2018 wurde die Fordatis Sandbox einem ersten Nutzerkreis vorgestellt.

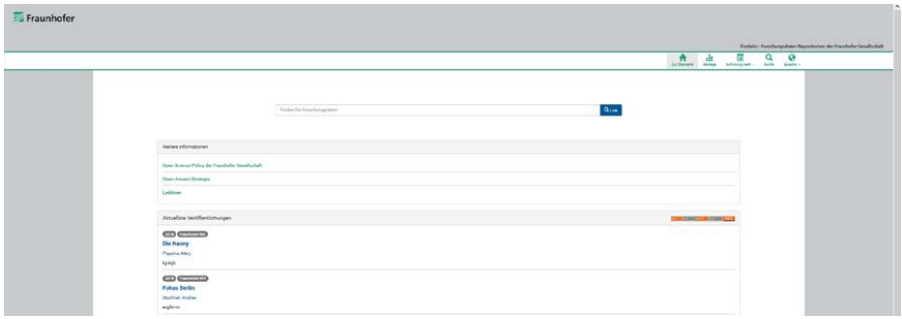


Abbildung 5: Layout von Fordatis

Von der technischen Entwicklung abgesehen, stellt die organisatorische Einbettung des Forschungsdaten-Repositoriums eine weitere Herausforderung dar. An 72 Instituten müssen Erfassungsworkflows etabliert werden, ein Datenkurator muss benannt werden. Sowohl Datenkuratoren als auch Wissenschaftler müssen befähigt werden mit Fordatis zu arbeiten. Um an den Instituten Erfassungsworkflows zu etablieren, wurden die Fachinformationsmanager gezielt angeschrieben und informiert. Damit sollte gewährleistet werden, dass die Rolle des Datenkurators in bestehende Fachinformations-Strukturen an den Instituten integriert wird und Synergieeffekte genutzt werden. Die Wissenschaftler und Fachinformationsmanager werden durch verschiedene Angebote bei ihrer Arbeit mit Fordatis unterstützt: Anleitungen zur Benutzung von Fordatis, Screencasts, Informationen zu den angebotenen Lizenzen. Des Weiteren wurde zur Kommunikation von »Fordatis« für verschiedene Stakeholder (vorrangig innerhalb der Fraunhofer-Gesellschaft) ein Kommunikationskonzept erstellt.

Begleitend zum Go-Live wird die Fraunhofer-Gesellschaft eine Forschungsdaten-Policy veröffentlichen. Diese sieht die Veröffentlichung Forschungsdaten vor, die zur Überprüfbarkeit von Forschungsergebnissen notwendig sind. Um die Interessen von Kooperationspartners aus der Industrie hinsichtlich Geheimhaltung zu berücksichtigten, sind Forschungsdaten, die Geheimhaltungsvereinbarungen unterliegen, von der Veröffentlichungspflicht ausgenommen.

Der Go-Live von »Fordatis« ist für Sommer 2019 geplant.

Diskussion

Mit dem Go-Live wird ein wichtiger Meilenstein des Projekts »FORDATIS« erreicht. Es stehen am Repositorium selbst noch weitere Aufgaben an: Ein

wichtiges Thema ist die Frage der Langzeitarchivierung der Forschungsdaten in »Fordatis«, für die ein Konzept zur Langzeitarchivierung erstellt und umgesetzt werden soll.

Mittelfristig soll die bibliografische Datenbank »Fraunhofer-Publica« auf DSpace CRIS migriert werden. Beide Systeme sollen so zusammengeführt werden, dass Normdaten, wie Projektdaten oder Autorendaten gemeinsam genutzt werden können. Nach außen soll dieser Verbund als »Fraunhofer-Open-Science-Cloud« sichtbar werden und geharvestet werden. Es wird in der Zukunft notwendig sein, »Fordatis« auf zukünftigen DSpace-Versionen zu aktualisieren und Anpassungen an die Anforderungen der Stakeholder vorzunehmen.

Das Projekt »FORDATIS« leistet einen wichtigen Beitrag zur Weiterentwicklung der Bibliothekare an den Institutsbibliotheken bzw. den Fachinformationsmanagern an den Instituten: Mit der Rolle „Datenkurator“ eröffnet sich ein in die Zukunft gerichtetes Betätigungsfeld, um die bestehenden Services am Institut mit dem neuen Aufgabenbereich des „Forschungsdatenmanagement“ zu erweitern. Langfristig können in diesem Bereich auch weitere Aufgaben, zum Beispiel die Beratung zu Anforderungen oder Datenmanagementplänen übernommen werden. Einige Institute nutzen diese Gelegenheit, um den „Datenkurator“ fest in der Institutsbibliothek zu verankern.

Das im Projekt entwickelte Beratungs- und Schulungsangebot wird sehr gut angenommen: Es wurden seit 2018 zahlreiche Anfragen zu Forschungsdaten und Forschungsdatenmanagement beantwortet. Häufige Themen waren dabei Datenmanagementpläne und die Anforderungen der Förderorganisationen. Mit den Arbeiten an »Fordatis« wurden auch zunehmend Fragen zum Repositorium gestellt. Das Schulungs-Angebot wird ebenfalls gut angenommen und regelmäßig von den Instituten nachgefragt.

Eine Herausforderung ist es, die Nutzung des Forschungsdatenrepositoriums über die kommenden Jahre im Institutsalltag zu verankern und Mehrwerte für die Wissenschaftler und die Fraunhofer-Gesellschaft zu erzeugen. Zunächst muss das Angebot bei Fraunhofer bekannt gemacht und ausgerollt werden. Da

es in direkter Konkurrenz zu anderen teilweise kostenlosen Repositorien wie Zenodo (CERN, 2019) und Figshare (Figshare, 2019) steht, und die Wissenschaftler im Hinblick auf ihr Veröffentlichungsmedium freie Wahl haben, muss es durch Mehrwerte überzeugen. Auch die Fraunhofer Forschungsdaten-Policy wird zur Nutzung des institutionellen Repositoriums beitragen, da sie die Relevanz des Themas unterstreicht und der Fraunhofer-Gesellschaft die Möglichkeit gibt, an zentraler Stelle den wichtigen Forschungsoutput der Forschungsdaten sichtbar zu machen.

Innerhalb der Fraunhofer-Gesellschaft stellen die Ergebnisse des Projekts »FORDATIS« den Anknüpfungspunkt zu einer weiteren Entwicklung dar: Im Rahmen des BMBF-Projekts HEFE – Heterogene Forschungsdaten im Stadtkontext wurde eine Data Governance konzipiert – ein System aus Regeln, wie mit Forschungsdaten innerhalb der Fraunhofer-Gesellschaft umgegangen werden soll. In diesem Rahmen wurde ein Domänenkonzept entwickelt, das in einem ersten Schritt für verschiedene Datenarten, wie Interviewdaten oder Befragungen Leitlinien und Checklisten erarbeitet. Langfristig möchte das Projekt den Anstoß zu einer digitalen Unterstützung der Data Governance über den gesamten Forschungsprozess hinweg geben. Hierbei soll »Fordatis« zum Veröffentlichen der Forschungsdaten dienen.

Im Hinblick auf eine Fortführung und Fortentwicklung der Forschungsdateninfrastruktur sind zwei Aspekte von besonderer Bedeutung: Ein wichtiger Impuls können die Beantwortung der bislang offenen rechtlichen Fragestellungen darstellen, beispielsweise wem die Forschungsdaten „gehören“ bzw. wer für die Qualität und die Weiterverwendung Forschungsdaten „haftet“ darstellen. Zum anderen sind die Entwicklungen der European Open Science Cloud und der Nationalen Forschungsdateninfrastruktur (NFDI) zu beobachten. »Fordatis« soll so weiterentwickelt werden, dass es an beide Infrastrukturen anschlussfähig ist.

Um die Arbeiten des Projekts fortzuführen, wurde der Betrieb des Repositoriums und der Service-Angebote zunächst bis Ende 2020 verstetigt. Das Projekt FORDATIS stellt im Kontext der Open Science einen wichtigen Schritt für die Umsetzung von Open Data in der Fraunhofer-Gesellschaft dar.

Daneben werden in der Fraunhofer-Gesellschaft Open Access sowie in ersten Ansätzen Open Software gefördert. Besonders im Hinblick auf das Spannungsfeld zwischen Forschung und Wirtschaft, in dem sich die Fraunhofer-Gesellschaft befindet, müssen innerhalb der Fraunhofer-Gesellschaft jeweils passende Umsetzungsstrategien entwickelt werden. (Küsters & Klages, 2019).

Literatur

Beyan, O., Wuchner, A., Eisengräber-Pabst, D., Quix, C., Zaschke, C., & Schumacher, O. (2018). Research Data in the Fraunhofer Digital Project : Creating a FAIR Research Data Infrastructure and Culture.

Büttner, S., Rümpel, S., & Hobohm, H.-C. (2011). Informationswissenschaftler im Forschungsdatenmanagement. In S. Büttner, H.-C. Hobohm, & L. Müller, Handbuch Forschungsdatenmanagement (S. 2010 - 2018). Bad Honnef: Bock + Herchen Verlag.

CERN. (2019). Zenodo. Abgerufen am 01. 04 2019 von https://www.zenodo.org/

DataCite. (2014). DataCite Standard 3.1. Von https://schema.datacite.org/meta/kernel-3.1/ abgerufen

Deutsche Forschungsgemeinschaft. (2013). Sicherung Guter Wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“, . Von http://doi.org/10.1002/9783527679188.oth1 abgerufen

Deutsche Forschungsgemeinschaft. (2015). Leitlinie zum Umgang mit Forschungsdaten. Online verfügbar unter http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/ richtlinien_forschungsdaten.pdf, zuletzt geprüft am 13.03.2017.

Deutsche Initiative für Netzwerkinformation e. V. (kein Datum). Webseite der DINI/Nestor AG Forschungsdaten. Abgerufen am 15. 04 2019 von https://dini.de/ag/dininestor-ag-forschungsdaten/

Dublin Core Metadata Initiative . (2014). Dublin Core Metadata Terms. Von http://www.dublincore.org/specifications/dublin-core/dcmi-terms/ abgerufen

Dublin Core Metadata Initiative. (2013). Dublin Core Element Set. Abgerufen am 28. 03 2019 von http://dublincore.org/documents/dces/

European Commission. (2016). Guidelines on FAIR Data Management in Horizon 2020. Abgerufen am 31. 03 2019 von http://ec.europa.eu/research/participants/ data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission. (2016). Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 . Abgerufen am 31. 03 2019 von http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi

European Commission. (2018). Implementation Roadmap for the European Science Cloud. Von https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf#view=fit&pagemode=none abgerufen

Fair Metrics Group . (2018). Von FAIR Metrics Org: http://fairmetrics.org/about.html abgerufen

Figshare. (2019). Figshare. Abgerufen am 01. 04 2019 von https://figshare.com/

FORCE 11. (2016). The fair data principles. Abgerufen am 31. 03 2019 von https://www.force11.org/group/fairgroup/fairprinciples

Fraunhofer-Gesellschaft. (2015). Fraunhofer Open-Access-Strategie. Abgerufen am 19. Februar 2019 von ttps://www.openaccess.fraunhofer.de/de/open-access-strategie.html

Hapke, H. (2016). Data Librarian: Das moderne Berufsbild. b.i.t online, S. 159 - 164.

Hauck, R., Kaps, R., Krojanski, H. G., Meyer, A., Neumann, J., & Soßna, V. (2016). Der Umgang mit Forschungsdaten an der Leibniz Universität Hannover: Auswertung einer Umfrage und ergänzender Interviews 2015/16. Abgerufen am 31. 03 2019 von https://doi.org/10.15488/265

Heidelberg, R. d. (Letzter Zugriff 26.03.2019). Webseite der E-Science-Tage Heidelberg. Von https://www.e-science-tage.de/ abgerufen

Heidelberg, U. (2017). E-Science-Tage Heidelberg. Abgerufen am 27. 03 2019 von https://www.e-science-tage.de

Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ. (2013). Forschungsdaten.org. Abgerufen am 27. 03 2019 von http://www.forschungsdaten.org

JERRI - Joining Efforts for Responsible Research & Innovation. (2017). Abgerufen am 19. Februar 2019 von https://www.jerri-project.eu/jerri/index.php

Karlsruher Institut für Technologie. (kein Datum). Research Data Management (RDM) am KIT. Abgerufen am 15. 04 2019 von http://www.rdm.kit.edu/

Karlsruher Institut. (kein Datum). Webseite des Karlsruher Institut für Technologie. Abgerufen am 15. 04 2019 von http://www.kit.edu/forschen/13557.php

KIT Karlsruhe. (2018). Projekt bwFDM-Info II. Abgerufen am 27. 03 2019 von https://bwfdm.scc.kit.edu/bwFDM-Info.php

Küsters, U., & Klages, T. (2019). Fostering Open Science@Fraunhofer. Procedia computer science(146), S. 39 -52. doi:10.1016/j.procs.2019.01.078

Leibniz Universität Hannover. (2011). Open-Access-Resolution. Abgerufen am 31. 03 2019 von https://www.uni-hannover.de/fileadmin/luh/content/webredaktion/universitaet/ziele/openaccess-resolution.pdf

Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. (2013). Webseite der Max Planck Computing & Data Facility. Abgerufen am 27. 03 2019 von https://www.mpcdf.mpg.de

Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V. (2019). Webseite der Max Planck Digital Library. Abgerufen am 27. 03 2019 von https://www.mpdl.mpg.de/

Meyer, A., Janna, N., & Soßna, V. (2017). Service durch Kompetenzbündelung - Das institutionelle Konzept zum Forschungsdatenmanagement der Leibniz Universität Hannover. Abgerufen am 2019. 03 31 von https://doi.org/10.11588/heibooks.285.377

Projekt Forschungsdatenmanagement und -Archivierung der Universität Marburg. (2015). Forschungsdatenmanagement an der Philipps-Universität Marburg. Die Ergebnisse der Umfrage zum Forschungsdatenmanagement im November 2014. Von https://10.17192/es2015.0019 abgerufen

RDA Deutschland e.V. (kein Datum). RDA Deutschland. Abgerufen am 31. 03 2019 von https://www.rda-deutschland.de/

Research Data Alliance. (kein Datum). Webseite der Research Data Alliance. Abgerufen am 15. 04 2019 von https://www.rd-alliance.org/

RfII - Rat für Informationsinfrastrukturen. (kein Datum). Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Abgerufen am 31. 03 2019 von urn:nbn:de:101:1-201606229098

Simucovic, E., Kindling, M., & Schirmbacher, P. (2013). Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin. Umfragebericht, Version 1.0. Von urn:nbn:de:kobv:11-100213001 abgerufen

Swan, A., & Brown, S. (2008). Skills, Role & Career Structure of Data Scientists. Report to the JISC.

Universität Heidelberg. (2018). EScience-Tage Heidelberg. Abgerufen am 27. 03 2019 von https://www.e-science-tage.de

Universität Konstanz. (2019). Forschungsdaten.info. Abgerufen am 27. 03 2019 von https://www.forschungsdaten.info

Wilkinson, M. D. (2018). A design framework and exemplar metrics for FAIR. Scientific Data.

Strategien

Implementierung der FAIR-Prinzipien im Forschungsdaten-management: eine Terminologiebasierte Strategie für die inhaltliche Beschreibung numerischer Faktendatensätze

Giacomo Lanza, Joachim Meier, Thomas Wiedenhöfer[1]
Ulrich Schwardmann[2]

[1]Physikalisch-Technische Bundesanstalt, Referat Q.11 „Wissenschaftliche Bibliotheken“, Braunschweig und Berlin

[2]Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen

Zusammenfassung

In der Open Science-Ökonomie stellen numerische Faktendaten eine große Herausforderung für die praktische Umsetzung der vier FAIR-Prinzipien dar. Dies resultiert einerseits aus einer unüberschaubar großen Anzahl von Datensätzen und andererseits aus einer großen Vielfalt an in unterschiedlichen Disziplinen verwendeten Datenstrukturen. Diese Heterogenität erschwert den Vergleich von Forschungsdatenstrukturen unterschiedlichen Ursprungs, sowie die Festlegung eines einheitlichen Standards zu ihrer Beschreibung mittels Metadaten. Beispielsweise sieht das gebräuchliche DataCite-Metadatenschema keine Felder für eine detaillierte Beschreibung zusätzlich zur Angabe frei zu vergebender – und damit unkontrollierter – Schlagworte vor. Vor diesem Hintergrund ist bereits die erste Stufe der FAIR-Prinzipien, die Auffindbarkeit (*Findability*), nur unzureichend zu realisieren. Zielgerichtetes, feingranulares Suchen und präzises Finden auf Datenrepositorien-übergreifender Ebene ist aktuell nicht möglich.

Zur Lösung des beschriebenen Problems ist es sinnvoll, bei typischen Eigenschaften numerischer Faktendaten anzusetzen. Diese Faktendaten sind gekennzeichnet durch Größen, Maßeinheiten, numerische Wertebereiche, Rollen (z.B. Messgröße, Messvariable, Messparameter) und, bei quantitativ bewerteter Zuverlässigkeit der Faktendaten, zusätzlich die Messunsicherheitsangabe und sofern bekannt das zugrundeliegende Messunsicherheitsmodell. Als Module eines aufzubauenden „Metrology Terminology Directory“ (MTD) werden in abgegrenzten Namensräumen kontrollierte Vokabulare für Messgrößen, Maßeinheiten, Messverfahren und verschiedene Charakteristiken von Messobjekten mehrsprachig entwickelt und jeweils über spezifische „Persistent Identifiers“ in einer sogenannten Data

Type Registry sprachübergreifend adressierbar gemacht. In neu entwickelten Faktendaten-spezifisch strukturierten Metadatenmodulen dienen diese Vokabulare zur Beschreibung der Faktendatensätze. Auf diese Weise werden die wesentlichen Eigenschaften numerischer Faktendatensätze für komplexes Suchen und Finden mittels geeigneter Retrieval-Werkzeuge zugänglich gemacht.

Durch Implementierung von Metadatenschema-Modulen seitens der Hersteller von digitalen Messgeräten, digitalen Sensoren oder Messverarbeitungs-Software, könnte zukünftig erreicht werden, dass eine vereinheitlichte Metadatenbeschreibung schon bei der erstmaligen analog/digital-Wandlung von Messdaten begonnen und über die weiteren Schritte der Messdatenverarbeitung „halbautomatisch“ sukzessive angereichert werden kann. Hierdurch würde nicht nur die Dokumentationsarbeit erleichtert, sondern es bestünde auch die Möglichkeit, die FAIR-Prinzipien vollständig umzusetzen.

Eine Pilotrealisierung der MTD und ausgewählter Metadaten-Module werden im Vortrag vorgestellt.

Abstract

Within the Open Science economy, a major challenge for the practical application of the FAIR principles is represented by numerical factual data. This is a result of the overwhelming big quantity of data sets, as well as of the big variety of used data structures currently used in the different disciplines. This heterogeneity makes it difficult to compare data structures of different origins, as well as the establishment of unique standards for their description through metadata: for instance, the common metadata schema DataCite doesn't include any fields for the detailed description of the study object, other than the inclusion of free-text – thus non controlled – keywords. Under these conditions there are limited possibilities to apply even the first of the FAIR principles, Findability: a directed advanced search of data over a plurality of repositories is currently not feasible.

We are seeking a solution to this problem in relation to the typical characteristics of numerical factual data, which are commonly described by reporting quantity names, units, numerical ranges, roles (measured quantity,

variable or parameter) and some estimate of the data reliability, such as the value uncertainty and the uncertainty model. Within a newly defined "Metrology Terminology Directory" (MTD), we are developing a collection of multilingual controlled vocabularies for physical quantities, measuring units, experimental techniques and selected information about research objects; the correct designation of each item takes place in a language-independent manner via a persistent identifier. These vocabularies are then applied within an *ad hoc* defined metadata module for numerical data for a thorough description of a data set. That way, the relevant features of numeric factual data are made accessible for a complex search and finding with suitable retrieval tools.

If the proposed metadata module is adopted and implemented by the producers of digital measuring instruments, digital sensors and software for data analysis, it will be possible in the future to implement a standardised metadata description already at the point of the first analog-to-digital conversion, and then propagate and enrich it somehow automatically during all data transformation steps. This would make documentation work tremendously easier, as well as contribute to the complete realisation of the FAIR principles.

Within the talk a pilot realisation of the MTD will be presented, along with selected metadata modules.

Hintergrund und Problemstellung

Der freie Austausch von Informationen und Wissen ist eine der tragenden Säulen der Wissenschaft. Mit zunehmender Digitalisierung in der Wissenschaft und mit der Verbreitung der Open-Science-Philosophie werden in wachsendem Ausmaß auch Forschungsdaten der Öffentlichkeit zur verfügbar gemacht. Diese Daten fallen in einer Vielzahl offener und proprietärer Formate an. Für deren Beschreibung werden Metadaten verwendet, deren Umfang i. d. R. von den Möglichkeiten der Datenportale oder von Fachgemeinschaften bestimmt wird. Die Qualität der Metadatenerfassung ist dabei abhängig von der fachlichen Expertise und dem Vollständigkeitsanspruch der einzelnen Kuratoren.

Eine Nachnutzung dieser Daten setzt die Fähigkeit voraus, sie mittels geeigneter Suchmaschinen selektiv recherchieren und filtern zu können. Eine feingliedrige, „erweiterte“ Suche bedarf einer gemeinsamen „Sprache“ zwischen den Suchmaschinen und den Datenrepositorien. Das wurde zum Teil erreicht mit der Festlegung standardisierter Metadatenschemata (DataCite) und Protokolle für das *Metadata Harvesting* (OAI-PMH); darüber hinaus wenden einige Fachdisziplinen zusätzliche Metadatenmodule an, die eine feingranulare Beschreibung ermöglichen. In gewissen Fällen wird auch eine Suche im Datenbestand selbst angeboten.

Angesichts der wachsenden Datenmengen und der anfallenden multidisziplinären Fragestellungen wird der Bedarf immer offensichtlicher an einer interoperablen Struktur für die dokumentarische Beschreibung der Daten, die das zuverlässige Finden, Filtern und Vergleichen von Forschungsdaten unterschiedlichen Ursprungs ermöglicht. Die weitgehend KI-unterstützten Methoden der *Big Data*-Analyseund des *Machine Learning* erfordern, dass die Daten nicht nur für Menschen zugänglich, sondern auch maschinenlesbar und -verständlich sind. Dies erfordert eine Auswahl standardisierter, langlebig lesbarer Datenformate, sowie eine eindeutige Kodierung der Metadatenbeschreibungen mit Standardisierung sowohl der Attributfelder als auch deren zulässiger Inhalte (Attributwerte). Die für Forschungsdatenmanagement eingeführten vier FAIR-Prinzipien, Die Daten sollen auffindbar (*Findable*), zugänglich (*Accessible*), interoperabel (*Interoperable*) und wiederverwendbar (*Reusable*) sein, geben die Ziele vor, machen aber keine Angaben für Lösungen dieser nicht trivialen Aufgabenstellung.

Sehr stark automatisierungs- und messtechnischgeprägte Industrieunternehmen sind sich dieser Herausforderungen seit längerem bewusst. Firmeneigene Schemata und Protokolle für die digitale Übertragung von Informationen sind bereits entwickelt und z. B. im Bereich künstlicher Intelligenz im Einsatz.

Lösungsansatz

Forschungsdaten ohne eine standardisierte, feingranulare, den

Suchmaschinen zugängliche Metadaten-Beschreibung sind schwierig zu finden bzw. mit zunehmender Komplexität schwer oder nicht immer eindeutig interpretierbar.

Um die Interoperabilität von Forschungsdaten zu gewährleisten, ist die Einführung einer gemeinsamen Grundlage für die Metadaten-Beschreibung unentbehrlich. Diese entsteht aus (1) einer definierten Palette von Attributen (Metadatenfelder) sowie (2) je einer kontrollierten Liste zulässiger Werte, die eine eindeutige, reproduzierbare und zweifelsfreie Erfassung ermöglicht. Das Ganze sollte international und womöglich fachübergreifend abgestimmt werden. Ein solcher Standard besteht aus folgenden Bausteinen:

- Ein Metadatenschema, also eine hierarchische Anordnung kontrollierter Metadatenfelder mit festgelegten Benennungen und vorgegebenen Regeln (Variabeltyp, Freitext / kontrollierter Text). Erstrebenswert ist hier ein modularer Aufbau, bei dem neben einem gemeinsamen Kernschema (z.B. eine Erweiterung des aktuell verwendeten DataCite-MDS) unterschiedliche fachspezifische oder methodenspezifische Metadaten-Module optional verwendet werden können.
- Mehrere kontrollierte Vokabulare, die für ausgewählte Metadatenfelder eine Liste zulässiger Werte (Terme) zur Verfügung stellen. Je nach Komplexität können die Vokabulare als Thesauri oder Ontologien realisiert werden. Um die Mehrdeutigkeiten der natürlichen Sprache zu überwinden, ist es ratsam, die einzelnen Begriffe mit langlebigen Kennzeichen (Persistent Identifiers) zu versehen. Hierdurch wird die Verfügbarkeit der Vokabulare in mehreren (menschliche) Sprachen unter Beibehaltung der logischen Struktur möglich.

Beispiel: Größen und Einheiten

In den quantitativen Wissenschaften (z. B. Chemie, Physik, Materialwissenschaften und andere technische Wissenschaftsgebiete) spielen numerische Faktendaten eine zentrale Rolle. Die grundlegenden Informationen für deren eindeutige Identifizierung sollten Angaben beinhalten über

- Versuchsobjekt: Probentyp, Proben-ID, chemische Identität, Auftraggeber;
- Datengenerierung: experimentelle Methode bzw. Simulationsprozedur,

Identifikation des Messplatzes, Zeitpunkt, Mitarbeiter…;

- Faktendaten-Merkmale: Größe, Maßeinheit, numerischer Wertebereich, numerische Auflösung, Messunsicherheit, Rolle (Messgröße, Variable oder Parameter).

Im vorgestellten Vorgehen wurde zunächst die Darstellung der Datenwerte behandelt. In der Folge wurde ein atomares Metadatenschema für numerische Faktendaten und Vokabulare für physikalische Größen und Einheiten entwickelt.

Ergebnisse: Metadatenschema

Im Rahmen des innerhalb Horizon 2020 europäisch geförderten Projektes *SmartCom* wurde für industrielle Anwendung ein atomares Metadatenschema (**D-SI**), basierend auf internationalen Richtlinien der Messtechnik, für die Beschreibung numerischer Faktendaten definiert. Das Schema sieht Felder für die Angabe eines Messwerts mit Maßeinheit und Messunsicherheit vor. Die Unsicherheit kann entweder als absolute erweiterte Unsicherheit samt Erweiterungsfaktor, oder als Intervallart samt Intervallgrenzen dargestellt werden; in beiden Fällen wird die Überdeckungswahrscheinlichkeit angegeben. Optionale Angaben dazu sind der Größenname, der Zeitstempel sowie die angenommene Wahrscheinlichkeitsdichteverteilung des Messunsicherheitswertes. Die Qualitätskontrolle legt den größten Wert auf die Angabe der Einheiten, wobei SI-Basiseinheiten ohne Präfixe stark empfohlen werden.

Für die Beschreibung von Forschungsdaten für Open Science-Anwendungen wird dieses Schema erweitert. Für jede im Datensatz auftauchende Größe werden sowohl der Wertebereich (Minimum und Maximum) als auch die Unsicherheit angegeben. Die Benennung der Größe ist verpflichtend und soll möglichst aus einem normierten Vokabular stammen. Zusätzlich wird die Rolle der Größe (Messgröße, Messvariable oder Messparameter) sowie eine wörtliche Erläuterung angegeben.

Das vorgeschlagene Schema bildet die hierarchische Anordnung der relevanten Felder ab. Die korrekte Eingabe der Metadatenwerte wird durch Anwendungsregeln unterstützt. Struktur und Regeln sind bewusst

syntaxneutral angelegt und ermöglichen deshalb die äquivalente Auszeichnung bzw. Export der Metadaten in XML, JSON, YAML oder einem anderen beliebigen Auszeichnungsformat bzw. Notation.

Ergebnisse: kontrollierte Vokabulare

Die entwickelten Vokabulare werden in einem „Metrology Terminology Directory“ gesammelt und sollen in maschinenlesbarem Format (JSON, RDF oder OWL) für die Wissenschaft, die Wirtschaft und die Öffentlichkeit zur Verfügung gestellt werden.
Das Vokabular für physikalische Größen besteht derzeit aus ca. 400 Einträge. Für jeden Eintrag werden die folgenden Attribute angeboten:

- Benennung, derzeit in 18 Sprachen.
- Symbol und Definitionsformel.
- Dimensionen nach BIPM Broschüre; Bezug zur vorgegebenen SI-Einheit.
- Kurzer Definitionstext und Anmerkungen.
- Notation aus einer mehrstufigen Klassifikation.

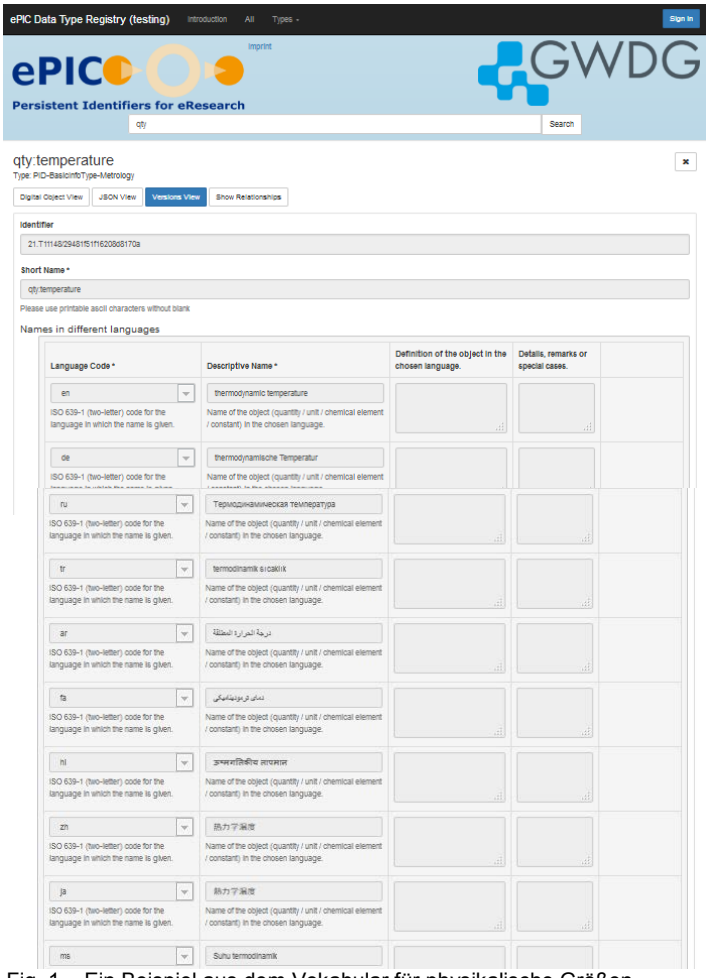


Fig. 1 – Ein Beispiel aus dem Vokabular für physikalische Größen.

Das Vokabular für Naturkonstanten, in Anlehnung an der Liste von CODATA, enthält derzeit 84 Einträge. Für jede Naturkonstante werden die schon bei Größen gelisteten Attribute angegeben, und zusätzlich:

- Der numerische Wert der Naturkonstante mit absoluter und relativer Standardunsicherheit und die Gültigkeitszeitspanne dieses Wertes (basiert auf den CODATA-Ausgaben von 1969, 1973, 1986, 1998, 2002, 2006, 2010, 2014 und 2018).
- Die Korrelationskoeffizienten zu den anderen Naturkonstanten, mit Gültigkeitszeitspanne.

Das Vokabular für Einheiten listet derzeit ca. 100 Einheiten, darunter alle in der BIPM-Broschüre gelisteten (SI-Einheiten und noch akzeptierte nicht-SI-

Einheiten) Einheiten sowie alle bei Messgrößen verwendeten Kombinationen von Einheiten (z. B. J/K, N.m). Für die Einheiten werden folgende Attribute vergeben:

- Bezeichnung, derzeit in 18 Sprachen.
- Symbol in lateinischem, kyrillischen und arabischen Alphabet, sowie LaTeX-Quellcode.
- Definitionsformel mit Umrechnungsfaktor, sofern notwendig.
- Dimensionen nach BIPM-Broschüre und Bezug zu den damit verbundenen Größen und Konstanten.
- Kurzer Definitionstext und Anmerkungen.

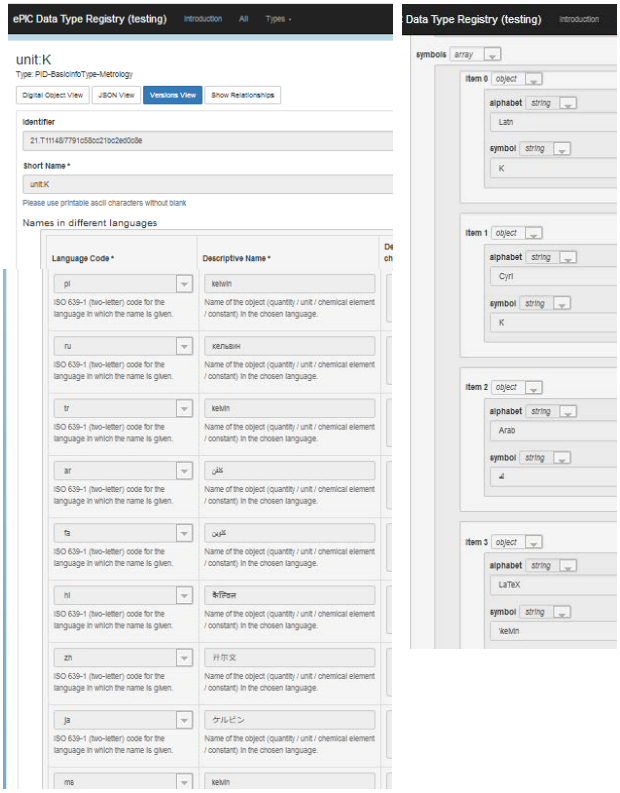


Fig. 2 – Ein Beispiel aus dem Vokabular für Einheiten.

Geplant ist ein viertes Vokabular, das auf Basis des VIM (Internationales Wörterbuch der Metrologie) und des GUM (Guide to the expression of uncertainty in measuremement) die grundlegenden metrologischen Begriffe mehrsprachig und maschinenlesbar darstellen wird. Unter anderen werden

die Grundbegriffe in Bezug auf Messverfahren, Messunsicherheit sowie die Rolle einer Größe aufgeführt.

Beispielanwendung

Das hier eingeführte Schema samt Vokabularen stellt eine mögliche Grundlage dar für die interoperable Erfassung von Forschungsdaten in einem Datenportal oder Repositorium, die eine feingranulare Suche ermöglicht. Dafür soll eine spezialisierte Suchmaschine eingerichtet werden, die nach Größennamen und Wertebereichen indexieren und filtern kann.

Als Beispiel könnte eine Recherche nach aktuellen Erkenntnissen über die Temperaturabhängigkeit verschiedener Materialeigenschaften eines Stoffs unter gewissen Umweltbedingungen ausgeführt werden. In diesem Beispiel sollen alle Datensätze relevant sein, in denen „Graphen“ als Substanz und „Temperatur“ als Größe vorkommen und bei denen Temperaturwerte zwischen vorgegebenen Grenzwerten (z. B. zwischen 200 K und 1000 K) vorliegen.

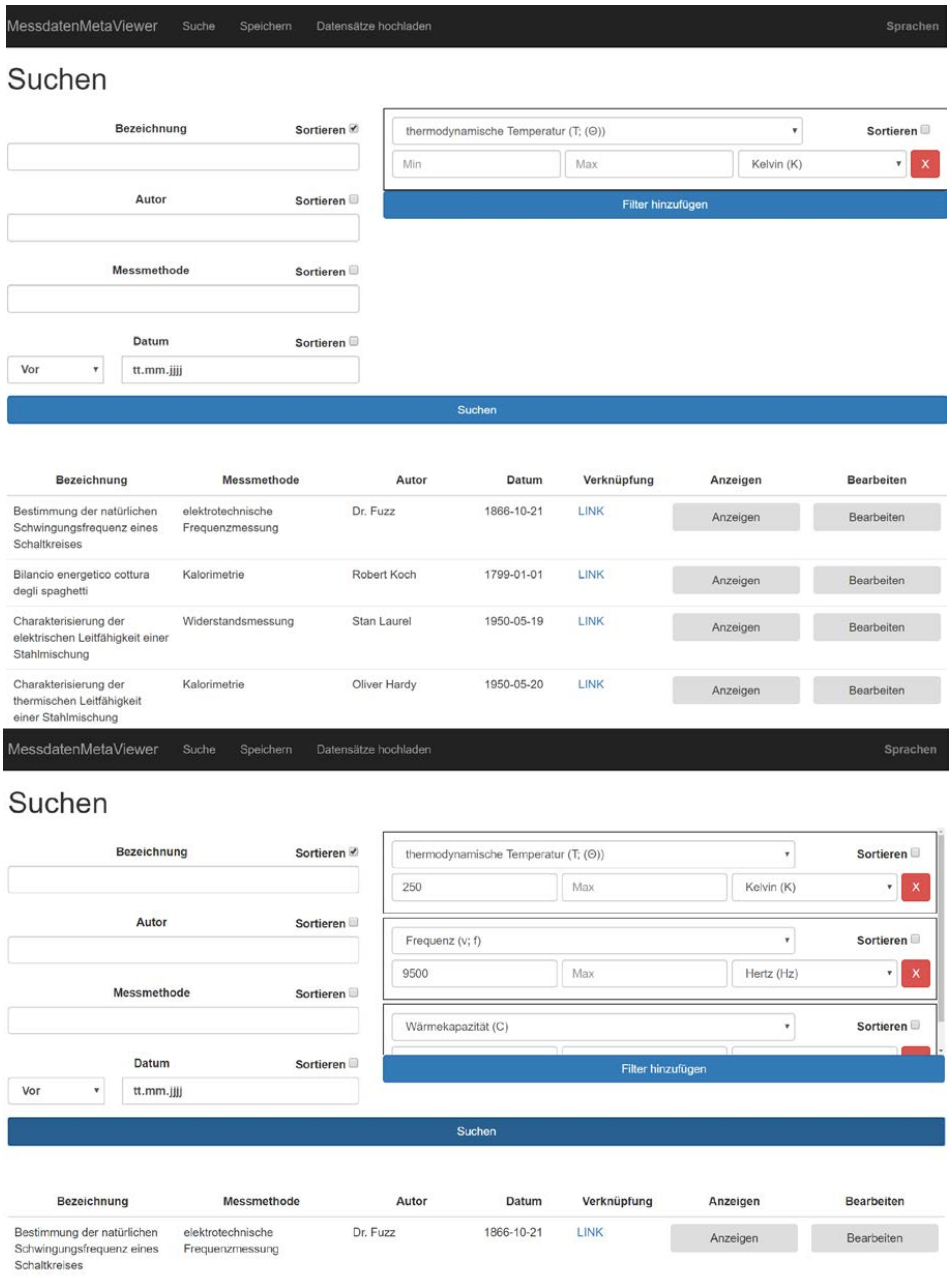


Fig. 3 – Beispielanwendung: eine Suchplattform für Forschungsdaten, die eine erweiterte Suche auf Basis der von uns definierten Felder und der festgelegten Vokabulare ausführt und nach Größen und Wertebereichen filtern kann.

Ausblick

Die Erfassung der Metadaten kann manuell oder automatisch erfolgen. Schon heute werden bei einer softwaregesteuerten Messung wichtige Parameter über den Messablauf von den Messgeräten selbst in den Kopfzeilen der erzeugten Rohdaten oder in zusätzlichen begleitenden Dateien protokolliert. Auch erlaubt bestimmte Software für die wissenschaftliche Datenauswertung die Speicherung von Informationen über enthaltene Variablen und deren numerischen Werte. Die Lesbarkeit und Nachnutzbarkeit dieser Metadaten hängt vom Format und von der verwendeten Syntax ab. Wären sie in einem universellen Format abgespeichert, so könnten sie leicht von anderer Software gelesen, interpretiert und genutzt werden. Somit würde die Metadatenbeschreibung eine Messobjektes entlang der Datenbearbeitungskette stetig angereichert werden und am Ende würde die gesamte Dokumentation über den Messprozess und seine Ergebnisse vorliegen. Metadatenerfassung durch einen Menschen wird so weitgehend vermieden und beschränkt sich daher auf diejenigen Felder, die nicht im Verlauf des Prozesses automatisch beschrieben werden können. Weitere wesentliche Vorteile dieser Vorgehensweise sind, dass fehleranfällige nachträgliche „händische“ Erfassung von Metadaten minimiert wird und die Kooperationsbereitschaft für Forschungsdatenmanagement mit Zielrichtung Open Science beim für die Datengewinnung zuständigen Personal wächst.
Um die Interoperabilität, eine breite Akzeptanz und die Praxistauglichkeit sicherzustellen, wird die Einbeziehung von Messgeräteherstellern und Herstellern wissenschaftlicher Software bei der Festlegung des Formats angestrebt.

Das DIAMANT-Model – Die Einführung eines multiperspektivischen Referenzmodells für die Implementierung von Forschungsdaten-management-Services und -Infrastrukturen

Katarina Blask[1], André Förster[2], Marina Lemaire[3]

[1]Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID)
[2]GESIS – Leibniz-Institut für Sozialwissenschaften
[3]Universität Trier, Servicezentrum eSciences

Abstract

Hochschulen und außeruniversitäre Forschungseinrichtungen stellen zunehmend Infrastrukturen und Services zum adäquaten Umgang mit Forschungsdaten bereit. Dennoch gibt es bislang kein fundiertes Konzept dazu, wie die idealen Bedingungen dieses Implementierungsprozesses aussehen. Mit dem DIAMANT-Modell versuchen wir, diese Lücke zu schließen. DIAMANT ist ein Referenzmodell, das einen Orientierungsrahmen für die Implementierung von Forschungsdatenmanagement (FDM) bereitstellt und dabei am Forschungsprozess ausgerichtet ist. Kern des Modells ist die Etablierung einer zentralen FDM-Informationseinheit, die den Informationsfluss zwischen anderen am FDM beteiligten Organisationseinheiten steuert. Aufgrund der Möglichkeit, Organisationseinheiten auszulagern, bleibt der Implementierungsprozess maximal flexibel und effizient.

Although research institutions take on increased responsibility for providing infrastructures and services around the proper handling of research data, there is no comprehensive framework addressing the ideal conditions of this implementation process. To overcome this gap, we present the DIAMANT model, a reference model aimed at providing an orientation framework for the implementation of research data management (RDM) guided by the research process itself. It builds upon a central RDM information unit controlling the information flow between all other organizational units involved in RDM. Due to the possibility of outsourcing organizational units, the implementation process is maximally flexible and efficient.

1. Einleitung

Hochschulen und außeruniversitäre Forschungseinrichtungen stellen zunehmend Strukturen und Services für den nachhaltigen Umgang mit Forschungsdaten bereit. Bislang gibt es jedoch kein *theoretisches* Grundkonzept zu den Rahmenbedingungen, unter denen der FDM-Prozess[1] maximale Effizienz und Effektivität erreicht. Dies ist zunächst überraschend, da Forschungsinstitutionen in den vergangenen Jahren verschiedene *praktische* Wege gefunden haben, FDM durchzuführen. Es mag jedoch darin begründet sein, dass die Ziele, die mit diesem Prozess assoziiert sind, hauptsächlich aus der Infrastrukturperspektive betrachtet wurden. Die Fokussierung hierauf ist dadurch bedingt, dass FDM als Bedingung für die Erfüllung von Qualitätskriterien gilt, die im Rahmen der Open-Science-Bewegung formuliert wurden (DFG, 2013; Wilkinson et al. 2016) und die Qualität des Prozesses hauptsächlich an der technischen Infrastruktur bemessen.

Ein Aspekt, der bei einer technischen Betrachtung außer Acht gelassen wird, ist, dass es sich beim FDM um einen Informationsverarbeitungsprozess handelt, dessen Implementierung die Integration verschiedener Perspektiven erfordert. Im Einzelnen erfordert die Umsetzung der mit dem FDM verbundenen Funktionen und Aktivitäten die Installation einer Aufbauorganisation[2], die die am FDM beteiligten Organisationseinheiten mit den zwischen ihnen bestehenden Kommunikations- und Weisungsbeziehungen beschreibt (Scheer, 2001). Diese Organisationssicht ist essentiell für die Implementierung eines optimierten FDM-Prozesses.

Im Folgenden skizzieren wir ein Referenzmodell für die Implementierung von FDM-Infrastrukturen und -Services an Forschungseinrichtungen. Dabei geht es zunächst um die Einführung der dem Modell zugrundeliegenden Konzepte. Beginnend mit der Darstellung des Forschungsprozesses als *proximalem* Referenzrahmen des FDM-Prozesses, erfolgt im Anschluss eine Darstellung des Konzeptes der „Architektur integrierter Informationssysteme“ (ARIS; Scheer, 2001), das als Grundlage für die Beschreibung des *distalen* Referenzrahmens vom FDM-Prozess dient. Proximal und distal beschreiben in diesem Kontext das Ausmaß, in dem die jeweiligen Rahmenbedingungen einen direkten (proximalen) beziehungsweise indirekten

[1] Eine prominente Variante für die Veranschaulichung des FDM-Prozesses liefert z.B. das DCC Curation Lifecycle Model Higgins (2008).
[2] Hiermit ist nicht die Schaffung einer speziellen Einrichtung, sondern der Aufbau der Organisation (also z.B. der Universität) hinsichtlich von Governance- bzw. Steuerungsstrukturen für FDM gemeint.

(distalen) Einfluss auf die Umsetzung der verschiedenen FDM-Funktionen haben. Mit Hilfe dieser beiden Referenzrahmen konzipieren wir ein FDM-Referenzmodell, das dazu dienen soll, den „Rohdiamant“ Forschungsdaten zu schleifen und dementsprechend auch so benannt ist: **D**esigning an **I**nformation **A**rchitecture for Data **Man**agement **T**echnologies (DIAMANT).

2 Forschungsdatenmanagement und seine verschiedenen Referenzrahmen

2.1 Der Forschungsprozess als proximaler Referenzrahmen von FDM

Die Notwendigkeit von FDM ergibt sich vor dem Hintergrund zweier im Zuge der Open-Science-Bewegung formulierter Zielzustände: Diese sind die Forderungen nach einer verbesserten Forschungs*ökonomie* und Forschungs*integrität*. Während die Forschungsökonomie vornehmlich auf eine Verbesserung der Wirtschaftlichkeit und somit Effizienz des Forschungsprozesses abzielt (Hinrichs-Krapels & Grant, 2016), wird mit einer Steigerung der Forschungsintegrität eine Verbesserung der Qualität des wissenschaftlichen Arbeitens angestrebt. Entsprechend dem Konzept der Forschungsintegrität (ALLEA, 2017) sollten Forschende bei ihrer Arbeit Reliabilität, Ehrlichkeit, Respekt und Verantwortung für alle mit dem Forschungsprozess assoziierten Aktivitäten sowie alle während des Prozesses erzeugten Forschungsergebnisse gewährleisten.

Der Forschungsprozess und die an ihn angelegten Qualitätskriterien definieren somit die Rahmenbedingungen, welche einen direkten Einfluss auf das FDM haben und somit den proximalen Referenzrahmen für die Implementierung von FDM-Infrastrukturen und –Services aufspannen. Die primäre Funktion von FDM ist demnach die Umsetzung der beiden Zielzustände (Forschungsintegrität und Forschungsökonomie) im Forschungsprozess. FDM erlangt diese Funktionalität mittels des Erzählens einer Geschichte über Forschungsdaten (Surkis & Read, 2015). Im Gegensatz zu der Veröffentlichung eines Zeitschriftenartikels, welcher hauptsächlich die Geschichte zu den Ergebnissen einer wissenschaftlichen Untersuchung erzählt, geht es beim FDM um das Erzählen der ganzen Geschichte in Anlehnung an die jeweiligen Stufen des Forschungsprozesses. Das heißt, dass FDM bereits mit der Entwicklung eines Forschungskonzeptes beginnt und während der Datenerhebung, Analyse, Veröffentlichung und Archivierung kontinuierlich andauert (Vardigan, Heus, & Thomas, 2008). Die Durchführung all dieser Aktivitäten

gewährleistet neben einer erhöhten Forschungsintegrität auch eine gesteigerte Effizienz des Forschungsprozesses, insofern als durch die vollständige Dokumentation Mehrarbeiten vermieden, Fehler umgangen und eine ressourcensparende Nutzung beziehungsweise Nachnutzung von Forschungsdaten gewährleistet werden (ICPSR, 2012; UK Data Archive, 2011).

2.2 Die Architektur integrierter Informationssysteme als distaler Referenzrahmen von FDM

Während der FDM-Prozess mit dem Forschungsprozess und dem darin enthaltenen Daten-Lebenszyklus einen sehr klar umrissenen proximalen Referenzrahmen besitzt, ist der distale Referenzrahmen, das heißt die Informationsarchitektur, innerhalb derer der Prozess abläuft, bislang nur unzureichend konzeptualisiert. Berücksichtigt man, dass FDM an Hochschulen und außeruniversitären Forschungseinrichtungen betrieben werden soll, sollte auch die Konzeptualisierung der Informationsarchitektur innerhalb dieses Rahmens gedacht werden.

Ein Konzept, welches sich in diesem Kontext besonders gut eignet, ist das Konzept der Architektur integrierter Informationssysteme (ARIS; Scheer, 2001). Dieses Konzept eignet sich vor allem deshalb, weil ein zentrales Spannungsfeld beim FDM, nämlich die Vereinbarkeit von IT und Management, der Ausgangspunkt für die Entwicklung dieses Konzeptes war. Im Detail ermöglicht das ARIS-Konzept aufgrund einer „gemeinsamen Sprache für IT und Management“ (Schwickert et al., 2011) eine Modellierung von Geschäftsprozessen, welche den Bedürfnissen aller Akteure gerecht wird. Darüber hinaus erlaubt das ARIS-Konzept eine Reduktion der wahrgenommenen Komplexität des FDM-Prozesses, indem Fragen der Organisation, der Funktionalität der benötigten Daten und der zu erbringenden Leistungen eines Geschäftsprozesses unabhängig voneinander betrachtet werden können. Dies wird ermöglicht durch die Aufteilung der dem ARIS-Konzept zugrundeliegenden Informationsarchitektur in fünf Sichten, von denen drei für die Anwendung von ARIS auf FDM relevant sind.[3] Während die Funktionssicht und die Organisationssicht die Betrachtung einer spezifischen Ebene des Prozesses ermöglichen, erlaubt die Steuerungssicht eine Darstellung der Relationen zwischen den einzelnen Sichten. Jede einzelne dieser Sichten wird durch drei Beschreibungsebenen dargestellt (Schwickert et al., 2011), von

[3] Da die Datensicht und die Leistungssicht im Kontext von FDM nur schwer zu generalisieren sind, beschränken wir uns auf die Funktionssicht, die Organisationssicht und die Steuerungssicht.

denen wir aufgrund des Verallgemeinerungsanspruchs unseres Referenzmodells lediglich die sogenannte Fachkonzeptebene nutzen.[4]

3 Das DIAMANT-Modell

DIAMANT steht für: **D**esigning an **I**nformation **A**rchitecture for Data **MAN**agement **T**echnologies. Das grundsätzliche Ziel des Modells besteht darin, Hochschulen und außeruniversitäre Forschungseinrichtungen zu befähigen, eine integrierte Informationsarchitektur für die optimierte Nutzung von FDM-Technologien zu gestalten. Konkret sollen sie in die Lage versetzt werden, eine Informationsarchitektur zu implementieren, welche Forschenden dabei hilft, ihre Forschungsdatensätze zu „hochkarätigen Diamanten“ zu schleifen. Wie hochkarätig ein Datensatz ist, bemisst sich an seiner Passung mit den im Rahmen der Open-Science-Bewegung geforderten Qualitätskriterien. Um eine optimierte Produktion solcher Hochkaräter an allen Hochschulen und außeruniversitären Forschungseinrichtungen zu ermöglichen, soll nachfolgend ein auf dem ARIS-Konzept beruhendes Fachkonzept für die Modellierung einer Informationsarchitektur skizziert werden, welche ein effektives und effizientes FDM innerhalb des Forschungsprozesses ermöglicht (für eine zusammenfassende Darstellung siehe Abbildung 1).

[4] Die Umsetzung der Fachkonzeptebene über die Modellierung konkreter Daten- und Implementierungsebenen ist durch die Forschungseinrichtungen selbst vorzunehmen.

Abbildung 1: Schematische Darstellung des DIAMANT-Modells mit den innerhalb der Steuerungssicht angedachten Beziehungen zwischen Funktions- und Organisationssicht innerhalb der verschiedenen Stufen des Forschungsprozesses. Aus Gründen der Übersichtlichkeit wurde auf die Darstellung der direkten Beziehungen zwischen FDM-Funktionen und ausführenden Organisationseinheiten (dunkelblau) verzichtet.

3.1 Das Soll-Konzept des FDM-Prozesses aus Funktionssicht

Die Funktionssicht des idealtypischen FDM-Prozesses dient dazu, den Gegenstandsbereich des Prozesses näher zu beschreiben. Demnach werden bei dieser perspektivischen Betrachtung des FDM-Prozesses diejenigen Funktionen beschrieben, welche eine Hochschule oder außeruniversitäre Forschungseinrichtung effizient ausführen sollte, um die übergeordneten Geschäftsziele einer verbesserten Forschungsintegrität und Effizienz des Forschungsprozesses erreichen zu können. Nachfolgend sollen deshalb die Funktionen dargestellt werden, welche innerhalb der verschiedenen Stufen des Forschungsprozesses ausgeführt werden sollten, um die entsprechenden Geschäftsziele erreichen zu können. Um die Erfüllung der skizzierten

idealtypischen FDM-Funktionen zu erleichtern, werden zudem Aktivitätenlisten (*To-Do*-Listen) formuliert. Die Dokumentation dieser Aktivitätenlisten erfolgt in Form von Entscheidungstabellen, um die entsprechenden Bedingungen darzustellen, unter denen die verschiedenen Aktivtäten auszuführen sind (King, 1967). Innerhalb der Entscheidungstabellen werden zum einen die Bedingungen beschrieben, welche für die Erfüllung einer Kompetenz gegeben sein müssen und zum anderen die Aktivitäten, welche bei Nichterfüllung einer Bedingung auszuführen sind. Ob eine Aktivität ausgeführt (gekennzeichnet durch „x“) werden muss oder nicht (gekennzeichnet durch „-“), wird dabei durch die auf die jeweilige Forschungseinrichtung zutreffende Entscheidungsregel (R1…Rn) festgelegt. Die Entscheidungsregeln beschreiben dabei jeweils, welche der relevanten Bedingungen für die Forschungseinrichtung zutreffen („J“ = ja), und welche nicht („N“ = nein). Dergestalt bestimmen sie die auszuführenden Aktivitäten im unteren Tabellenteil.[5]

Der FDM-Prozess startet typischerweise mit der Konzeptualisierungsphase des Forschungsprozesses. Bei der Entwicklung eines neuen Forschungskonzeptes bestehen die primären FDM-Funktionen in der Informationsrecherche (siehe Tabelle 1), der Methodenauswahl (siehe Tabelle 2) sowie der Auswahl und Festlegung auf einen adäquaten Förderrahmen (siehe Tabelle 3). Die Erfüllung dieser drei Funktionen stellt grundsätzlich das Fundament für die Beantragung von Drittmitteln für ein jedes Forschungsvorhaben dar und ist damit essentieller Bestandteil der Wertschöpfungskette einer jeden Hochschule oder außeruniversitären Forschungseinrichtung. Entsprechend der Klassifikation von Porter (1985) handelt es sich hierbei um primäre Funktionen beziehungsweise Aktivitäten. Forschungseinrichtungen sollten somit eine Unterstützung dieser drei Funktionen gewährleisten, um über eine gesteigerte Effizienz bei ihrer Ausführung einen Wettbewerbsvorsprung erzielen beziehungsweise halten zu können.

Im Anschluss an die Phase der Konzeptualisierung eines Forschungsprojektes folgen die Phasen der Primärdatenerhebung sowie der Analyse und Auswertung. Die relevanten FDM-Funktionen sind in beiden Phasen eine adäquate technische Unterstützung des FDM-Prozesses sowie eine Beratung bei der Methodenauswahl in der Auswertungsphase (siehe Tabelle 4). Diese haben keinen direkten Einfluss auf die

[5] Bsp. zu Tabelle 1: An einer Forschungseinrichtung ist der Zugang zu aktuellen Informationsrecherchediensten gewährleistet (j), die nötige Informationsrecherchekompetenz ist allerdings noch unzureichend (n). Dies entspricht der Entscheidungsregel R2: Als nötige Aktivität ist die Bereitstellung von Schulungsangeboten vorzusehen, während bei den Informationsrecherchediensten kein Handlungsbedarf besteht.

Forschungsdatenerzeugung, sondern haben einen unterstützenden Charakter und sind somit den sekundären Aktivitäten zugeordnet (Schwickert et al., 2011). Sekundäre Aktivitäten können grundsätzlich ausgelagert werden, ohne Einbußen im Wertschöpfungsprozess zu erleiden.

In den letzten beiden Phasen des Forschungsprozesses, nämlich der Publikationsphase und der Archivierungsphase, geht es darum, Forschungsergebnisse zu veröffentlichen sowie Forschungsdaten zu archivieren und gegebenenfalls zugänglich zu machen (siehe Tabellen 5 und 6). Bei beiden Funktionen handelt es sich wieder eher um sekundäre Funktionen, womit auch hier die Möglichkeit einer Auslagerung der mit ihnen assoziierten Aktivitäten besteht. Wenn Unterstützungsfunktionen ausgelagert werden, ist jedoch darauf zu achten, dass entsprechende Informationen zu den externen Anbietern und ihren Produkten (z.B. auf der eigenen Homepage) bereitgestellt werden, damit Forschende einen schnellen Zugang zu relevanten Infrastruktur- und Serviceanbietern erhalten.

Entscheidungstabelle – Informationsrecherche		**Regeln**			
		R1	R2	R3	R4
Bedingungen	Zugang zu aktuellen Informationsrecherchediensten	J	J	N	N
	Informationsrecherchekompetenz	J	N	J	N
Aktivitäten	Zugang zu aktuellen Informationsrecherchediensten einrichten	-	-	x	x
	Schulungsangebote bereitstellen	-	x	-	x

Tabelle 2: Entscheidungstabelle-Informationsrecherche

Entscheidungstabelle – Methodenauswahl		Regeln							
		R1	R2	R3	R4	R5	R6	R7	R8
Bedingungen	Methodenkompetenz	J	J	J	N	J	N	N	N
	Zugang zu relevanter Soft-/Hardware	J	J	N	J	N	J	N	N
	Zugang zu fachspezifischer Methodenberatung	J	N	J	J	N	N	J	N
Aktivitäten	Methodenschulungen bereitstellen	-	-	-	x	-	x	x	x
	Soft-/Hardwarekomponenten bereitstellen	-	-	x	-	x	-	x	x
	Zugang zu fachspezifischer Methodenberatung	-	x	-	-	x	x	-	x

Tabelle 2: Entscheidungstabelle-Methodenauswahl

Entscheidungstabelle – Auswahl und Festlegung eines Förderrahmens		Regeln							
		R1	R2	R3	R4	R5	R6	R7	R8
Bedingungen	Kenntnis über die relevanten rechtlichen und ethischen Rahmenbedingungen	J	J	J	N	J	N	N	N
	Kenntnis über die relevanten fachspezifischen und fächerübergreifenden FDM-Leitlinien	J	J	N	J	N	J	N	N
	Kenntnis über die relevanten wissenschaftlichen Netzwerkstrukturen	J	N	J	J	N	N	J	N
Aktivitäten	Einrichtung einer Rechtsberatung	-	-	-	x	-	x	x	x
	Einrichtung einer Ethikkommission	-	-	-	x	-	x	x	x
	Einrichtung einer Informationsseite über relevante FDM-Leitlinien	-	-	x	-	x	-	x	x
	Einrichtung einer Informationsseite über relevante Drittmittelgeber	-	x	-	-	x	x	-	x

Tabelle 3: Entscheidungstabelle-Auswahl und Festlegung eines Förderrahmens

Entscheidungstabelle – Technische Unterstützung		Regeln							
		R1	R2	R3	R4	R5	R6	R7	R8
Bedingungen	Sichere Speicherung	J	J	J	N	J	N	N	N
	Aufbereitung und Dokumentation der Daten im Lebenszyklus	J	J	N	J	N	J	N	N
	Dauerhafte Auffindbarkeit von Daten	J	N	J	J	N	N	J	N
Aktivitäten	Bereitstellung notwendiger Speicherkapazitäten	-	-	-	x	-	x	x	x
	Zugang zu relevanten Unterstützungstools für die (fachspezifische) Dokumentation von Daten	-	-	x	-	x	-	x	x
	Bereitstellung von Informationen und Zugang zu Fortbildungen zur Handhabung der Dokumentationssoftware	-	-	x	-	x	-	x	x
	Bereitstellung von Infrastrukturen, welche eine dauerhafte Auffindbarkeit von Daten gewährleisten (z.B. Bitstream Preservation, Persistente Identifikatoren)	-	x	-	-	x	x	-	x

Tabelle 4: Entscheidungstabelle-Technische Unterstützung

Entscheidungstabelle – Auswahl von Publikationsorganen		Regeln			
		R1	R2	R3	R4
Bedingungen	Wissen über geeignete (Open-Access-)Publikationsorgane	J	J	N	N
	Zugang zu (Open-Access-)Publikationsorganen	J	N	J	N
Aktivitäten	Bereitstellung von Informationsmaterialien zu relevanten (fachspezifischen) Publikationsorganen für Forschungsdaten und forschungsdatenbezogene Publikationen	-	-	x	x
	Gewährleistung des Zugangs zu (Open-Access-)Publikationsorganen (z.B. Open-Access-Publikationsfonds)	-	x	-	x

Tabelle 5: Entscheidungstabelle-Auswahl von Publikationsorganen

Entscheidungstabelle – Auswahl eines Datenarchivs		Regeln							
		R1	R2	R3	R4	R5	R6	R7	R8
Bedingungen	Wissen über und Zugang zu existierenden Datenzentren und Repositorien	J	J	J	N	J	N	N	N
	Gewährleistung der dauerhaften Auffindbarkeit von Daten	J	J	N	J	N	J	N	N
	Passung zwischen Nutzungsbedingungen und projekt-, fachspezifischen sowie rechtlichen Anforderungen der Daten und Texte	J	N	J	J	N	N	J	N
Aktivitäten	Bereitstellung von Informationen zu relevanten existierenden Datenzentren und Repositorien	-	-	-	x	-	x	x	x
	Unterstützung von Mechanismen zur Gewährleistung der dauerhaften Auffindbarkeit (z.B. Auszeichnung mit persistenten Identifikatoren)	-	-	x	-	x	-	x	x
	Bereitstellung von Unterstützungsangeboten für die Einschätzung der Passung zwischen Nutzungsbedingungen und projekt-, fachspezifischen sowie rechtlichen Anforderungen der Daten und Texte	-	x	-	-	x	x	-	x

Tabelle 6: Entscheidungstabelle-Auswahl eines Datenarchivs

3.2 Das Soll-Konzept des FDM-Prozesses aus Organisationssicht

Die Organisationssicht liefert eine dezidierte Darstellung der Organisationseinheiten und Akteure einer Forschungseinrichtung sowie ihrer Beziehungen und Strukturen. Damit stellt sie auch einen genauen Überblick über die Kommunikations- und Weisungsbeziehungen zwischen den verschiedenen am FDM-Prozess beteiligten Organisationseinheiten bereit. Betrachtet man den FDM-Prozess aus einer Organisationssicht, so lassen sich grundsätzlich acht Organisationseinheiten identifizieren, welche für ein optimiertes FDM an Forschungseinrichtungen von Relevanz sind. Ihre Beteiligung ergibt sich dabei auf Grundlage ihrer Bedeutsamkeit für die Umsetzung der beschriebenen FDM-Funktionen. Relevante Organisationseinheiten in diesem Zusammenhang sind eine Ethikkommission, eine Rechtsberatung, eine Methodenberatung, eine Informationsrechercheeinheit, Soft- und Hardwareanbieter, eine Informationseinheit, Publikationsorgane und ein Datenarchiv. Es sei an dieser Stelle angemerkt, dass es sich bei den Benennungen der Organisationseinheiten lediglich um Typbeschreibungen handelt. Demnach ist die

genaue strukturelle und personelle Ausgestaltung einer Organisationseinheit an dieser Stelle nicht definiert. Diese ist durch die jeweilige Forschungseinrichtung auf Grundlage der ihr zur Verfügung stehenden Ressourcen selbst zu bestimmen.

Um die inhaltlichen Zusammenhänge zwischen den verschiedenen Organisationseinheiten entlang des FDM-Prozesses zu illustrieren, wird im DIAMANT-Modell eine prozessorientierte Sicht auf die dem FDM-Prozess zugrundeliegende Aufbauorganisation (im Sinne einer Organisationsstruktur) eingenommen. Entsprechend dieser Modellierung der für den FDM-Prozess spezifischen Aufbauorganisation sind die drei Organisationseinheiten mit ausschließlich menschlichen Leistungsträgern, das heißt die Rechtsberatung, die Methodenberatung und die Ethikkommission im Zusammenspiel mit der Informationseinheit auf den oberen Hierarchieebenen angesiedelt. Sie bilden demnach die funktionalen Einheiten für die Ausführung der in den frühen Phasen des Forschungsprozesses erforderlichen FDM-Aktivitäten. Die unteren Hierarchieebenen, deren Funktionalität auf dem Zusammenspiel zwischen menschlichen und informationsverarbeitenden Leistungsträgern beruht und deren Attribute maßgeblich durch die oberen Hierarchieebenen mitbestimmt werden, bilden die Soft- und Hardwareanbieter, die Publikationsorgane und das Datenarchiv. Sie korrespondieren mit den späteren Phasen des Forschungsprozesses.

3.3 Das Soll-Konzept des FDM-Prozesses aus Steuerungssicht

Die bisherigen Darstellungen der Funktionssicht und der Organisationssicht zielten vornehmlich darauf ab, die Komplexität der Modellierung eines optimierten FDM-Prozesses zu reduzieren, indem nur eine Perspektive des Prozesses in den Blick genommen wird. Für eine optimierte Implementierung des FDM-Prozesses und damit verbunden die Implementierung relevanter FDM-Infrastrukturen und -Services bedarf es jedoch auch eines grundlegenden Verständnisses über das Zusammenspiel dieser beiden Prozesssichten. Die Verbindung einzelner Prozesssichten erfolgt im Rahmen der ARIS-Prozessmodellierung typischerweise mittels der Steuerungssicht. In der Steuerungssicht werden zum einen die strukturellen Beziehungen zwischen den einzelnen Prozesssichten beschrieben und zum anderen Zustandsänderungen. Um die Generalisierbarkeit des DIAMANT-Modells zu gewährleisten, beschränkt sich die Beschreibung der Steuerungssicht auf die Darlegung der strukturellen Beziehungen

zwischen den FDM-Funktionen und den Organisationseinheiten, welche optimale Rahmenbedingungen für die effektive und effiziente Durchführung der verschiedenen FDM-Aktivitäten bieten sollen.[6] Der Grad der Beteiligung verschiedener Organisationseinheiten an der Ausübung der einzelnen FDM-Funktionen wurde auf Grundlage bestehender Darstellungen in der Literatur sowie den durch Forschende geäußerten Bedarfen (Blask & Förster, 2018) an ein optimiertes FDM modelliert.

Die Darstellung der strukturellen Beziehungen zwischen den FDM-Funktionen und den Organisationseinheiten erfolgt in einer Matrix (siehe Tabelle 7). Innerhalb dieser Matrix werden klare Verantwortungs- und Zuständigkeitsbereiche definiert, sodass für Forschungseinrichtungen eindeutig erkennbar wird, welche Informationsverarbeitungs- beziehungsweise Kommunikationswege für die Realisierung der einzelnen Funktionen vonnöten sind. Zu diesem Zweck werden drei Beziehungstypen zwischen Funktionen und Organisationseinheiten definiert. Der erste Beziehungstyp beschreibt, welche Organisationseinheiten verantwortlich sind für die Initiierung von Unterstützungsmaßnahmen, welche die Umsetzung einer bestimmten FDM-Funktion zum Gegenstand haben. Diese Organisationseinheiten sind weisungsbefugt gegenüber den Organisationseinheiten des zweiten Beziehungstyps, welche als sogenannte Exekutiveinheiten zu bezeichnen sind. Exekutiveinheiten sind an der Bereitstellung und Durchführung der jeweiligen Unterstützungsmaßnahmen beteiligt, sind jedoch nicht verantwortlich für die Initiierung der entsprechenden Ereignisprozessketten. Die Überwachung ihrer Arbeit erfolgt über die verantwortlichen Organisationseinheiten. Diese Form der Supervision ist sinnvoll, um die Effektivität und damit die Genauigkeit und Vollständigkeit der Durchführung der einzelnen FDM-Funktionen zu gewährleisten. Der letzte Beziehungstyp dient als Bindeglied zwischen den funktionsspezifischen Organisationseinheiten. Organisationseinheiten, welche diesem Beziehungstyp angehören, sind über relevante Ergebnisse, welche aus dem Ablauf bestimmter funktionsspezifischer Ereignisprozessketten hervorgegangen sind, zu informieren. Dergestalt sollen Unterstützungsmaßnahmen weiterentwickelt sowie weitere Mehrarbeit reduziert und der FDM-Prozess insgesamt transparenter und flexibler gestaltet werden.

[6] Üblicherweise erfordert ARIS neben der Beschreibung der strukturellen Beziehungen auch die des dynamischen Verhaltens des Systems. Das dynamische Verhalten eines Systems lässt sich jedoch nur unter Kenntnis der konkreten ereignisgesteuerten Prozessketten (EPK) modellieren. Da eine Darstellung solcher EPK nur bei gleichzeitiger Reduktion der Allgemeingültigkeit dieses Referenzmodells erreicht werden kann, wird an dieser Stelle auf eine derartige Darstellung verzichtet.

Organisationseinheiten / FDM-Funktionen	Informations-recherchedienst	Methodenberatung	Ethikkommission	Rechtsberatung	**Informations-einheit**	Soft- und Hardware	Publikationsorgane	Datenarchiv
Informationsrecherche	b				v x			
Methodenauswahl		b			v x	b		
Auswahl und Festlegung eines Förderrahmens			b	b	v x			
technische Unterstützung					v x	b		x
Auswahl von Publikationsorganen					v x		b	
Auswahl eines Datenarchivs					v x			b

Tabelle 7: Darstellung der Beteiligung der verschiedenen Organisationseinheiten an der Initiierung, Weiterentwicklung, Bereitstellung und Durchführung der für die Umsetzung der FDM-Funktionen relevanten Unterstützungsmaßnahmen. Die funktionalen Beziehungen werden wie folgt gekennzeichnet: v = ist verantwortlich, b = stellt bereit und führt durch, x = erhält Ergebnisse

Aufgrund der in Tabelle 7 vorgenommenen Funktionszuordnung und der Darstellung der verschiedenen Beziehungstypen zwischen Funktionen und Organisationseinheiten wird schnell deutlich, dass die Einleitung der FDM-Funktionen in den verschiedenen Stufen des Forschungsprozesses immer über einen Kontakt des Forschenden mit der Informationseinheit initiiert wird. Der FDM-Prozess sowie der Forschungsprozess selbst obliegen somit der Eigenverantwortung des Forschenden. Die zentrale Informationseinheit ist hingegen verantwortlich für die Versorgung des Forschenden mit relevanten Informationsangeboten (z.B. Schulungen, Workshops, oder der Bereitstellung von Informationsmaterialien) sowie gegebenenfalls die Weiterleitung des Anliegens an die relevanten Exekutiveinheiten. Es sei an dieser Stelle angemerkt, dass der Forschende für die ihm in ausgewählten FDM-Prozessen zugewiesenen Verantwortlichkeiten auch die Möglichkeit hat, sich direkt an die relevanten Exekutiveinheiten zu wenden. Dieser direkte Weg bietet sich natürlich vor allem bei an der eigenen Forschungseinrichtung institutionalisierten Organisationseinheiten an. Diese sollten den Forschenden wiederum an die Informationseinheit verweisen, falls sie für das von dem Forschenden vorgebrachte Anliegen nicht zuständig sind beziehungsweise nicht über die notwendigen Mittel

und/oder Kompetenzen verfügen. Werden Unterstützungsangebote hingegen über ausgelagerte Organisationseinheiten angeboten, ist in jedem Fall eine Vermittlung über die Informationseinheit zu empfehlen. Eine ausschließliche Beschränkung der Verantwortlichkeit auf eine Organisationseinheit, nämlich die Informationseinheit, hat dabei vor allem eine Schnittstellenreduktion und damit eine Steigerung der Effizienz des FDM-Prozesses zum Ziel. Des Weiteren kommt man mit einem solchen Prozessmodell der Forderung vieler Forschender nach einem FDM-Service-Angebot *aus einer Hand* nach (Blask & Förster, 2018). Die Effektivität und Effizienz des FDM-Prozesses werden zudem über eine Feedback-Schleife von den ausführenden Organisationseinheiten zurück zu der Informationseinheit gesteigert. Über die Rückmeldung der Ergebnisse, welche aus der Ausübung der verschiedenen FDM-Funktionen resultieren, kann für Forschende das bestehende Informationsangebot fortlaufend erweitert werden.

4. Zusammenfassende Darstellung des DIAMANT-Modells und seiner Implikationen

Das DIAMANT-Modell ist ein FDM-Referenzmodell, welches Hochschulen und außeruniversitären Forschungseinrichtungen einen Orientierungsrahmen für die Implementierung von FDM-Infrastrukturen und -Services geben soll. Dabei soll der durch die Infrastrukturen und Services aufgespannte Rahmen vor allem eine effiziente und effektive Ausführung des FDM-Prozesses ermöglichen. Mit anderen Worten, es sollen optimierte Rahmenbedingungen für die Ausführung FDM-bezogener Aufgaben geschaffen werden, um den Aufwand für eine Integration von FDM in den Arbeitsalltag der Forschenden möglichst gering zu halten.

Durch die unserem Modell immanente Reduktion von Schnittstellen und Insellösungen ebnet es den Weg für ein optimiertes FDM an Forschungseinrichtungen und unterstützt somit auch bei der Etablierung bestehender Infrastrukturen und ihrer Organisation in einer nationalen Forschungsdateninfrastruktur (NFDI; RfII, 2017). Diese stärkere Strukturierung und Organisation bestehender Forschungsdateninfrastrukturen fördert zudem die Umsetzung allgemeiner Qualitätskriterien sowie auch die Entwicklung und Etablierung fachspezifischer Standards.

Zusammengefasst bietet das DIAMANT-Modell eine FDM-Informationsarchitektur, die am Forschungsprozess selbst ausgerichtet ist. Da die FDM-Anforderungen an Infrastruktur- und Servicelandschaften aus den Bedürfnissen der Forschenden entspringen, trägt unser Modell zu einer Optimierung des gesamten FDM-Prozesses bei. Diese Optimierung ist auf die grundlegende Funktion von FDM ausgerichtet: Den „Rohdiamant“ Forschungsdaten zu schleifen.

Referenzen

ALLEA - All European Academies (2017). The European Code of Conduct for Research Integrity. https://www.allea.org/wp-content/uploads/2017/05/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf

Blask, K., & Förster, A. (2018). Problembereiche der FDM-Integration (Daten der 1. Interviewwelle des BMBF-Projekts PODMAN; PODMAN W1): Version 1.0.0 [Data set]. http://doi.org/10.5281/zenodo.1492182

DFG (2013). *Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Denkschrift : Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft"*. Weinheim: Wiley-VCH.

Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, *3*, 134–140. https://doi.org/10.2218/ijdc.v3i1.48

Hinrichs-Krapels, S., & Grant, J. (2016). Exploring the effectiveness, efficiency and equity (3e's) of research and research impact assessment. *Palgrave Communications*, *2*. https://doi.org/10.1057/palcomms.2016.90

ICPSR. (2012). Guide to Social Science Data Preparation and Archiving. http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf

King, P. J. H. (1967). Decision Tables. *The Computer Journal*, *10*, 135–142.

Klump, J., & Ludwig, J. (2013). Forschungsdaten-Management. In H. Neuroth, N. Lossau, A. Rapp, & Heike Neuroth, Norbert Lossau, Andrea Rapp (Hrsg.), *Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft* (257–275). Glückstadt: VWH Verlag Werner Hülsbusch.

Porter, M. E. (1985). *Competitive advantage: creating and sustaining superior performance*. New York: Free Press.

RfII (2017). *Entwicklung von Forschungsdateninfrastrukturen im internationalen Vergleich*. Göttingen.

Scheer, A.-W. (2001). *ARIS - Modellierungsmethoden, Metamodelle, Anwendungen* (4. Aufl.). Berlin: Springer.

Schwickert, A. C., Müller, L., Bodenbender, N., Mader, M., Kirchhof, J., & Blöcher, N. (2011). Geschäftsprozessmodellierung mit ARIS. https://wiwi.uni-

giessen.de/dl/down/open/Schwickert/f1dbb464f72be49391c38d2e5328678cd7653e95e908312925e82e84c835203f7020665f52efdafe51443bc769f387fb/Apap_WI_GI_2011_01_PDF_Download_gesch_tzt.pdf

Surkis, A., & Read, K. (2015). Research data management. *Journal of the Medical Library Association JMLA*, *103*, 154–156. https://doi.org/10.3163/1536-5050.103.3.011

UK Data Archive (2011). Managing and sharing data: Best practice for researchers. http://www.data-archive.ac.uk/media/2894/managingsharing.pdf

Vardigan, M., Heus, P., & Thomas, W. (2008). Data Documentation Initiative: Toward a Standard for the Social Sciences. *The International Journal of Digital Curation*, *1*, 107–113. https://doi.org/10.2218/ijdc.v3i1.45

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018 EP. https://doi.org/10.1038/sdata.2016.18

FD-Strategieentwicklung mit dem RISE-DE Framework – Lessons Learned und weitere Anwendungsszenarien

Boris Jacob, Niklas K. Hartman, Nadin Weiß

UB/ZIM Universität Potsdam FIT

Zusammenfassung
Mit RISE-DE liegt als FDMentor-Projektergebnis der Universität Potsdam ein Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement vor. RISE-DE bietet einen Bewertungsrahmen zur Selbstevaluation und Zielbestimmung und eignet sich als Werkzeug zur Gestaltung einer strukturierten, stakeholder-orientierten Strategieentwicklung für das Forschungsdatenmanagement an Hochschulen und Forschungseinrichtungen.

RISE-DE basiert auf dem lose an Reifegradmodellen orientierten Research Infrastructure Self-Evaluation Framework des Digital Curation Centre (RISE v1.1), wurde aber für den Einsatz in partizipativen Prozessen deutlich überarbeitet sowie inhaltlich an den deutschen Wissenschaftskontext und Entwicklungen in der guten Praxis im FDM angepasst.

Abstract
RISE-DE, an FDMentor project result by the University of Potsdam, is a reference model for strategy processes in institutional research data management. RISE-DE provides a framework for self-evaluation and definition of objectives and is useful as a tool to design a structured, stakeholder-oriented strategy process for research data management in higher education and research institutions.

RISE-DE is based on the Digital Curation Centre's Research Infrastructure Self-Evaluation Framework (RISE v1.1), which is loosely based on capability maturity models. However, RISE-DE has been significantly revised for use in participatory processes and adapted in content to the German scientific context and developments in good practice in RDM.

Einleitung

Die Universität Potsdam ist mit über 20.000 Studierenden die größte Hochschule in Brandenburg. Im Jahr 1991 gegründet, belegt sie den 17. Platz beim 2018 Times Higher

Education Young University Ranking und ist beim 2019 World University Ranking unter den 250 besten Hochschulen (Times Higher Education 2019). Das Thema Forschungsdatenmanagement spielte bisher nur in einzelnen Forschungsbereichen der Universität Potsdam eine Rolle, erreichte aber 2017 mit dem Symposium *Qualitätsentwicklung und Professionalisierung des Forschungsdatenmanagements* (Seckler 2017) eine hochschulweite Dimension, welche durch die Teilnahme im BMBF-finanzierten Verbundprojekt FDMentor weitergeführt wurde.

Fünf Partneruniversitäten aus Berlin und Brandenburg haben sich in FDMentor (Laufzeit vom 01.05.2017 bis zum 30.04.2019) gemeinsam mit der Frage beschäftigt, wie institutionelles Forschungsdatenmanagement schnell aufgebaut sowie breit und nachhaltig an den universitären Zentraleinrichtungen verankert werden kann. Die Universität Potsdam hat sich in Arbeitspaket 1 mit Strategieentwicklung beschäftigt und diesen Prozess direkt durch ein Team der am Projekt beteiligten Zentraleinrichtungen umgesetzt, der Universitätsbibliothek (UB) und dem Zentrum für Informationstechnologie und Medienmanagement (ZIM).

In den letzten Jahren haben universitäre Zentraleinrichtungen vermehrt ausgewiesene Forschungsdatendienste aufgebaut. Diese reichen von Beratung und Training bis zu neuen, umfassend betreuten IT-basierten Diensten. Der Aufbau dieser Dienste folgt allerdings nicht immer einer strategischen Planung. Zur Bedarfsermittlung werden vor allem Online-Befragungen, teils auch qualitative Interviews eingesetzt. Aus methodologischer Sicht und praktischer Erfahrung bestehen Zweifel über die Eignung dieser Verfahren zur Gewinnung der benötigten Informationen. Zudem bleibt ihre Rolle in der Strategieentwicklung meist unklar. Mehr Potenzial, Bedarfsermittlung und Strategie zu verzahnen, können Methoden zur Selbstbewertung bieten. Im Bereich Forschungsdaten wurden zur Selbstbewertung bisher einerseits sehr einfache Werkzeuge nach dem Ampel-Prinzip vorgelegt (LEARN Project 2013), andererseits das Konzept von Reifegradmodellen aus der Softwareentwicklung und dem IT-Portfoliomanagement auf das institutionelle FDM übertragen. Zur strukturierten Selbstbewertung und Zielfestlegung für FD-Dienste stehen mittlerweile eine ganze Reihe von Reifegradmodellen zur Verfügung (Oppenländer et al. 2017). Außerdem wurden Modelle zum quantitativen Benchmarking bestehender Dienste entwickelt (LERU Research Data Working Group 2013).

Entwicklung und Pilotanwendung von RISE-DE

Diese verschiedenen Ansätze wurden beim ersten FDMentor-Workshop *Wo bitte geht's zum FDM: Selbstevaluationsmaterialien zur Unterstützung von Strategieprozessen* (Hartmann & Weiß 2018) mit der Community geteilt und besprochen. In den Diskussionen wurde das lose an Reifegradmodellen orientierte Research Infrastructure Self-Evaluation Framework (Rans & Whyte 2017) des Digital Curation Centre (DCC) als beste Grundlage für eine strukturierte Strategieentwicklung für institutionelle Forschungsdatendienste identifiziert. RISE v1.1 umfasst 10 Themenfelder ('areas'), die sich am Forschungsdaten-Lebenszyklus orientieren (Abb. 1) und in 21 Themen ('capabilities') unterteilt sind.

Meyer (2019) berichtet nach der Anwendung des englischsprachigen RISE-Frameworks an der Universität Hannover von dessen Vorteilen, dazu gehören Zeitersparnis, die Nutzung eines standardisierten Verfahrens und die Möglichkeit sich dazu international austauschen zu können. Beim Prozess der Evaluation ist es außerdem hilfreich sich auf die Reputation des Digital Curation Centre berufen zu können. Es wird aber auch argumentiert, dass eine Anpassung an den deutschen Wissenschaftskontext fehlt und die organisatorische Einbettung der im Research Data Lifecycle angeordneten Module fehlt.

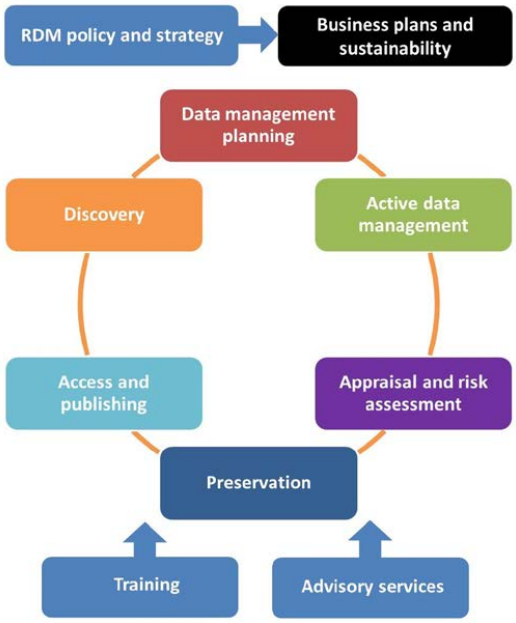


Abb. 1: DCC Research Data Service Model (Rans & Whyte 2017)

Beide Aspekte wurden bei der Verwendung von RISE v1.1 an der Universität Potsdam ebenfalls als Nachteile identifiziert und wie folgt adressiert: RISE v1.1 wurde als RISE-DE ins Deutsche übersetzt und an den deutschsprachigen Hochschulkontext angepasst. Berücksichtigt wurden dabei insbesondere die Empfehlungen der Hochschulrektorenkonferenz (2015) und der League of European Research Universities (LERU Research Data Working Group 2013).

RISE v1.1 wurde verwendet durch die Technische Universiteit Delft (Boehmer 2017) und die Leibniz Universität Hannover (Kaps et al. 2018). Beide verfügten zum Zeitpunkt der Evaluation bereits über etablierte FDM-Services, welche über die eigene Universität hinaus gehende Aufgaben wahrnehmen. Die Selbstevaluation wurde durch diese Teams durchgeführt. Da an der Universität Potsdam aber zentrale FDM-Dienste von Grund auf etabliert werden sollten, erschien es sinnvoll, wichtige universitätsweite Stakeholder von vornherein einzubinden. Zu diesem Zweck wurden die Materialien optimiert für den Einsatz in einem umfassenden, strukturierten und stakeholder-orientierten dialogischen Prozess, der sich an Prinzipien von Total Quality Management-Verfahren (TQM) für den öffentlichen Sektor wie dem Common Assessment Framework (Deutsches CAF-Zentrum 2013) orientiert (Abb. 2, siehe Hartmann, Jacob & Weiß 2019). Solche Prozesse integrieren Elemente der Bedarfsermittlung und Informationsgewinnung, Bewertung bestehender Dienste und Prozesse sowie der Strategieentwicklung.

Zur Selbstevaluation und Zielbestimmung wurde durch die Forschungskommission des Senats eine Arbeitsgruppe bestehend aus Mitgliedern der Fakultäten sowie Verwaltung und Infrastruktur (Vizepräsident für Forschung und wissenschaftlichen Nachwuchs, CIO, UB, ZIM) eingesetzt, welche in regelmäßigen Sitzungen mit RISE-DE arbeitete.

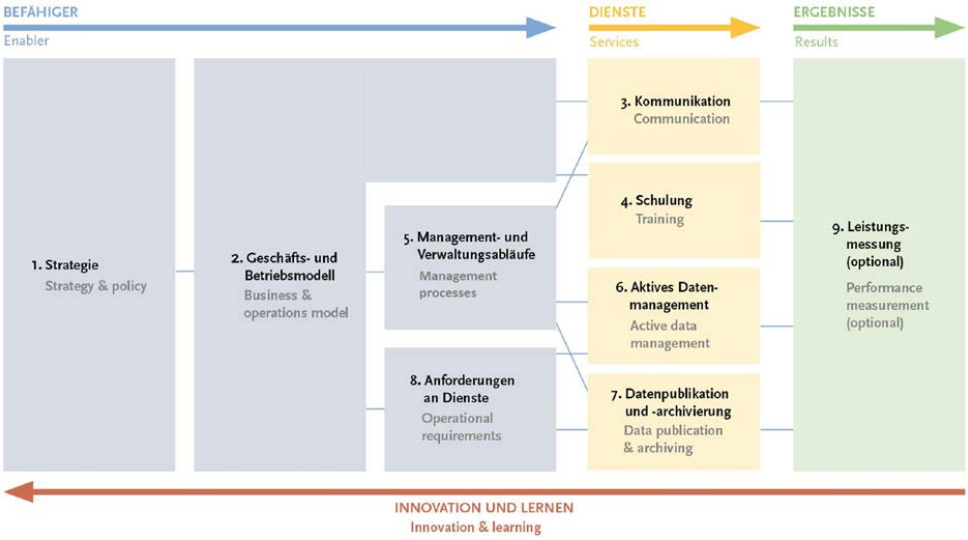


Abb. 2: RISE-DE Research Data Governance & Service Areas

Das deutschsprachige Framework umfasst acht Themenfelder mit insgesamt 25 Themen, welche jeweils mit Hilfe der vierstufigen, vom DCC für RISE v1.1 entwickelten Skala bewertet werden können:

- Stufe 0: Keine (nennenswerte) Aktivität.
- Stufe 1: Minimale Aktivität zum Erreichen externer Anforderungen wie Erhalt der Drittmittelfähigkeit.
- Stufe 2: Engagierte Aktivität am Bedarf der Forschenden an der Einrichtung ausgerichtet.
- Stufe 3: Herausragende Aktivität, national oder international branchenführend.

Ziel der Bewertung sollte nicht sein, eine hohe Stufe zu erreichen, sondern eine, die im Kontext der jeweiligen Institution sinnvoll ist. Die Technische Universiteit Delft (Boehmer 2017) und die Leibniz Universität Hannover (Kaps et al. 2018) haben die Ergebnisse ihrer mit dem englischsprachigen RISE-Framework durchgeführten Selbstevaluation und Zielbestimmung publiziert. Die Ergebnisse der Selbstevaluation und Zielbestimmung an der Universität Potsdam wurden in eine Roadmap eingearbeitet, welche die Schritte zur Etablierung von Diensten zum FDM in den folgenden Jahren beschreibt. Danach wurden eine Forschungsdaten-Strategie und eine Forschungsdaten-Policy erarbeitet, letzte unter Verwendung der “Empfehlungen zur Erstellung institutioneller Forschungsdaten-Policies” (Hiemenz & Kuberek 2018), ebenfalls ein FDMentor-Projektergebnis.

Diskussion

RISE-DE ist in der der Lage, die Entwicklung von Diensten abzubilden und eignet sich daher für den Einsatz an Hochschulen, die beim institutionellen Forschungsdatenmanagement noch ganz am Anfang stehen ebenso, wie für bereits weiter fortgeschrittene Einrichtungen. Es ist für den Hochschulkontext entwickelt, aber auch an das deutsche Wissenschaftssystem angepasst und eignet sich daher auch zur Verwendung in außeruniversitären Forschungs- und Infrastruktureinrichtungen.

Das Thema Leistungsmessung, bisher dargestellt im Ergebnisse-Themenfeld in Abb.2 ist bisher weder in RISE v1.1 noch in RISE-DE weiter ausgearbeitet. Der bisherige FDMentor-Projektverbund plant hierzu Folgearbeiten im Rahmen eines weiteren Verbundprojekts.

RISE-DE ist als Working Paper in der Version 0.9 veröffentlicht, die finale Version wird das Feedback aus der Community berücksichtigen.

Referenzen

Boehmer, Jasmin. 2017. „2017 Self Assessment of Research Data Services and 4TU.Centre for Research Data Services with RISE“. TU Delft. https://openworking.tudl.tudelft.nl/2017/06/28/2017-self-assessment-of-research-data-services-and-4tu-centre-for-research-data-services-with-rise/.

Deutsches CAF-Zentrum. 2013. „Common Assessment Framework – Verbesserung öffentlicher Organisationen durch Selbstbewertung“. Bundesverwaltungsamt. https://www.verwaltung-innovativ.de/SharedDocs/ExterneLinks/DE/Download/CAF_Brosch%C3%BCre_2013.pdf?__blob=publicationFile&v=3.

Hartmann, Niklas K., Boris Jacob & Nadin Weiß. 2019. „RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement“. FDMentor. Universität Potsdam. https://doi.org/10.5281/zenodo.2549344.

Hartmann, Niklas K. & Nadin Weiß. 2018. „Wo bitte geht‘s zum FDM: Selbstevaluationsmaterialien zur Unterstützung von Strategieprozessen“. gehalten auf der Nachnutzbare Strategien und Werkzeuge für das Forschungsdatenmanagement – universitäre Zentraleinrichtungen als Akteure im FDM, Humboldt-Universität zu Berlin. http://www.forschungsdaten.org/images/3/35/FDMentor-Inputworkshop-AP1Strategie_-_public.pdf.

Hiemenz, Bea & Monika Kuberek. 2019. „Strategischer Leitfaden zur Etablierung einer institutionellen Forschungsdaten-Policy. Das Forschungsdaten-Policy-Kit als generischer Baukasten mit Leitfragen und Textbausteinen für Hochschulen in Deutschland“. FDMentor. http://dx.doi.org/10.14279/depositonce-8412.

Hochschulrektorenkonferenz. 2015. *Wie Hochschulleitungen die Entwicklung des Forschungsdatenmanagements steuern können: Orientierungspfade, Handlungsoptionen, Szenarien - Empfehlung der 19. Mitgliederversammlung der HRK am 10. November 2015 in Kiel*. 2016. Bd. 1/2016. Beiträge zur Hochschulpolitik. Bonn https://www.hrk.de/positionen/beschluss/detail/wie-hochschulleitungen-die-entwicklung-des-forschungsdatenmanagements-steuern-koennen-orientierungsp/.

Kaps, Reiko, Anke Krüger, Anneke Meyer, Janna Neumann, Jessika Rücknagel, Volker Soßna & Frauke Ziedorn. 2018. „Research Data Management Services at Leibniz University Hanover : A Self-Assessment by the RDM Service Team Based on the RISE Questionnaire (v1.1)“. Leibniz Universität Hannover. https://doi.org/10.15488/4205.

LEARN Project. 2013. „LEARN RDM Readiness Survey“. http://learn-rdm.eu/en/rdm-readiness-survey/.

LERU Research Data Working Group. 2013. „LERU Roadmap for Research Data“ 14. Advice Paper. https://www.leru.org/files/LERU-Roadmap-for-Research-Data-Full-paper.pdf.

Meyer, Anneke. 2019. „Using RISE for Service Evaluation - Experiences, Findings and Differences“. gehalten auf der 14th International Digital Curation Conference, Melbourne. https://doi.org/10.15488/4205.

Oppenlaender, Jonas, Falko Glöckler, Jana Hoffmann & Claudia Müller-Birn. 2017. „Reifegradmodelle für ein integriertes Forschungsdatenmanagement in multidisziplinären Forschungsorganisationen“. In *Forschungsdaten Managen*, 53–63. Heidelberg. https://publikationen.bibliothek.kit.edu/1000078228/5549805.

Rans, Jonathan & Angus Whyte. 2017. „Using RISE the Research Infrastructure Self-Evaluation Framework“. Version 1.1. Digital Curation Centre. http://www.dcc.ac.uk/sites/default/files/documents/publications/UsingRISE_v1_1.pdf.

Seckler, Robert. 2017. „Grußwort“. gehalten auf der 1. Symposium zum Forschungsdatenmanagement an der Universität Potsdam, Potsdam. https://www.uni-potsdam.de/de/zim/wir-ueber-uns/projekte/fdm-symposium.html.

Times Higher Education. 2019. „World University Rankings“. https://www.timeshighereducation.com/world-university-rankings/2019/world-ranking.

Strukturen und Schnittstellen

Management digitaler Forschungsdaten im akademischen Umfeld – Lessons learned aus der Einführung von RADAR

Kerstin Soltau, Matthias Razum [1]
Dorothea Strecker[2]

[1] FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur
[2] Humboldt-Universität zu Berlin

RADAR: vom DFG-geförderten Projekt zum verstetigten Dienst

Mit RADAR nahm 2017 ein disziplinübergreifender Cloud-Dienst für die langfristige Archivierung und Publikation digitaler Forschungsdaten den Betrieb auf. RADAR wird von FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur betrieben und wendet sich derzeit primär an Hochschulen und außeruniversitäre Forschungseinrichtungen, die keine eigene Forschungsdateninfrastruktur betreiben oder die RADAR ergänzend zu existierenden disziplinspezifischen Angeboten nutzen möchten.

Die Entwicklung von RADAR erfolgte im Rahmen eines DFG-geförderten Projekts (2013-2016) unter Beteiligung von fünf Forschungseinrichtungen aus den Natur- und Informationswissenschaften. (RADAR, 2016) Zu den Projektpartnern zählten neben FIZ Karlsruhe die Technische Informationsbibliothek (TIB), die Fakultät für Chemie und Pharmazie der Ludwig-Maximilians-Universität München (LMU), das Leibniz-Institut für Pflanzenbiochemie Halle (IPB), sowie das Steinbuch Centre for Computing (SCC) des Karlsruher Instituts für Technologie (KIT). Die Kooperation von Informationsinfrastruktureinrichtungen mit Forschenden an Hochschulen und außeruniversitären Forschungseinrichtungen zielte darauf ab, ein bedarfsgerechtes Angebot „aus der wissenschaftlichen Community für die wissenschaftliche Community“ zu ermöglichen. Bereits während der Projektlaufzeit wurde ein nachhaltiges Geschäftsmodell erarbeitet, das nach einer fünfjährigen Anlaufphase ein sich selbst tragendes Dienstangebot sichern soll. Ein wichtiger Aspekt ist dabei der Verzicht auf Projektförderung für den laufenden Betrieb von Anfang an.

Seit Ende der Projektphase wird RADAR von FIZ Karlsruhe betrieben, das als Vertragspartner für die nutzenden Einrichtungen auftritt. Die technische Infrastruktur von RADAR wird vom SCC und dem Zentrum für Informationsdienste und Hochleistungsrechnen (ZIH) der TU Dresden bereitgestellt.

Die Speicherung der Forschungsdaten erfolgt in drei Kopien an geographisch getrennten Standorten in den Rechenzentren des SCC und des ZIH. Als deutsches DataCite-Mitglied registriert die TIB Hannover DOIs für Datensätze, die über RADAR publiziert werden. Der komplette RADAR-Service und seine Infrastruktur unterliegen damit deutschem Recht.

Seit Projektbeginn zielte RADAR insbesondere auf das institutionelle Forschungsdatenmanagement (FDM) im „Long Tail" der Forschung ab, also auf Disziplinen, die mit kleineren Datenmengen umgehen und für die bisher keine disziplinspezifischen Forschungsdateninfrastrukturen zur Verfügung stehen. Als disziplinübergreifender Dienst steht RADAR somit vor der Herausforderung, fachspezifische Anforderungen in einer generischen Infrastruktur zusammenzuführen. Bei der Sammlung von Bedarfsanforderungen verfolgte das Projektteam verschiedene Ansätze.

Die Konzeption während der Projektphase wurde durch Anforderungsanalysen begleitet, die die bedarfsgerechte, standardkonforme und dem Stand der Technik entsprechende Entwicklung der Infrastruktur und die Definition der zentralen Diensteigenschaften, wie z. B. Dienstleistungsmodell, Rechtekonzept Datenmanagement oder Systemarchitektur sicher stellten. (Hahn, Neumann, & Razum, 2014) Die Analyse fachwissenschaftlicher Anforderungen umfasste sowohl einen Vergleich mit etablierten internationalen Dienstleistern als auch eine Analyse nationaler, überwiegend disziplinspezifischer Forschungsdatenrepositorien und zeigte Lücken in der bestehenden Infrastrukturlandschaft auf, beispielsweise das Fehlen disziplinübergreifender Angebote unter deutschem Recht. Insbesondere die Projektpartner aus den Bereichen Chemie und Biochemie trugen dazu bei, dass das RADAR-Dienstleistungsspektrum bestmöglich an die Bedarfe und Workflows von Forschenden an akademischen Einrichtungen angepasst wurde. Das RADAR-Metadatenschema wurde neben den Projektpartnern auch von Akteurinnen und Akteuren aus weiteren Fachdisziplinen (Material-, Geistes-, Sport- und Gesundheitswissenschaften) mit vorhandenen Forschungsdaten auf Anwendbarkeit geprüft. Die Konzeption und Entwicklung von RADAR orientierte sich zudem an den Ergebnissen nationaler und internationaler Initiativen.

Um die nutzergerechte Ausrichtung von RADAR zu gewährleisten, zielte das Projektteam auf eine enge Abstimmung mit den späteren Nutzenden und Institutionen ab.

Während der Projektphase wurden drei öffentliche Workshops mit mehr als 80 Teilnehmenden durchgeführt. Zum Ende des zweiten Projektjahres erfolgte die Freischaltung eines Testsystems, über das mehr als 30 Einrichtungen das System evaluieren und Anregungen für dessen Weiterentwicklung geben konnten. Für die Kommunikation mit der wissenschaftlichen Community wurde vielfältige Publikationen zum Projekt veröffentlicht und Präsentationen gehalten sowie eine zweisprachige Website mit Informationsmaterialien (beispielsweise in Form eines Glossars, FAQs oder der Dokumentation von Metadatenschema und API) erstellt.[1] Auf Anregung des Ausschusses für wissenschaftliche Bibliotheken und Informationssysteme (AWBI) wurde für das RADAR-Projekt ein wissenschaftlicher Beirat etabliert, dessen Vertreterinnen und Vertreter nationale Infrastruktureinrichtungen repräsentieren. Mit dem Beirat wurde insbesondere das Geschäftsmodell, die Integration des Dienstes in das wissenschaftspolitische Umfeld und die funktionale Ausgestaltung von RADAR diskutiert.

RADAR - Dienstleistungen, Dienstmerkmale und Geschäftsmodell

Funktional bietet RADAR seit Aufnahme des Dienstes drei zentrale Dienstleistungen an:

Datenarchivierung: Die Datenarchivierung dient der langfristigen und formatunabhängigen Aufbewahrung von Forschungsdaten für die jeweils festgelegten Haltefristen (z. B. 10 Jahre, wie von der DFG im Rahmen guter wissenschaftlicher Praxis empfohlen, aber auch abweichende Haltefristen sind möglich). Forschungsdaten werden in Form von paketierten Zusammenstellungen gesichert (bitstream preservation) und erhalten einen eindeutigen Identifier. Sofern von der Datengeberin bzw. dem Datengeber nicht anders vorgesehen, werden archivierte Forschungsdaten und zugehörige Metadaten nicht veröffentlicht. Durch eine flexible Zugriffsverwaltung können Datengeberinnen und Datengeber Datensätze gezielt mit benannten Dritten oder der Allgemeinheit teilen.

Datenpublikation: Publizierte Datensätze werden für mindestens 25 Jahre gesichert. Jedes publizierte Datenpaket erhält einen persistenten Identifier (DOI), wird

[1] https://www.radar-service.eu/de

automatisch bei DataCite indexiert und über standardisierte Protokolle (OAI-PMH) zum Harvesting angeboten. Dies sorgt für maximale Verbreitung und Auffindbarkeit der Forschungsdaten.

Über die DOI ist das Datenpaket eindeutig und dauerhaft identifizierbar, zitierfähig und kann mit wissenschaftlichen Publikationen verknüpft werden. Bei Bedarf können DOIs bereits vor der Datenpublikation reserviert werden. Falls Forschungsdaten nicht sofort veröffentlicht werden sollen, kann ein Embargo von bis zu einem Jahr festgelegt werden. Das Datenpaket ist dann erst nach Ablauf dieser Sperrfrist öffentlich zugänglich, die Metadaten sind sofort nach der Veröffentlichung einsehbar. Für jedes publizierte Datenpaket muss die Datengeberin bzw. der Datengeber eine Lizenz angeben (z. B. Creative Commons 4.0).

Peer-Review von Daten: Forschungsdaten können vor der Veröffentlichung im Rahmen eines Peer-Review-Prozesses begutachtet werden, beispielsweise durch externe Gutachterinnen und Gutachter, Verlage oder wissenschaftliche Zeitschriften. Über die Peer-Review-Funktion wird eine sichere URL generiert, die an die Gutachterinnen und Gutachter weitergeleitet werden kann. Nach Abschluss des Peer-Review-Prozesses verliert der Link seine Gültigkeit. Während des Peer-Review-Prozesses kann der betreffende Datensatz nicht verändert werden.

Zu den wesentlichen Dienstmerkmalen von RADAR zählen die Systemarchitektur, das Metadatenschema sowie das Rollen- und Rechtemodell. Eine weitere Eigenschaft des Dienstes ist das Geschäftsmodell.

Die RADAR-Systemarchitektur (Abb. 1) ist modular aufgebaut und besteht aus dem User Interface, der Management Layer und der Storage Layer (Archiv). Die Schichten kommunizieren über Application Programming Interfaces (API) miteinander. Dieser offene Aufbau ermöglicht die Integration von RADAR in bestehende Systeme und Arbeitsprozesse, wobei einzelne Komponenten von RADAR gegen eigene Lösungen ausgetauscht oder parallel betrieben werden können. Anwendungsfälle umfassen beispielsweise den Betrieb eines institutionseigenen Frontends, das automatisierte Hochladen von Forschungsdaten in die Management Layer oder die Übertragung beziehungsweise den Abruf von (Meta-)Daten aus anderen Anwendungen.

Die gesamte entwickelte Software ist ausschließlich aus Open Source Komponenten aufgebaut und implementiert ein OAIS-konformes Langzeitarchivierungssystem.

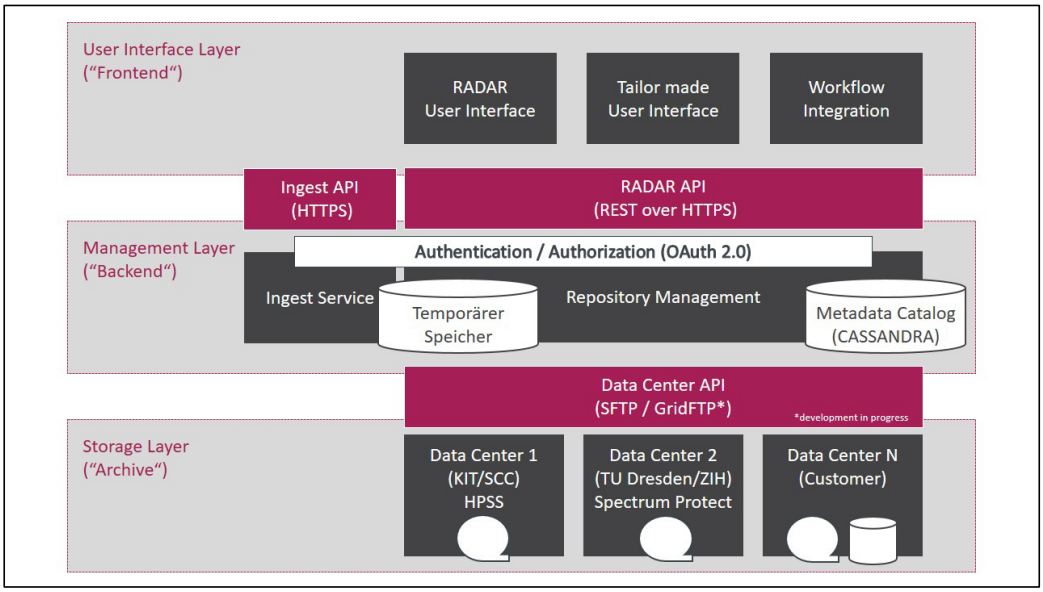


Abbildung 1: RADAR-Systemarchitektur

Das RADAR Metadatenschema[2] ist disziplinübergreifend angelegt und kompatibel mit dem DataCite Metadata Schema und DublinCore. Es erlaubt die Verwendung von Normdaten für Personen (ORCID iD) und Förderorganisationen (CrossRef Open Funder Registry). Das Metadatenschema umfasst neben 10 Pflichtfeldern, die für die DOI-Registrierung notwendig sind, 13 optionale Felder zur Beschreibung der Datenerhebung und -aufbereitung. Durch die Kombination aus kontrollierten Vokabularen und Freitext-Einträgen ermöglicht RADAR die Interoperabilität der beschriebenen Forschungsdaten, während gleichzeitig der Heterogenität der Daten aus einer Vielzahl von Disziplinen Rechnung getragen wird.

Ein klar definiertes Rollen- und Rechtemodell (Abb. 2) erlaubt die delegierte Administration durch die nutzende Einrichtung. Damit können entsprechend der jeweiligen Bedürfnisse Prozesse strukturiert, Aufgaben verteilt und interne Verantwortlichkeiten definiert werden. Von der nutzenden Einrichtung eingesetzte

[2] https://www.radar-service.eu/de/radar-schema

Administratorinnen und Administratoren verwalten die RADAR-Arbeitsbereiche, die als zentrale Einstiegspunkte für Forschende einer Arbeitsgruppe oder eines Projekts dienen. Administratorinnen und Administratoren können Kuratorinnen und Kuratoren bestimmen, welche Forschungsdaten in einem Arbeitsbereich ablegen, mit Metadaten beschreiben und nach qualitätssichernden Maßnahmen archivieren oder publizieren. Optional können Subkuratorinnen und Subkuratoren benannt werden, die Forschungsdaten bearbeiten und beschreiben, aber nicht archivieren oder publizieren können. Über eine integrierte Registrierung wird die Nutzung von RADAR durch Forschende administrativ erleichtert. Sofern die Einrichtung an DFN-AAI teilnimmt, erfolgt die Authentifizierung mit der institutionellen Nutzerkennung.

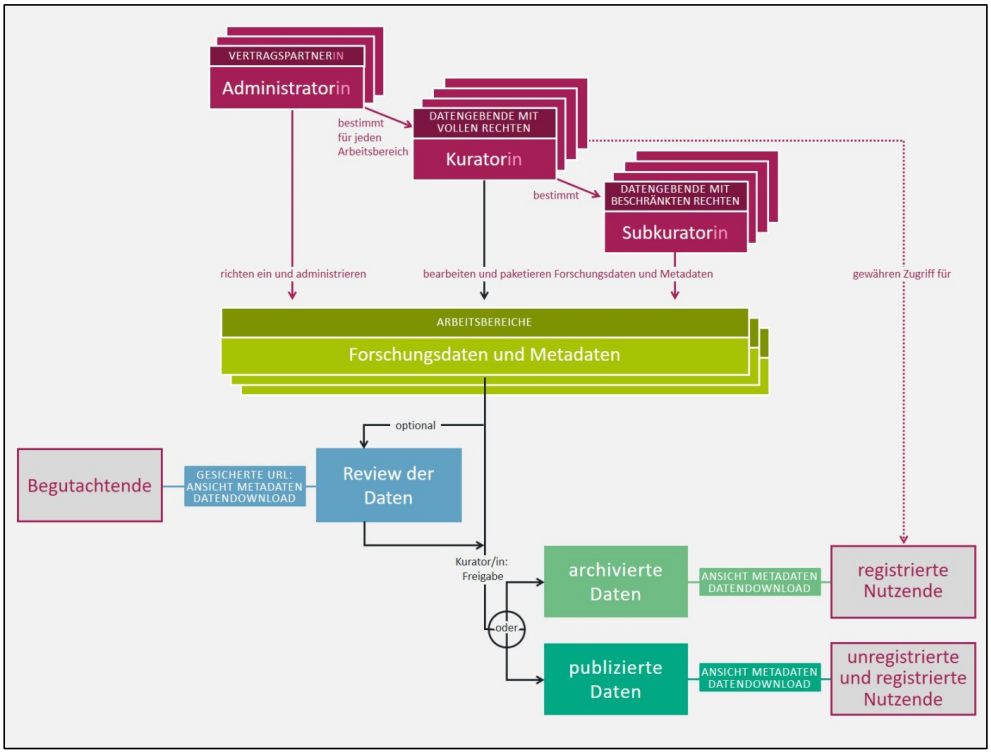


Abbildung 2: RADAR Rollen- und Rechtemodell

Geschäftsmodell: RADAR ist auf langfristigen Betrieb ausgelegt und operiert nicht gewinnorientiert. Die Nutzung von RADAR setzt den Abschluss eines Dienstleistungsvertrags voraus, für den eine jährliche Grundgebühr anfällt.

Die Dienstleistungen Datenarchivierung und Datenpublikation werden nutzungsbasiert abgerechnet. Für archivierte Daten fallen innerhalb der gewählten Haltefrist jährliche Kosten an. Nutzende Einrichtungen können archivierte Daten nach Vertragsende aufgrund der verwendeten Standardformate für die Datenpakete einfach an andere Dienstleister weitergeben und damit das Risiko eines „Vendor Lock-ins“ ausschließen. Publizierte Daten müssen unabhängig von der Vertragssituation verfügbar bleiben. Aus diesem Grund bietet RADAR ein Modell basierend auf Einmalzahlungen für veröffentlichte Daten mit einer garantierten Haltefrist von mindestens 25 Jahren, selbst nach Vertragsende.

Mit RADAR können nutzende Einrichtungen Kosten einsparen, die für den Aufbau, die Weiterentwicklung und den Unterhalt eigener Forschungsdatendienste anfallen. Gleichzeitig ermöglicht die delegierte Administration an der nutzenden Einrichtung die Verwaltung von Forschungsdaten durch vorhandene Stellen, beispielsweise durch Bibliothekspersonal, die auch die Beratung der Forschenden vor Ort übernehmen können. Durch die Transparenz des Preismodells können Kosten, die für die Nutzung von RADAR-Diensten anfallen, vorab kalkuliert und über Drittmittel mit eingeworben werden. Über Quotas lassen sich zudem Kostenobergrenzen flexibel festlegen, um notwendige Mittel besser planen zu können.

Aktuelle Entwicklungen im FDM-Umfeld

Seit RADAR 2017 den Produktivbetrieb aufnahm, hat sich das Umfeld für Forschungsdatenrepositorien dynamisch weiterentwickelt. Im neu ausgerichteten Förderprogramm "Fachinformationsdienste für die Wissenschaft" der DFG entwickeln einige FIDs disziplinspezifische Forschungsdatendienste. (Deutsche Forschungsgemeinschaft. Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme, 2019; Harbeck & Kaun, 2019) Disziplinübergreifende Forschungsdatenrepositorien verzeichneten in den letzten Jahren einen deutlichen Anstieg der abgelegten Datensätze. (Assante, Candels, Castelli & Tani, 2016) Allerdings gibt es bisher wenige disziplin- und einrichtungsübergreifende Repositorien, die unter deutscher Jurisdiktion stehen.

Auf Landesebene haben sich Initiativen gebildet, die einrichtungsübergreifende Ansätze für das FDM entwickeln, beispielsweise HeFDI in Hessen oder FoDaKo und

die Landesinitiative NFDI der Digitalen Hochschule in Nordrhein-Westfalen. (Krause et al., 2018; Hess, von Rekowski, Roller, & Walger, 2019; Grasse, López, & Winter, 2018) Parallel soll mit der Nationalen Forschungsdateninfrastruktur (NFDI) ein übergreifendes FDM für das deutsche Wissenschaftssystem entstehen. (Gemeinsame Wissenschaftskonferenz, 2018). Dabei wird die wissenschaftsgeleitete Etablierung disziplinärer Konsortien angeregt, die Fächerkompetenz mit der Etablierung geeigneter Infrastrukturlösungen verknüpfen. Daneben gibt es auf internationaler Ebene Bestrebungen zu einer intensiveren Vernetzung bestehender Infrastrukturkomponenten im Rahmen der European Open Science Cloud (Europäische Kommission, 2018).

2016 wurden mit den FAIR Principles Ziele für Forschungsdaten formuliert: "findable, accessible, interoperable, reusable". (Wilkinson et al., 2016) Indem immer mehr Interessensgruppen Bezug auf die FAIR Principles nehmen, werden sie zunehmend zu einem Quasi-Standard für den Umgang mit Forschungsdaten. (Rat für Informationsinfrastrukturen, 2019; Deutsche Bundesregierung, 2018)

Mit der Bewusstseinsschärfung für die Notwendigkeit von FDM wächst auch das Verständnis für damit verbundene Kosten. Neben der Datenspeicherung verursachen diese insbesondere personalintensive Arbeitsschritte wie Übernahme in das Datenarchiv, Erstellung von Metadaten und Maßnahmen zur digitalen Langzeitarchivierung. (Beagrie, Chruszcz & Lavoie, 2008) (Wollschläger & Dickmann, 2016) Eine aktuelle Studie, die von der Europäischen Kommission in Auftrag gegeben wurde, macht jedoch auch auf Opportunitätskosten in Höhe von mindestens 10,2 Milliarden Euro pro Jahr aufmerksam, die durch das Fehlen von FAIR Data in Europa entstehen. (Europäische Kommission. Generaldirektion Forschung und Innovation, 2019, S. 26) Vor diesem Hintergrund ist die Entwicklung nachhaltiger Geschäftsmodelle nach wie vor eine zentrale Herausforderung für Forschungsdatenprogramme. Das zeigt beispielsweise die Bewertung des Förderprogramms „Informationsinfrastrukturen für Forschungsdaten“ durch die Deutsche Forschungsgemeinschaft - demnach wird der langfristige Betrieb von Infrastrukturen, die im Rahmen von Drittmittelprojekten entstehen, insbesondere durch das Auslaufen drittmittelfinanzierter, befristeter Stellen erschwert. (Deutsche Forschungsgemeinschaft, 2019a, S.56)

RADAR im Produktivbetrieb

Mit dem Abschluss mehrerer Dienstleistungsverträge innerhalb weniger Monate nach Aufnahme des Produktivbetriebs wurde die Geschäftsplanung im ersten Betriebsjahr übertroffen. Auch in den Folgejahren konnten weitere Einrichtungen für RADAR gewonnen werden. Funktionale Erweiterungen, wie beispielsweise die Sicherstellung der Barrierefreiheit, die Einführung einer optionalen Datenintegritätsprüfung oder erste Branding-Optionen wurden im Rahmen der laufenden Softwarepflege umgesetzt. Die RADAR-Website[3] wird kontinuierlich um Informationsmaterialien, Dokumente und News erweitert. Darüber hinaus wird ein eigener Twitter-Account betrieben. Die Fachöffentlichkeit wird regelmäßig über den aktuellen Funktionsumfang informiert und durch Workshops und Präsentationen in die Weiterentwicklung des Dienstes eingebunden. Die jährlichen Workshops richten sich sowohl an bestehende Nutzende als auch an Einrichtungen, die Interesse am Einsatz von RADAR haben. Um den Austausch mit den Teilnehmenden zu fördern und deren Bedarfe festzustellen, werden interaktiven Elementen wie Fragerunden, Beiträgen der Teilnehmenden und Diskussionen viel Raum gegeben. Das RADAR-Testsystem steht weiterhin zur Verfügung und bietet interessierten Einrichtungen die Möglichkeit, den Dienst unverbindlich kennenzulernen und Rückmeldung zu geben. Zudem werden in Vertragsverhandlungen Bedarfe und Anforderungen geäußert, die in Entscheidungen über die Weiterentwicklung des Dienstes einfließen. Trotz des breiten Interesses überrascht die Zahl über RADAR publizierter Datensätze, die derzeit im niedrigen zweistelligen Bereich liegt. Andererseits deckt sich diese Beobachtung mit Umfrageergebnissen, die auf eine bislang eingeschränkte Bereitschaft von Forschenden zum öffentlichen Teilen von Forschungsdaten hinweisen. (Fecher, 2018) Gleichzeitig zeigt sich, dass zwischen Vertragsabschluss und der Publikation erster Datensätze oftmals mehrere Monate liegen, in denen interne Abläufe in den nutzenden Einrichtungen etabliert und Qualitätssicherungsprozesse durchlaufen werden.

[3] https://www.radar-service.eu/de

Positionierung von RADAR und geplante Erweiterungen

Disziplinübergreifende Forschungsdatenrepositorien spielen eine wichtige Rolle bei der Datenpublikation in Disziplinen, die bisher keine eigenen Forschungsdateninfrastrukturen und -dienste anbieten. Allerdings müssen sie auch Lösungen für vielfältige Datenformate und -typen, Forschungsgemeinschaften und -praktiken anbieten. (Assante, Candela, Castelli & Tani, 2016)

RADAR wurde als disziplin- und formatunabhängiges Repositorium konzipiert und ermöglicht Einrichtungen darum, Angebote für eine Vielzahl von Datentypen zu entwickeln. Durch den modularen und offenen Aufbau kann RADAR flexibel an lokale Bedürfnisse angepasst werden. Die oben beschriebenen Entwicklungen des wissenschaftspolitischen Umfelds machen allerdings deutlich, dass disziplinübergreifende Angebote zukünftig stärker auf disziplinspezifische Besonderheiten eingehen und die Interoperabilität befördern müssen.

Forschungsdateninfrastrukturen erzeugen ein "Ökosystem" digitaler Dienste und Ressourcen und wirken als Ermöglicher bestimmter Forschungspraktiken. (Bowker, Baker, Millerand, & Ribes, 2010) Als Schnittstelle zwischen Datengeberinnen bzw. Datengebern und Nutzenden fördern sie den Austausch innerhalb der "Designated Community" und damit die Akzeptanz von Forschungsdaten als eigenständige wissenschaftliche Objekte. RADAR verpflichtet sich der Umsetzung der FAIR Principles und bietet eine Reihe von Maßnahmen zu ihrer Implementierung an (Tab. 1). Zukünftige Weiterentwicklungen des Dienstes sollen daher auch die Auffindbarkeit, Zugänglichkeit, Interoperabilität und Nachnutzbarkeit von Forschungsdaten verbessern.

Findable	**F1:** (meta)data are assigned a globally unique and eternally persistent identifier	• Registrierung von DOIs für publizierte Datenpakete, RADAR-IDs für archivierte Datenpakete
	F2: data are described with rich metadata.	• Metadatenschema basierend auf dem DataCite Metadata Schema • Überprüfung der Metadaten auf Vollständigkeit
	F3: (meta)data are registered or indexed in a searchable resource.	• Indexierung bei DataCite, Google und B2FIND • Harvesting über OAI-PMH
	F4: metadata specify the data identifier.	• Feld <identifierType> wird bei Ausstellung des Identifiers automatisch ausgefüllt
Accessible	**A1:** (meta)data are retrievable by their identifier using a standardized communications protocol.	• (Meta-)Daten sind über eine API und OAI-PMH zugänglich
	A1.1: the protocol is open, free, and universally implementable.	• Verwendete Protokolle und Schnittstellen sind weit verbreitet und gut dokumentiert
	A1.2: the protocol allows for an authentication and authorization procedure, where necessary.	• Rollen- und Rechtemodell erlaubt Datengeberinnen bzw. Datengebern Kontrolle über Zugriff auf Datensätze
	A2: metadata are accessible, even when the data are no longer available.	• Landing page bleibt nach der Löschung eines Datensatzes erhalten
Interoperable	**I1:** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	• Metadatenschema basierend auf XML
	I2: (meta)data use vocabularies that follow FAIR principles.	• Personenidentifikation über ORCID iDs • Angaben zu Förderorganisationen über Crossref Funder Registry
	I3: (meta)data include qualified references to other (meta)data.	• Referenzierung anderer digitaler Ressourcen über persistente Identifier
Re-Usable	**R1:** meta(data) have a plurality of accurate and relevant attributes.	• Umfassende Beschreibung der Eigenschaften von Forschungsdaten über das Metadatenschema
	R1.1: (meta)data are released with a clear and accessible data usage license.	• Für jedes Datenpaket muss eine Lizenz (z.B. Creative Commons 4.0) vergeben werden
	R1.2: (meta)data are associated with their provenance.	• Metadatenschema erlaubt Angaben zur Provenienz von Forschungsdaten
	R1.3: (meta)data meet domain-relevant community standards.	• Metadatenschema wurde auf Anwendbarkeit in verschiedenen Disziplinen getestet

Tabelle 1: Umsetzung der FAIR Principles mit RADAR

Einer der Hauptmerkmale von RADAR war bisher die entfallende Notwendigkeit, eine eigene Dateninfrastruktur aufzubauen und zu betreiben. Die Kommunikation mit interessierten Einrichtungen zeigte, dass dieses Modell nicht alle Anwendungsfälle adäquat abdecken kann. Eine Öffnung des Betriebs- und Geschäftsmodells soll darum die flexible Anpassung an alternative Einsatzszenarien ermöglichen.

Im Folgenden werden Ansätze der bedarfsgesteuerten Weiterentwicklung des Dienstes geschildert.

Einbindung eigener Rechenzentren

Bisher ist RADAR ausschließlich als Cloud-Dienst verfügbar, der die dauerhafte Archivierung und Verfügbarkeit der Forschungsdaten gewährleistet. In Gesprächen mit größeren Hochschulen, außeruniversitären Forschungseinrichtungen und Konsortien wurde der Wunsch nach der Einbindung eigener Rechenzentren in die Storage Layer geäußert, entweder exklusiv oder ergänzend zu bestehenden Datenzentren (SCC und ZIH). Diese Option würde nutzenden Einrichtungen mehr Kontrolle über Forschungsdaten einräumen und könnte gleichzeitig Kosten senken. Derzeit werden technische Anpassungen vorgenommen, die die Umsetzung institutionsspezifischer Vorgaben für die Speicherung von Archivkopien unterstützen. Diese Anpassungen erfordern auch eine Weiterentwicklung des Geschäftsmodells (beispielsweise eine Erweiterung um Pauschalpreise) sowie der Verträge und Haftungsregelungen. Die RADAR-Software wird in diesem Modell nutzenden Einrichtungen weiterhin als gehosteter Dienst zur Verfügung gestellt, dabei aber entweder eine oder alle Kopien auf lokaler Speicherinfrastruktur abgelegt werden. Darüber hinaus ist ein Szenario vorgesehen, in dem die nutzende Einrichtung die RADAR-Software selbst auf eigener Hardware betreibt, aber über einen Wartungsvertrag dabei unterstützt wird und regelmäßig Software-Updates erhält.

Unterstützung großer Datenvolumen

Mit der Adressierung von Konsortien und größeren Hochschulen als Zielgruppe gehen Anforderungen hinsichtlich der Verarbeitung großer Datenmengen (im zwei- bis dreistelligen Terabyte-Bereich) einher, die effizientere Übertragungs- und Verarbeitungsverfahren erfordern. Für solche Datenmengen wurde RADAR als Repositorium für den “Long Tail” nicht konzipiert. Um große Datenvolumina besser zu unterstützen, soll zukünftig ein Datentransfer auch per GridFTP möglich sein. Das bereits in 2018 eingeführte Rabattpreismodell für Datenvolumen über 50 TB greift diesen Aspekt bereits auf.[4]

[4] https://www.radar-service.eu/de/preise

Customizing-Optionen

Neben der Einbindung von eigener Speicherinfrastruktur ermöglicht RADAR in Zukunft weitere Anpassungen für nutzende Einrichtungen. Dazu zählen die Vergabe von DOIs mit einem institutionseigenen Präfix, Anpassungen an das Corporate Design, Verweise auf institutionseigene Unterstützungsangebote sowie die Möglichkeit einer institutionellen Sicht, über die nur die von der jeweiligen Institution publizierten Datensätze angezeigt werden. Diese Anpassungen befinden sich aktuell in der Umsetzung, initiiert durch die Anforderungen des Karlsruher Institut für Technologie (KIT), das als Projektpartner aktuell ein auf RADAR basierendes eigenes Forschungsdatenrepositorium aufbaut. Diese Anpassungen sollen auch anderen nutzenden Institutionen noch 2019 zur Verfügung stehen.

Disziplinspezifische Anpassungen

FIZ Karlsruhe bringt sich mit RADAR in mehreren sich formierenden NFDI-Konsortien als Infrastrukturpartner für die Bereitstellung eines bedarfsgerechten Forschungsdatenrepositoriums ein. RADAR kann sich hier als generischer Baustein für disziplinspezifische Angebote etablieren. Dazu wird RADAR über entsprechende Erweiterungen an disziplinspezifische Bedarfe angepasst. Dies betrifft z. B. die optionale Beschreibung von Datenpaketen mit zusätzlichen, fachspezifischen Metadaten in Form von XML-Dateien. Sofern ein geeignetes Schema referenziert wird, kann die mitgelieferte XML-Datei automatisch validiert werden. Mittels Harvesting von Metadaten über OAI-PMH und einer entsprechenden Indexierung lassen sich fachspezifische Suchen aufbauen. In diesem Zusammenhang könnten NFDI-Konsortien wertvolle Impulse für die Verwendung geeigneter Metadatenschemata geben, beispielsweise unter Anwendung des Research Data Alliance Metadata Directory.[5]

Verknüpfung mit anderen Forschungsdatendiensten

Weiterhin prüft RADAR die Kooperation mit anderen Forschungsdatendiensten, um Forschenden bereits früher im Lebenszyklus von Forschungsdaten sinnvolle Dienste anzubieten. Hier ist insbesondere RDMO zu nennen: ein Werkzeug, das Forschende bei der Erstellung und Umsetzung von Datenmanagementplänen (DMP) unterstützt

[5] https://rd-alliance.github.io/metadata-directory/standards/

und das bereits von vielen Einrichtungen genutzt wird.[6] (Neuroth et al., 2018) Durch die Integration von RDMO könnte RADAR den gesamten Lebenszyklus von Forschungsdaten begleiten, vom Projektstart über die Datenerhebung und -aufbereitung bis zur Datenpublikation beziehungsweise -archivierung. Durch den Import von Metadaten aus RDMO könnten Doppeleingaben vermieden und die Nutzerfreundlichkeit von RADAR verbessert werden. Perspektivisch wäre auch eine stärkere Integration von RDMO und RADAR im Sinne von "machine actionable Data Management Plans" möglich. (Miksa, Simms, Mietchen & Jones, 2019)

Personenbezogene Daten

Gemäß dem RADAR-Dienstleistungsvertrag ist die Archivierung beziehungsweise Publikation personenbezogener, nicht anonymisierter Daten ohne Einwilligung der Betroffenen nicht zulässig. Bei Beratungsgesprächen mit interessierten Einrichtungen wird dieses Thema regelmäßig adressiert. Aus diesem Grund wird sich RADAR zukünftig stärker mit möglichen Angeboten für personenbezogene Daten befassen. Basierend auf Fallbeispielen soll aus juristischer und technischer Perspektive untersucht werden, ob unter Anwendung bestehender Möglichkeiten der Pseudonymisierung eine Erweiterung des Angebots möglich ist.

Zertifizierung

Da Forschungsdatenrepositorien die Verantwortung für die langfristige Verfügbarkeit von und den Zugriff auf Forschungsdaten übernehmen, müssen sie Qualitätsansprüchen aller beteiligter Interessensgruppen gerecht werden. Viele der am FDM beteiligten Akteure, darunter Forschende, Verlage und Forschungsförderer, stellen ähnliche Ansprüche an Forschungsdatenrepositorien. (Husen, de Wilde, de Ward & Cousijn, 2017) Zertifikate wie das CoreTrustSeal[7] formalisieren diese Anforderungen und zielen darauf ab, die Vertrauenswürdigkeit von Forschungsdatenrepositorien zu bewerten und sichtbar zu machen. (Dillo & Leeuw, 2018) Dadurch fördern sie Transparenz und können Datengeberinnen und Datengebern Orientierung bei der Entscheidung für geeignete

[6] https://rdmorganiser.github.io/kooperationen/
[7] https://www.coretrustseal.org/

Forschungsdateninfrastrukturen geben. Bisher ist die Zertifizierung von Forschungsdatenrepositorien allerdings noch nicht weit verbreitet.

(Kindling et al., 2017) RADAR lässt sich derzeit nach dem CoreTrustSeal zertifizieren. Damit möchte RADAR einerseits das Vertrauen (potenzieller) Vertragspartner sicherstellen beziehungsweise gewinnen. Andererseits können Vertragspartner durch den zertifizierten Dienst die Vertrauenswürdigkeit an Datengeberinnen und Datengeber "weiterreichen".

Zusammenfassung und Ausblick

Die Betrachtung des nationalen und internationalen Umfelds macht deutlich, wie dynamisch sich die wissenschaftspolitischen Rahmenbedingungen des Forschungsdatenmanagements verändern. Infrastrukturanbieter müssen darum in der Lage sein, flexibel auf veränderte Anforderungen zu reagieren und gleichzeitig ein Modell für die langfristige Verfügbarkeit ihrer Dienstleistung implementieren.

Seit der Überführung von RADAR in den Produktivbetrieb wird der Dienst bedarfsorientiert weiterentwickelt. Bisherige Erfolge bei der Zusammenarbeit mit Forschungseinrichtungen und das breite Interesse von Einrichtungen schlagen sich bisher noch nicht in einer intensiven Nutzung durch Forschende nieder. Als Dienstleister für wissenschaftliche Einrichtungen kann RADAR auf die Nutzung durch Forschende nur indirekt einwirken. Die hier vorgestellten Anpassungen und Erweiterungen zielen darauf ab, Forschungsdatenmanagement und die Publikation von Forschungsdaten für Forschende einfacher und attraktiver zu gestalten. Darum setzt RADAR zukünftig verstärkt auf disziplinspezifische Angebote und den Kontakt zu forschungsnahen Diensten, beispielsweise über Konsortien im Rahmen der NFDI.

Zu diesem Zweck öffnet sich RADAR neuen Kooperationspartnern und -formen. In diesen Kooperationen können sich RADAR und FIZ Karlsruhe mit umfassender technischer, juristischer und informationswissenschaftlicher Expertise einbringen.

Der Ausbau von Customizing-Optionen, der auch die Einbindung institutionseigener Speicherkapazität einschließt, ermöglicht den Einsatz von RADAR für ein breiteres Anwendungsspektrum. Zusätzlich soll der Funktionsumfang erweitert werden, beispielsweise durch die Unterstützung großer Datenvolumen und die Verknüpfungen mit anderen Forschungsdatendiensten.

Durch die Zertifizierung nach dem CoreTrustSeal sollen die Qualität und Vertrauenswürdigkeit des Dienstes gesichert und an die Fachcommunity weitergereicht werden.

Literatur

Assante, M., Candela, L., Castelli, D., & Tani, A. (2016). Are Scientific Data Repositories Coping with Research Data Publishing? *Data Science Journal*, *15*, 6. https://doi.org/10.5334/dsj-2016-006

Beagrie, N., Chruszcz, J., & Lavoie, B. (2008). *Keeping Research Data Safe. A Cost Model and Guidance for UK Universities*. Abgerufen von https://www.webarchive.org.uk/wayback/archive/20140613220103/http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger, L. Klastrup, & M. Allen (Hrsg.), *International Handbook of Internet Research* (S. 97–117). https://doi.org/10.1007/978-1-4020-9789-8_5

Deutsche Bundesregierung. (2018). *Die Hightech-Strategie 2025 – Forschung und Innovation für die Menschen*. Abgerufen von http://dip21.bundestag.de/dip21/btd/19/041/1904100.pdf

Deutsche Forschungsgemeinschaft. (2019). *Bewertung des Förderprogramms „Informationsinfrastrukturen für Forschungsdaten“*. http://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/studien/studie_forschungsdaten.pdf

Deutsche Forschungsgemeinschaft. Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme. (2019). *Fachinformationsdienste für die Wissenschaft. Von den Sondersammelgebieten zu den Fachinformationsdiensten: Zwischenbilanz der Umstrukturierung der Förderung*. Abgerufen von https://www.dfg.de/download/pdf/foerderung/programme/lis/fid_zwischenbilanz_umstrukturierung_foerderung_sondersammelgebiete.pdf

Dillo, I., & Leeuw, L. de. (2018). CoreTrustSeal. *Mitteilungen Der Vereinigung Österreichischer Bibliothekarinnen Und Bibliothekare*, *71*(1), 162–170. https://doi.org/10.31263/voebm.v71i1.1981

Europäische Kommission. (2018). *Implementation Roadmap for the European Open Science Cloud*. Abgerufen von https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

Europäische Kommission. Generaldirektion Forschung und Innovation. (2019). *Cost of not having FAIR research data. Cost-benefit analysis for FAIR research data*. Abgerufen von https://doi.org/10.2777/02999

Fecher, B. (2018). *Eine Reputationsökonomie*. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-20895-0

Gemeinsame Wissenschaftskonferenz. (2018). *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur*. Abgerufen von https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf

Grasse, M., López, A., & Winter, N. (2018). Landesinitiative NFDI – a Central Point of Contact for RDM for Higher Education Institutions in the German State of North Rhine-Westphalia. *Data Science Journal*, *17*, 25. https://doi.org/10.5334/dsj-2018-025

Hahn, M., Neumann, J., & Razum, M. (2014). RADAR – Ein Forschungsdaten-Repositorium als Dienstleistung für die Wissenschaft. *Zeitschrift Für Bibliothekswesen Und Bibliographie*, *61*(1), 18–27. https://doi.org/10.3196/186429501461150

Harbeck, M., & Kaun, M. (2019). Forschungsdaten und Fachinformationsdienste – eine Bestandsaufnahme. *Bibliothek Forschung und Praxis*, *43*(1), 35–41. https://doi.org/10.1515/bfp-2019-2015

Hess, V., von Rekowski, T., Roller, S., & Walger, N. (2019). Synergieeffekte durch Kooperation: Hintergründe, Aufgaben und Potentiale des Projekts FoDaKo. *Bibliothek Forschung und Praxis*, *43*(1), 98–104. https://doi.org/10.1515/bfp-2019-2009

Husen, S. E., de Wilde, Z. G., de Waard, A., & Cousijn, H. (2017). Recommended versus Certified Repositories: Mind the Gap. *Data Science Journal*, *16*, 42. https://doi.org/10.5334/dsj-2017-042

Kindling, M., Pampel, H., van de Sandt, S., Rücknagel, J., Vierkant, P., Kloska, G., … Scholze, F. (2017). The Landscape of Research Data Repositories in 2015: A re3data Analysis. *D-Lib Magazine*, *23*(3/4). https://doi.org/10.1045/march2017-kindling

Krause, E., Brand, O., Deppe, A., Krähwinkel, E., Jagusch, G., Müllerleile, T., … Wolff-Wölk, A. (2018). FDM vernetzt und kooperativ. *o-bib. Das offene Bibliotheksjournal*, *5*(4), 220–236. https://doi.org/10.5282/o-bib/2018H4S220-236

Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2019). Ten principles for machine-actionable data management plans. *PLOS Computational Biology*, *15*(3), e1006750. https://doi.org/10.1371/journal.pcbi.1006750

Neuroth, H., Engelhardt, C., Klar, J., Ludwig, J., & Enke, H. (2018). Aktives Forschungsdatenmanagement. *ABI Technik*, *38*(1), 55–64. https://doi.org/10.1515/abitech-2018-0008

RADAR. (2016). *Abschlussbericht an die Deutsche Forschungsgemeinschaft*. Abgerufen von https://www.radar-projekt.org/download/attachments/753675/Abschlussbericht_DFG-Projekt_RADAR_Ver%C3%B6ffentlichung.pdf?version=1&modificationDate=1485414994000&api=v2

Rat für Informationsinfrastrukturen. (2019). *Stellungnahme des Rates für Informationsinfrastrukturen (RfII) zu den aktuellen Entwicklungen rund um Open Data und Open Access*. Abgerufen von http://www.rfii.de/?p=3748

Wollschläger, T., & Dickmann, F. (2016). Kosten. In H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, & K. Huth (Hrsg.), *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. (Version 2.3, Kapitel 14.2). http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949

CiTAR – Citable Scientific Software and Software Methods

Klaus Rechert, Rafael Gieschke, Susanne Mocken, Oleg Stobbe, Oleg Zharkov[1]
Felix Bartusch[2] , Kyryll Udod[3]

[1]Universität Freiburg
[2]Universität Tübingen
[3]Universität Ulm

Bewahrung virtueller Forschungsumgebungen: CiTAR

Moderne Forschungsvorhaben stützen sich nicht nur auf computergestützte Methoden und digitale Ressourcen, sondern entwickeln häufig auch spezifische Software und software-gestützte Prozesse. Ein wesentliche Säule der Wissenschaft ist die Nachvollziehbarkeit sowie gegebenenfalls die Reproduktion oder Verifikation veröffentlichter Erkenntnisse. Für computergestützte Forschung wird dies zunehmend zu einer Herausforderung, insbesondere durch die komplexe Rekonstruktion software-basierter Methoden, ihrer spezifischen Konfiguration und technischen Abhängigkeiten. Dieser Beitrag stellt CiTAR[1] - *Citing and Archiving Research* vor, ein dreijähriges Projekt gefördert durch das Ministeriums für Wissenschaft und Kunst des Landes Baden-Württemberg, das eine Infrastruktur zur Unterstützung computergestützter Forschung entwickelt. Im Fokus stehen dabei die Methoden, d.h. softwaregestützte Prozesse oder Modelle zur Erstellung oder Auswertung von Daten. Das wesentliche Projektziel ist es, diese Methoden nachvollziehbar und nachnutzbar nachzuweisen, so dass diese ebenso zitierbar und publizierbar werden wie derzeit schon Forschungsdaten (vgl. z.B. Brase 2009).

Sowohl die Erhaltung als auch die Zitierfähigkeit von (wissenschaftlicher) Software als weiterer grundlegender Aspekt einer Forschungsdatenstrategie ist im wesentlichen unbestritten[2] [3], diverse Akteure und Organisationen haben sich bereits dem Thema bereits gewidmet.[4] [5] Für eine erfolgreiche Reproduktion oder Verifikation (z.B. definierbar als "*the ability for someone to replicate a computational experiment that*

[1] http://citar.eaas.uni-freiburg.de/
[2] National Institutes of Health (NIH), https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-015.html
[3] https://danielskatzblog.wordpress.com/2017/02/22/creation-publication-and-citation-issues-in-software- citation-versus-paper-and-data-citation/
[4] Software Preservation Network http://www.softwarepreservationnetwork.org
[5] UNESCO PERSIST Initiative, https://www.unesco.nl/digital-sustainability

was done by someone else, using the same software and data, and then to be able to change part of it (the software and/or the data) to better understand the experiment and its bounds[.]"[6]) ist das Vorhalten von einzelnen Softwarekomponenten nicht ausreichend. Software-gestützte Prozesse bestehen meist aus mehreren Softwarekomponenten, die einer spezifischen Konfiguration und Orchestrierung bedürfen. Für zitierfähige und nachnutzbare Softwaremethoden sind also eigenständig nutzbare Einheiten notwendig, die eine Kapselung spezifischer Softwareinstallationen ermöglichen. Im Rahmen von CiTAR wurden virtuelle Maschinen (VMs) und Container (Meng et al. 2015, Boettinger 2015) als geeignete Einheiten identifiziert und ein Prozess entwickelt, um diese als eigenständige Objekte aus den Labor- und Arbeitsumgebungen der Entwickler und Forschenden (Produzenten) herauszulösen und von diesen unabhängig funktional zu erhalten. Ein publiziertes Zitat impliziert langfristig aber auch Verfügbarkeit, so dass die Langzeiterhaltung der so nachgewiesenen Methoden eine wichtige Nebenbedingung für die im Projekt entwickelten Konzepte und Infrastruktur darstellen. Ein wesentliches Ziel war somit, für VMs und Container jeweils eine homogene, technische Objektdefinition zu entwickeln, so dass die Erhaltungsplanung sowie spätere Erhaltungsmaßnahmen -- mit nur wenigen verschiedenartigen Objektarten -- effizient erfolgen können.

Erhaltung und Nachnutzung virtueller Maschinen

Als Beispiel für die Aufrechterhaltung eines langfristigen Zugangs zu softwarebasierten Forschungsressourcen haben wir unter anderem die Ergebnisse des Projekts SlaVaComp (2013-2015) genutzt, das ein elektronisches Meta-Glossar mit regionalen und diachronen Varianten des Kirchenslavischen erstellt hat -- eine Sprache, die im orthodoxen slavischen Raum zwischen dem 10. und 16. Jahrhundert verwendet wurde. Bis zur Erstellung einer digitalen Datenbank hatten Forscher nur die Möglichkeit, gedruckte Wörterbücher zu konsultieren, so dass selbst einfache Recherchen sehr viel Zeit in Anspruch nehmen konnten. Fünfzehn gedruckte kirchenslavisch-griechische und griechisch-kirchenslavische Glossare mit verschiedenen Herkunftsregionen wurden zu einer einfach zu bedienenden, webbasierten Online-Anwendung zusammengefasst.

[6] https://danielskatzblog.wordpress.com/2017/02/07/is-software-reproducibility-possible-and-practical/

Die Kosten für die Wartung eines Servers und der Software nach Projektende werden in der Regel nicht durch den Mittelgeber übernommen, insbesondere wenn diese Projekte nur für eine kleine und spezialisierte Forschungsgemeinschaft von Interesse sind. Der Betrieb einer nicht gewarteten, veralteten Maschine, die mit dem Internet verbunden ist, stellt jedoch ein latentes und zunehmendes Sicherheitsrisiko dar. Da der Support für das zugrunde liegende Serverbetriebssystem in naher Zukunft ausläuft, ist die Zukunft des SlaVaComp-Dienstes ungewiss. Selbst wenn das Betriebssystem auf die nächste Serverversion mit Langzeitsupport aktualisiert wird, ist damit nicht sichergestellt, dass andere Softwareabhängigkeiten, wie z.B. die zugrundeliegende Datenbank, kompatibel bleiben oder Sicherheitsupdates langfristig zur Verfügung stehen werden. Dies ist jedoch für einen öffentlichen Online-Dienst entscheidend. Eine Migration des Services auf eine modernen Softwarebasis ist aufwendig und bedarf der Mithilfe der ursprünglichen Entwickler, die aber in der Regel nach dem Ende des Projekts nicht mehr da sind und somit das umfassende, spezifische Wissen über die von ihnen erstellte Software mitgenommen haben. Dieses Schicksal teilen zahlreiche Softwareentwicklungen, die aus wissenschaftlichen Projekten hervorgehen.

Publikation einer virtuellen Forschungsumgebung

Virtuelle Maschinen sind in der Lage, sogenannte virtuelle Forschungsumgebungen (VRE) zu kapseln, die dann interaktiv oder mittels technischer Schnittstellen (API) von Forschenden genutzt werden können. Beispiele hierfür sind beispielsweise (Web-)Services, Content Management Systeme, Datenbanken, Simulationen oder ähnliche Dienste, die im Rahmen von Forschungsprojekten entwickelt wurden. Da jeder Forschungsbereich unterschiedliche Software-Stacks, spezifische Konfigurationsmöglichkeiten und Nutzungsmuster verwendet, ist ein generischer und vollautomatischer Ansatz erforderlich, der im Idealfall alle virtuellen Maschinen (VMs) gleich behandelt. Als zugrunde liegende Erhaltungsstrategie wird Emulation (bzw. Virtualisierung) auf Basis des Emulation-as-a-Service (EaaS)-Frameworks mit einer entsprechenden Integrationsstrategie für generalisierte/normalisierte Disk-Images und VMs eingesetzt (vgl. Rechert et al. 2012, Rechert et al. 2016).

Im Rahmen dieses Prozesses kann ein CiTAR-Publisher ein Festplattenabbild (Disk-Image) einer VM hochladen, das anschließend analysiert und technisch normiert wird, um die Kompatibilität mit der Emulations- oder Virtualisierungsinfrastruktur und deren Langzeitarchivierungsoptionen sicherzustellen. Darüber hinaus muss der Publisher die Maschine und deren Einsatzmöglichkeiten so beschreiben, dass eine zukünftige Nachnutzung (z.B. Inbetriebnahme) einfach möglich ist (s. Abb 1.). Anschließend ist der Publisher in der Lage, die Maschine zu testen und ggf. weiter anzupassen, beispielsweise um die Benutzerfreundlichkeit zu verbessern, Passwörter zu ändern, Anwendungen automatisch zu starten etc. und dann abschließend die korrekte Funktionsweise abzunehmen. Alle Änderungen werden (technisch) dokumentiert und erzeugen versionierte Revisionen der publizierten VM. Schließlich wird der veröffentlichten Maschine eine persistente Kennung von *Handle.net* zugewiesen, die auf eine *Landingpage* der archivierten Maschine zeigt.

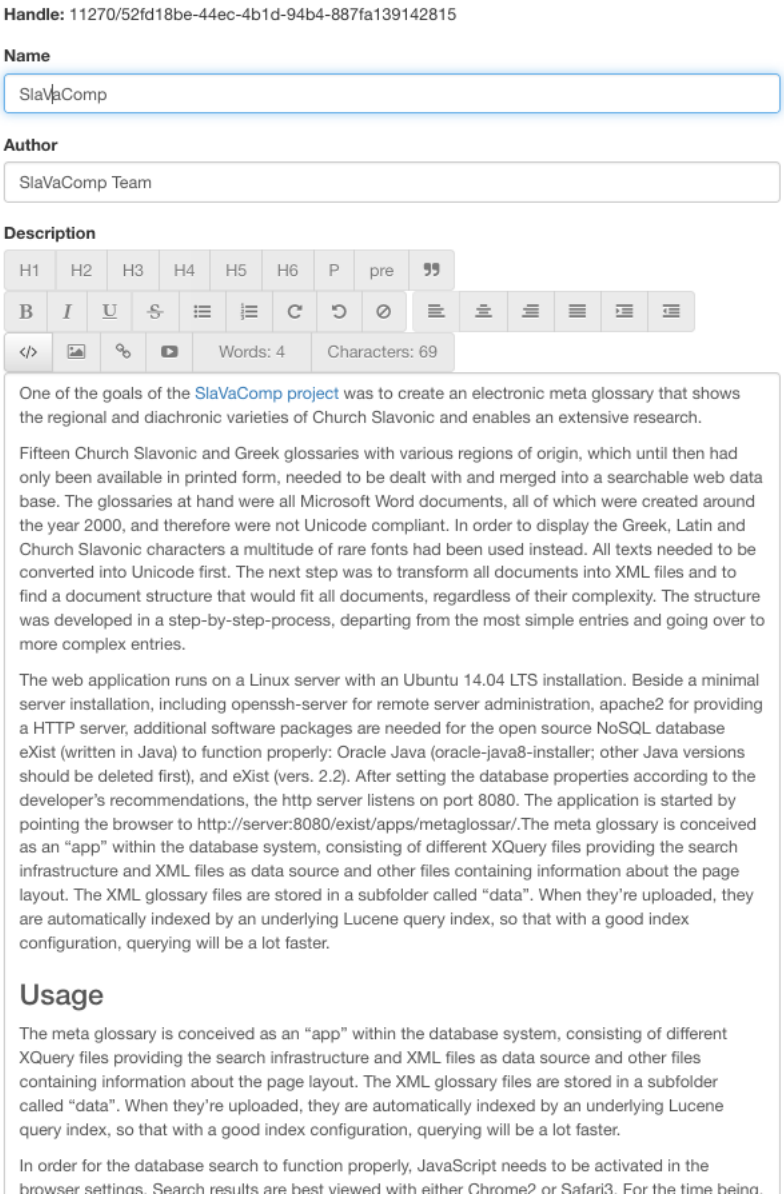


Abb. 1: Beschreibung einer Virtuellen Maschine.

Effizienter Betrieb und Zugriff

Eine veröffentlichte, archivierte und insbesondere zitierbare Forschungsumgebung erfordert auch eine langfristige technische Nutzbarkeit. Das *Handle*, das der archivierten VM zugeordnet ist, leitet den Benutzer zu einer *Landingpage* weiter, die

die virtuelle Umgebung und mögliche Interaktionen beschreibt. Die meisten dieser VMs richten sich an kleine Zielgruppen und werden nicht oft aufgerufen, so dass die VMs erst bei Bedarf gestartet werden.

Die größte Herausforderung betreffend des öffentlichen Zugriffs ist die Sicherheit der archivierten Maschine. Da ein solches System in seinem ursprünglichen Zustand archiviert wurde, ist die VM im Internet gefährdet. Um dennoch einen (sicheren) Netzwerkzugriff zu ermöglichen, werden alle archivierten Maschinen in einem eigenen privaten Netzwerk bereitgestellt. Die Brücke zwischen dem privaten Netzwerk und dem Netzwerk des Benutzers bildet das CiTAR-Gateway. Je nach Sicherheitsanforderungen und/oder Zugangsinfrastruktur stehen derzeit drei verschiedene Netzwerkzugangsoptionen zur Verfügung:

1. *Port forwarding* Für lokale Netzwerkimplementierungen können einzelne Netzwerkports eines emulierten Servers an ein dem Benutzer zugängliches Netzwerk weitergeleitet werden.

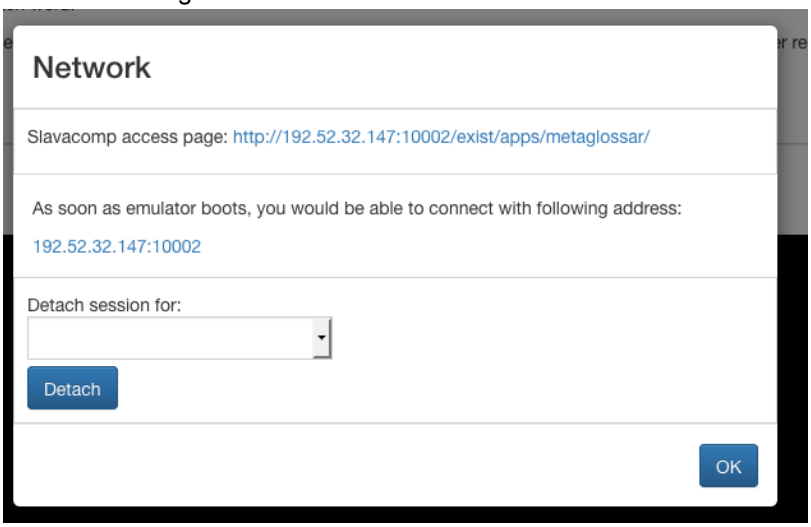


Abb. 2: Das CiTAR SlaVaComp Landingpage-Beispiel (hdl: 11270/52fd18be-44ec-4b1d-94b4-887fa139142815) zeigt ein Fenster mit Verbindungsinformationen an: Der Benutzer kann sich direkt mit der Webanwendung, die auf dem internen Port 8080 läuft, verbinden.

2. *SOCKS* Wenn die Maschine mehrere Dienste über TCP-Ports bereitstellt, kann der SOCKS5-Modus verwendet werden. Das Fenster “Verbindungsinformationen” zeigt die notwendigen Informationen für eine lokale Proxy-Konfiguration an. Zusätzlich kann ein (einfacher!) Passwortschutz konfiguriert werden.

Nach der Einrichtung des SOCKS-Proxy wird das SlaVaComp-Beispiel unter seiner ursprünglichen IP unter http://10.0.2.100 verfügbar sein und die Ports 80 und 8080 freigeben. Beispiel CiTAR SlaVaCom Landingpage (SOCKS mode) (hdl:11270/62fd18be-44ec-4b1d-94b4-887fa139142815)

3. Local-mode In manchen Situationen ist eine vollständig private Verbindung erforderlich. Zusätzlich soll es möglich sein, eine archivierte Maschine in aktuelle (virtuelle) Forschungsumgebungen einzubinden. Hierfür bietet CiTAR einen lokalen Verbindungsmodus an. Der Benutzer muss ein kleines Programm (verfügbar für Linux, MacOS und Windows) installieren, das einen CiTAR-Netzwerk URL-Handler registriert.

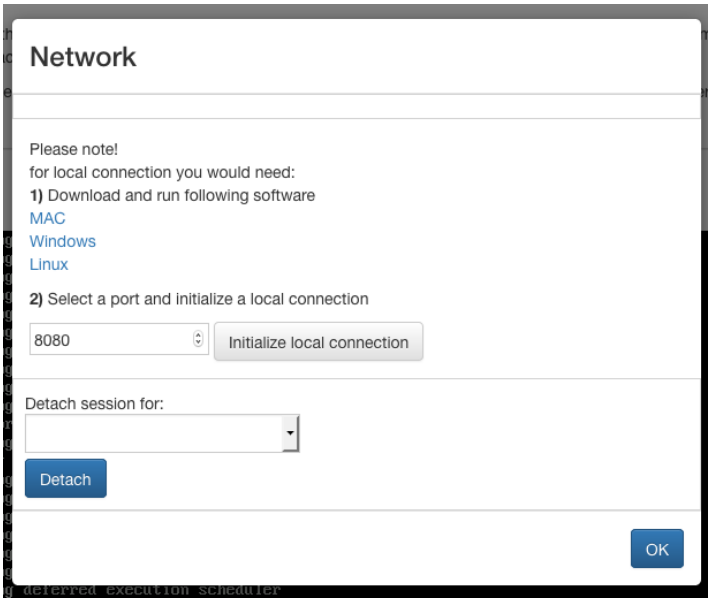


Abb. 3: Einrichten einer privaten Netzwerkverbindung.

Der Benutzer kann dann mit den archivierten Netzwerkumgebungen interagieren, z.B. eine aktuelle Excel-Tabelle mit einem archivierten Datenbankserver verbinden oder im Falle von SlaVaComp mit der REST-Schnittstelle über lokale Programme oder Skripte interagieren. CiTAR SlaVaComp Landigpage (local mode) (hdl: 11270/62ddddddd-44ec-4b1d-94b4-887fa139142815)

Erhaltung von Containern

Die Containertechnologie (auch Betriebssystemvirtualisierung genannt) wurde als leichtgewichtige Alternative zu virtuellen Maschinen entwickelt. Docker und (in den wissenschaftlichen Communities insbesondere) Singularity[7], wurden recht schnell von Forschenden aufgegriffen, da Container in der Lage sind, eine Softwareumgebung, z.B. eine komplexe Softwareinstallation, in eigenständige portable Einheiten zu kapseln. Um wissenschaftliche Methoden in Containern langfristig zugänglich, nutzbar und zitierbar zu machen, muss deren Langlebigkeit als neue digitale Objektklasse gewährleistet sein.

Aufgrund der strengen Isolation (Portabilitätsanforderung) müssen Container eigenständig sein, d.h. alle Softwareabhängigkeiten (Bibliotheken, Anwendungen etc.) müssen innerhalb des Containers aufgelöst sein. Somit ist die Hauptkomponente eines Containers ein in sich abgeschlossenes Dateisystem mit installierten und konfigurierten Softwarekomponenten.

Die einzige verbleibende externe Softwareabhängigkeit ist das zugrunde liegende Betriebssystem, d.h. das Application Binary Interface (ABI) des Betriebssystems. Die zweite Containerkomponente ist die Laufzeitkonfiguration, die Dateisystem-Mappings, z.B. gemeinsame Ordner zwischen Container und Hostsystem, sowie die Definition eines Einstiegspunktes innerhalb des Containers (z.B. ein Skript oder Programm) definiert.

[7] https://www.sylabs.io/docs/.

Struktur eines Containerobjekts

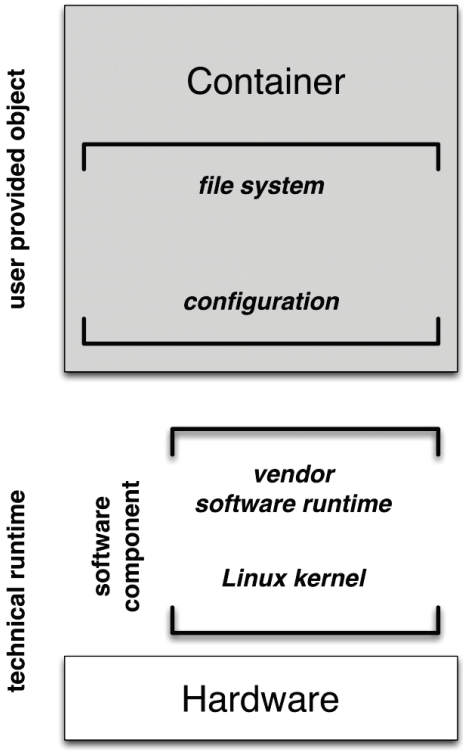


Abb. 4: Container. Eine neue Klasse von digitalen Objekten.

Die technische Laufzeitumgebung eines Containers setzt sich also aus zwei Komponenten zusammen: einer Hardwarekomponente (dem Computer) und einer Softwarekomponente. Im Allgemeinen stellt die Softwarekomponente eine Linux-Installation mit einer installierten und konfigurierten (herstellerspezifischen) Container-Laufzeitumgebung dar. Die Erhaltung von Containern und deren technischen Laufzeit kann somit entkoppelt werden, einerseits durch das Erstellen und Erhalten eines (generischen) Container-Laufzeit-Setups und andererseits dem Ersetzen veralteter Computer durch virtuelle oder emulierte Hardware. Für letzteres wird ebenfalls EaaS als grundlegende Technologie eingesetzt.

Obwohl Container auf den gleichen technischen Grundlagen aufbauen, verwenden die verschiedenen Containerimplementierungen wie Docker oder Singularity unterschiedliche Containerformate und Konfigurationsformen.

Diese Formatunterschiede erfordern in der Regel herstellerspezifische technische Laufzeitumgebungen. Im Idealfall sollte jedoch eine kleine Anzahl von Software-Laufzeitumgebungen ausreichen, um eine große Anzahl von Containern zu betreiben und damit den zukünftigen Verwaltungs- und Wartungsaufwand reduzieren. Da man davon ausgehen kann, dass veröffentliche und insbesondere archivierte Container final sind, werden die Containerformate verallgemeinert, indem herstellerspezifische Komponenten in eine gemeinsame, (teilweise) standardisierte Struktur migriert wurden, so dass aus Erhaltungssicht alle archivierten Containerobjekte dieselben technischen Eigenschaften, d.h. technischen Abhängigkeiten, aufweisen. Darüber hinaus kann man davon ausgehen, dass die Betriebssystemschnittstelle selten inkompatibel wird, so dass nur wenige Variationen solcher Laufzeiten überhaupt notwendig sind.

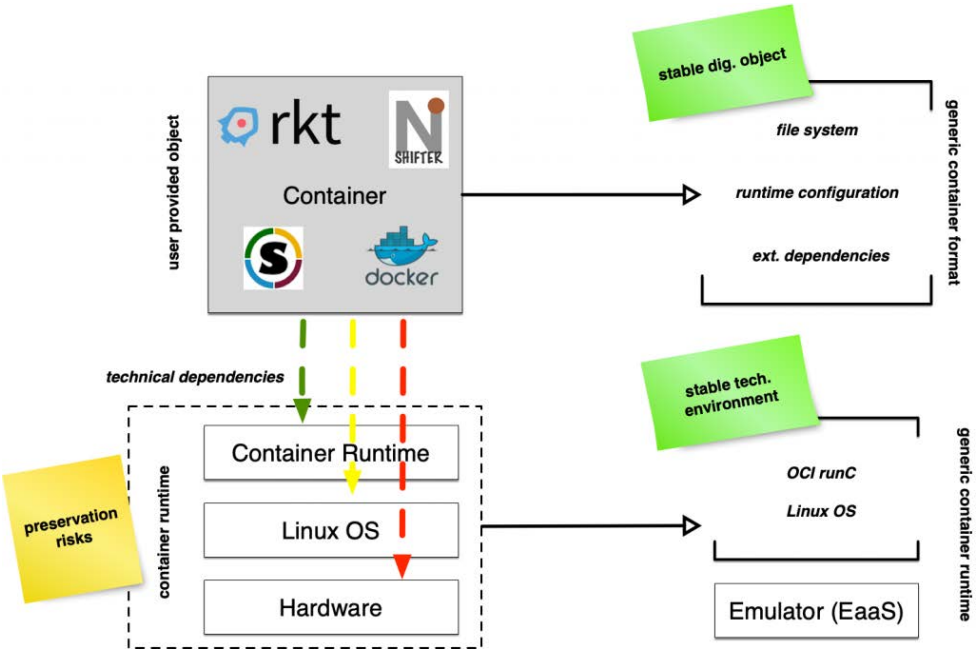


Abb. 5: Aufbau eines Containers.

Publikation zitierfähiger Container

Um zu veranschaulichen, wie Container in CiTAR publiziert und nachgenutzt werden können, soll hier ein Tutorial-Workflow, der von den Entwicklern von OpenMS für die

Workflow-Engine KNIME (Konstanz Information Miner) zur Verfügung gestellt wird, beschrieben werden.

OpenMS ist eine Open-Source-Software zur Verarbeitung von Massenspektrometrie-Daten. Mit KNIME kann man Workflows erstellen, indem man *Knoten* hinzufügt und verbindet. Knoten bieten konfigurierbare Funktionen zur Verarbeitung von Daten im Workflow. Der Benutzer kann zusätzliche Knoten installieren, um KNIME um Funktionen zu erweitern. Die OpenMS-Entwickler haben KNIME-Knoten erstellt, die in dem hier vorgestellten Beispiel verwendet werden.

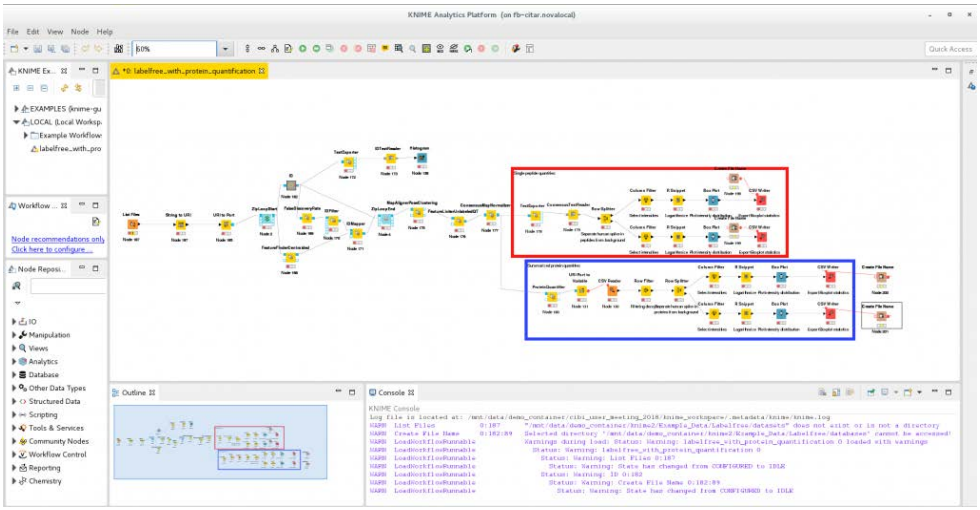


Abb. 6: KNIME-Workflow innerhalb eines Singularity-Containers.

Der in diesem Beispiel erstellte Workflow wird als Singularity Container exportiert und kann anschließend in CiTAR durch den Forschenden selbst importiert werden (vgl. Abb. 6). Hierfür muss der Container zunächst technisch beschrieben werden, insbesondere, wie der enthaltene Prozess gestartet werden soll, wo die Eingabedaten erwartet werden und wo die Ausgabedaten hinterlegt werden. Zusätzlich soll die Funktionsweise des Containers für spätere Nutzer beschrieben werden. Diese Beschreibung bildet dann die Grundlage für die von CiTAR erzeugte Landingpage.

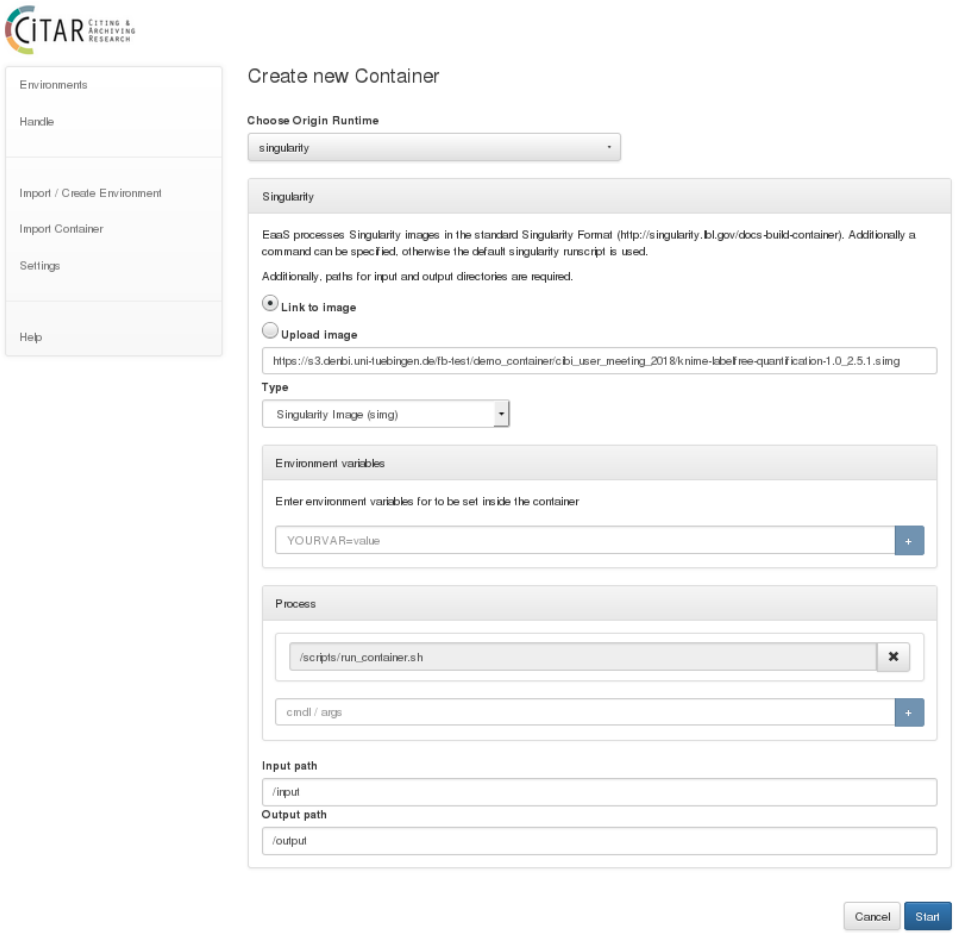


Abb. 7: Erstellung eines neuen Containers und Übermittlung an CiTAR.

Nachnutzung der publizierten Container

Der importierte Container beinhaltet einen Workflow der OMSSA, ein Software Framework zur Massenspektrometrie, die auf einer Sequenzdatenbank basiert. Die von CiTAR erstellte Landingpage (http://hdl.handle.net/11270/188b6194-ea68-4e59-88c2-61d652941e29) beschreibt dessen Funktionsweise und beinhaltet unter anderem einen Link zu einem Beispieldatensatz und eine Beispielausgabe des Workflows. Die notwendigen Datensätze werden nicht von CiTAR selbst verwaltet und sollen als Parameter beim Start des Container als Input-Parameter angegeben werden.

Abb. 8: Reproduktion von Ergebnissen.

Ausblick

Die Erhaltung von softwarebasierten wissenschaftlichen Methoden stellt eine wesentliche Grundlage für reproduzierbare Forschung dar. Im Projekt CiTAR wurden dazu erste Lösungsansätze entwickelt, um Container und andere virtuelle Forschungsumgebungen (u.a. virtuelle Maschinen) langfristig zugänglich zu machen und insbesondere in aktuelle Prozesse zu integrieren. Nachdem die Software fertiggestellt und sie anhand ausgewählter Beispiele erprobt wurde, wird derzeit ein öffentliches Portal entwickelt, das im Laufe des Jahres Interessierten zur Verfügung stehen wird.

Literatur

Garijo, D. / Kinnings, S. / Xie, L. / Xie, L. / Zhang, Y., Bourne, P. E. / Gil, Y. (2013): *Quantifying reproducibility in computational biology: The case of the tuberculosis drugome.* PLoS ONE, 8(11), 1–11. https://doi.org/10.1371/journal.pone.0080278

Brase, J. (2009): “DataCite -- A global registration agency for research data.” In: *Cooperation and Promotion of Information Resources in Science and Technology, 2009. COINFO'09. Fourth International Conference on* IEEE (pp. 257-261).

Meng, Haiyan/Kommineni, Rupa/Pham, Quan/Gardner, Robert/Malik, Tanu / Thain, Douglas: "An invariant framework for conducting reproducible computational science." In: *Journal of Computational Science* 9 (2015): 137-142.

Boettiger, Carl (2015): "An introduction to Docker for reproducible research." In: *ACM SIGOPS Operating Systems Review* 49, 1 (2015): 71-79.

Rechert, Klaus/Valizada, I. /Suchodoletz, D. von/Latocha, J.: "bwFLA – A functional approach to digital preservation." In: *PIK-Praxis der Informationsverarbeitung und Kommunikation* 35, 4 (2012): 259.

Rechert, Klaus/Falcao, P. / Emson, T.. “Towards a Risk Model for Emulation-based Preservation Strategies: A Case Study from the Software-based Art Domain.” In: *Digital Preservation (iPRES) 2016 13th International Conference on*, 139 - 148 (2016).

Rechert, Klaus/Liebetraut, T. / Kombrink S. / Wehrle, D./ Mocken, S. / Rohland, M.: “Preserving Container”. In Kratzke, J. und Heuveline, V. (Hrsg.)*: E-Science-Tage 2017: Forschungsdaten managen,* Heidelberg: heiBOOKS 2017.

The CiTAR Team

Rafael Gieschke, Susanne Mocken, Klaus Rechert, Oleg Stobbe, Oleg Zharkov (University Freiburg), Felix Bartusch (University Tübingen), Kyryll Udod (University Ulm)

A Comprehensive Approach to Support Research – Processes in the CRC 1270 ELAINE

Max Schröder[1,2], Frank Krüger[1], Robert Zepf[2], Ursula van Rienen[3], Sascha Spors[1]

[1]Institute of Communications Engineering, University of Rostock
[2]University Library, University of Rostock
[3]Institute of General Electrical Engineering, University of Rostock

Virtual research environments play a central role in today's interdisciplinary research. They provide a secure and convenient way for collaboration and traceable research. While all-in-one solutions exist that enable researchers to collaborate during the entire research data lifecycle, we argue that a flexible virtual research environment can also be composed of off-the-shelf services. In that vein, we introduce the virtual research environment of the CRC 1270 ELAINE and discuss different implementation aspects.

1 Introduction

Managing digital research artefacts (such as data, models and source code) in a convenient, comprehensive, flexible and secure way is an important issue in today's research (Hanson et al., 2011). Many approaches, however, focus on storing, archiving and publishing only the research data. Other artefacts of the research process such as the model, the source code of the analysis, and the computing environment are often not considered. In the Collaborative Research Centre (CRC) (German: Sonderforschungsbereich) 1270 ELAINE supported by the German Research Foundation (German: Deutsche Forschungsgemeinschaft, DFG), instead, we aim at supporting the overall research process that begins at the definition of the hypothesis and ends at the dissemination and publication of the results and findings. While all-in-one software exists that aims at supporting the entire lifecycle, different reasons prevent researchers from using such solutions. Firstly, such all-in-one solutions are not very flexible when it comes to the individual needs of a multidisciplinary and heterogeneous research group. Secondly, if the software will no longer be developed, great effort is required to re-use and migrate the research artefacts to another solution. Furthermore, hosting all-in-one software at one own's institution, often is not easy when it comes to installation, maintenance, and extension. Finally, researchers are often not allowed to use "Software as a Service" offers that are hosted abroad. Reasons for this are for instance data privacy issues and governmental guidelines.

In the CRC ELAINE, the needs of the researchers range from documenting simulation studies and wetlab experiments, to managing and sharing data and models in different lab environments, to supporting versions, provenance and archival of research artefacts.

In contrast to all-in-one software solutions, we propose a Virtual Research Environment (VRE) that builds upon simple individual services, particularly tailored for the specific needs of each research field. This enables the easy replacement of single services if e.g. requirements change or service development is discontinued. In order to determine the requirements of the individual researchers, we performed a systematic assessment by use of a questionnaire (Krüger and Spors, 2018) in addition to visits of the research groups. For the appropriate selection of each service, we additionally interviewed domain experts that are target users of the service.

With respect to the research artefacts, the research process can be structured into three major tiers: 1. Study planning and data collection, 2. Collaborative modelling and data analysis, and 3. Reproducibility, provenance and archival. For each of these tiers, a large number of simple but specialised Open Source web-services are available. By the use of standardised interfaces, the services can be integrated into a comprehensive, flexible and secure VRE. The VRE for the CRC ELAINE which we currently implement is illustrated in Figure 2. The particular services, e.g. electronic lab notebooks, Jupyter notebooks and Docker are result of the requirements analysis with the researchers.

The next section introduces the general idea of VREs. Section 3 presents the details of the VRE within the CRC 1270 ELAINE followed by a discussion of the VRE.

2 Virtual Research Environments

Candela et al. (2013) define VREs as “innovative, web-based, community-oriented, comprehensive, flexible, and secure working environments conceived to serve the needs of modern science.” Thus, they enable more efficient, open, and reproducible research (Barker et al., 2019). Agreeing with this definition, we would like to add another aspect. The overall research process is driven by research artefacts beginning at the formulation of a hypothesis; ending at the dissemination and publication of the data and source code. Research artefacts in this case are data such as spreadsheets, sensor data and survey results. Furthermore, simulation models, source code, and

images are also research artefacts, e. g., CT scans. However, not all artefacts are digital, e. g., materials, prototypes, and lab notes. A VRE, then, has to incorporate and track the creation, modification, and share of these research artefacts that are increasingly (re-)used in the research lifecycle (Briney, 2015).

Several solutions for specific research fields exists, such as for the structural biology (Morris et al., 2019) and genomics (Codó et al., 2019), astronomy (Herwig et al., 2018), archaeology (Meghini et al., 2017), and earth sciences (Marelli et al., 2016). Especially in highly multidisciplinary research projects, such specifically tailored solutions, however, cannot be employed for all researchers. Thus, a more generally applicable solution is needed. The Open Science Framework[1] is an example for such all-in-one software. Instead of using a single software for the overall life cycle, simple tools tailored for specific use cases integrated into each other are another option. This increases the effort for integration, but at the same time increases flexibility of the VRE. Services can be easily replaced if requirements change or development is discontinued. Khoonsari et al. (2019) follow a similar direction by providing a scalable data analysis platform for applications in metabolomics that can be launched on demand using Docker[2] and Jupyter[3] notebooks.

Central aspects of a VRE are the support for FAIR principles Wilkinson et al. (2016), powerful programming interfaces which are needed in order to integrate the services into the research workflow and existing software systems as well as an effective user interface. Ardito et al. (2016) proposed the creation of a VRE addressing the diversity of users based on the meta-design approach. In order to support the FAIR principles—i. e., findability, accessibility, interoperability, and reusability—Dimitrov and Stoyanov (2018) proposed a discovery service for the CKAN4[4] data repository engine.

[1] see https://osf.io/
[2] see https://www.docker.com/
[3] see https://jupyter.org/
[4] see https://ckan.org/

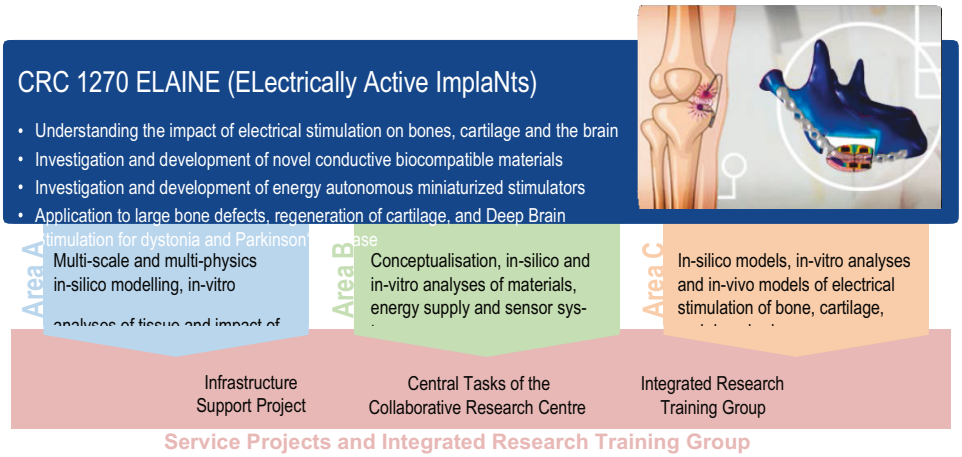


Figure 1: Schematic overview of the research aspects within the CRC 1270 ELAINE

3 The CRC 1270 ELAINE Virtual Research Environment

The next section introduces the research objectives of the CRC 1270 ELAINE in order to better understand the needs and requirements for its VRE. Afterwards, the details of the CRC 1270 VRE are discussed.

3.1 The CRC 1270 ELAINE

A CRC is a large interdisciplinary collaboration project funded by the German Research Foundation (DFG) with a strong focus on fundamental research on a particular topic. Objective of the CRC 1270 ELAINE is to investigate different aspects of electrically active implants and put novel solutions forward. The ultimate goal is to provide implants for electrical stimulation in bone, cartilage, and brain that are energetically autonomous. For this purpose researchers from medicine and biology, electrical and mechanical engineering, material and computer sciences, and physics work in close cooperative on the following aspects:

1. Establish innovative energy autonomous implants that allow a feedback-controlled electrical stimulation,
2. Efficient multi-scale simulation models to enable rapid progress in targeted im- plant improvements and patient-specific therapies, and
3. Analyse the basic mechanisms of electrical stimulation in bone, cartilage and brain, and translate this knowledge into clinical practice.

A schematic overview of these aspects is given in Figure 1.

Typical investigations within the CRC 1270 ELAINE involve experiments and simulations. The results of the experiments are used to calibrate and validate simulation models, whereas the results of the simulation are used to determine optimal parameters for the electrical stimulation of the tissue. As this research is conducted by different research teams at different locations (partly even different cities), the usage of a VRE that enables researchers to share data and results in a traceable way is of central interest. As previously discussed, further aspects of the VRE are usability, security and privacy.

The VRE of the CRC 1270 is also a testbed for novel services that are later de- ployed for all members of the university. To support this target, we are working closely together with the university library and the IT and Media center. To be applicable for all researchers of the university, the VRE service cannot be domain specific.

The VRE of the CRC 1270 ELAINE is built upon a set of microservices, each of them addressing particular objectives of the research data lifecycle. To this end, these services are ordered accordingly, resulting in the three tiers for data collection, data analysis and data archival. A schematic representation of the VRE is given in Figure The selection of the services based on the results of the questionnaire (Krüger and Spors, 2018) that was used as an initial analysis of the requirements within the CRC 1270 and subsequent visits of the research groups. A strong focus during the selection of potential services, however, lay on web based open source software in order to decrease proprietary dependencies and increase convenient usage of the researchers avoiding the install of software on their computers. In the following, an overview of the three tiers is provided, the different aspects are discussed and the specific microservices introduced

3.2 Tier 1: Study Planning and Data Collection

The first tier of the VRE is concerned with the beginning of a new research investigation: the documentation of the research question or hypothesis. Furthermore, also the subsequent documentation of the experiment respectively study as well as the initial data storage has to be afforded at this stage.

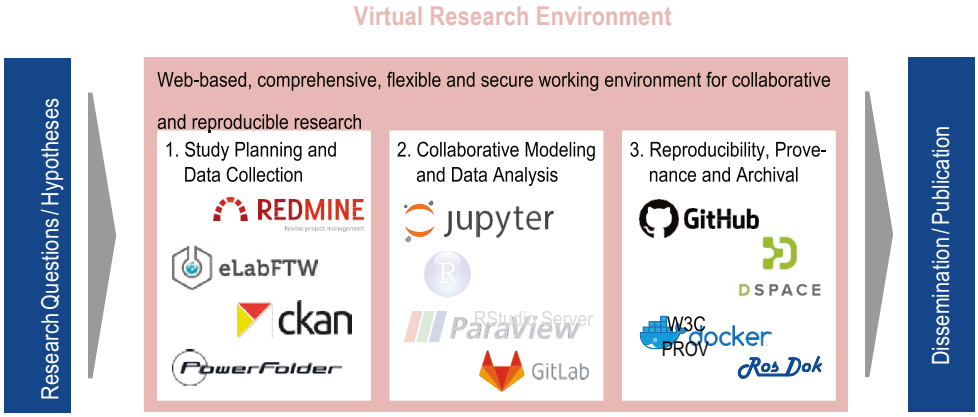


Figure 2: The current VRE in the CRC ELAINE including the main microservices. RStudio Server and ParaView are ideas for further extensions.

General Project Management

A general-purpose project management service is needed for the project management tasks such as central storage of protocols and newsletters, and the documentation of the research questions and hypotheses arising within the CRC. Furthermore, the structure of the CRC should be implementable into this tool so that individual projects of the CRC have particular workspaces that can be shared with other researchers and groups. For the general communication and documentation of the CRC, further workspaces are needed.

In the CRC 1270, we use Redmine[5] for this purpose. Although Redmine was originally aimed at tracking software development, the research project documentation has some similar requirements. Basically, two mechanisms are required: (1) store research artefacts in a versioning system, and (2) store textual documentation. Redmine provides both mechanisms in a simple but functional user interface and provides great community support including a multitude of extensions. Other software that provides similar functionality is e. g., OpenProject[6], basecamp[7], and trac[8].

Experiment Documentation Tool

In laboratory environments such as in the biological and medical research, lab notebooks are common to use in order to document particular experiments

[5] see https://redmine.org/
[6] see https://www.openproject.org/
[7] see https://basecamp.com/
[8] see https://trac.edgewall.org/

respectively studies. Digital solutions for such white paper doc- umentations exists. Their main advantage is searchability and readability of multiple researchers at the same time if configured appropriately. Although general project management tools can be employed for the documentation of laboratory experiments, electronic lab notebook (ELN) solutions might better suited. ELN tools often provide additional features such as materials databases, laboratory management features, and the possibility to define and share Standard Operation Procedures (SOP). Such functionality can also be used to define the set of minimum information that are required to make experiments reproducible (Budde et al., 2019).

For these reasons, we decided to deploy elabFTW[9] as ELN solution in the CRC 1270. This tool is not tailored for a specific domain but rather for the documentation of many kinds of experiments providing a white-paper based documentation style. While the main demand of this tool is for in-vitro and in-vivo documentation, some researchers try this tool for their general research documentation. Other prominent examples of electronic laboratory notebooks include labfolder[10], openBIS[11], and RSpace[12].

Data Storage and Internal Data Sharing

The second major aspect of this tier is the storage and sharing within the research consortium of the raw data created during an experiment respectively a simulation study. Despite the actual storage of arbitrary data formats, several additional requirements have to be met. The provenance of data has to be tracked, which not only includes meta information such as the author of the data and a description; but also revisions of the data and the meta data i. e., versions. The data artefacts have to be findable, accessible, interoperable, and reusable (Wilkinson et al., 2016) for the original creator but also other researchers for whom this dataset is shared. The internal storage platform ideally is integrated into the publication service in order to automatically create a public version of the data set on request. In the CRC 1270 ELAINE, two services for the data storage were selected that com- plement each other: CKAN[13] and PowerFolder[14]. CKAN is a tool that enables the stor- age of meta information along with a set of resources which can be files or urls. The tool then provides a web interface for the search and simple inspection of the data sets.

[9] see https://www.elabftw.net/
[10] see https://www.labfolder.com/
[11] see https://labnotebook.ch/
[12] see https://www.researchspace.com/
[13] see https://ckan.org/
[14] see https://www.powerfolder.com/

It supports a basic versioning of the datasets by providing meta data fields for version information and the possibility of creating relations between datasets. Although this satisfies the requirements for the storage of data from research investigations, the software PowerFolder—a file synchronisation and sharing tool—(along with OnlyOffice[15]) is employed for the collaborative editing of text, spreadsheets and presentations. The PowerFolder service is already deployed by the central IT and Media Center for all members of the university.

3.3 Tier 2: Collaborative Modelling and Data Analysis

After the experimental study has been planned and the raw data collected, the modelling and data analysis phase begins. During this phase, researchers often iteratively develop models or perform data analyses. Different versions of data, models and scripts evolve, which have to be maintained appropriately. Furthermore, during the development and discussion of the data analysis, it is often necessary to share intermediate results and insights. This indicates the need for version control and collaborative environments for modelling as well as data analysis and exploration. Instead of using click-and-point software, the analysis should be developed as source code in order to enable reproducibility. Furthermore, in order to enable other researchers understand the analysis and models, the method of literate programming recently gained attention. This method combines explanatory text and source code in a single document. In the CRC 1270 ELAINE, a Jupyter notebook[16] service is deployed in order to support literate programming. This service provides an interactive web user interface for the development of source code and documentation as well as the execution of the analysis using the computing resources of the server. Similar software include RStudio Server[17] with the R packages knitr[18] or Sweave[19]. The Jupyter service, however, does not track different versions of the model respectively the analysis. To incorporate this functionality in the CRC 1270, a local Gitlab[20] instance is deployed, that provides GIT repository management as core functionality. The Gitlab, furthermore, provides continuous integration functionality that can be used in order to control data quality as well as reproducibility (Pietsch et al., 2019). In combination both services enable to collaboratively work on the data analysis and keep track of modifications and results.

[15] see https://www.onlyoffice.com/
[16] see https://jupyter.org/
[17] see https://www.rstudio.com/products/rstudio/
[18] see https://yihui.name/knitr/
[19] see https://leisch.userweb.mwn.de/Sweave/
[20] see https://about.gitlab.com/

3.4 Tier 3: Reproducibility, Provenance and Archival

This last tier is concerned with the tasks of completing the research investigation artefacts, providing a comprehensive and integrated overview of the investigations, and providing convenient access to them. As such this task is divided into three parts: Reproducibility, Provenance, and Archival.

Reproducibility

To support reproducibility on a technical level a comprehensive doc- umentation, tracking of the research artefacts, and programming of the research analysis has already been discussed. Additionally, it is crucial to document the software stack that is used for the computation of the results i. e., the computing environment. For the comprehensibility of the research investigations it, however, is not enough to provide an executable image of this computing environment. Furthermore, a description of the environment with its software dependencies including the software versions is needed. This enables both easy execution as long as the image is working and a specification enabling to rebuild the computing environment if the virtualisation solution does no longer exist.

For this purpose, containerisation solutions recently are gaining attention. Within the CRC 1270, we use Docker[21] as a containerisation tool. This enables to build the computing environment based on the specification in so-called Dockerfiles. The Docker tool exists for many operating systems providing an easily executable environment on the personal computer of the researchers. Furthermore, continuous integration chains such as in Gitlab support the usage of these container images. Thus, they can be incorporated to automatically check data quality and consistency. Other solutions include Singularity[22].

Provenance

For the reproducibility but also the comprehensibility of the research investigations even after years, it is crucial to document the provenance of research artefacts and their relations (McClatchey, 2018). Provenance information should be integrated into all parts of the VRE so that it is both understandable for other researchers and machine

[21] see https://www.docker.com/
[22] see https://www.sylabs.io/singularity/

interpretable. The latter is crucial in order to provide consistency checks and automatic notifications about new revisions of artefacts.

In the CRC 1270, we incorporate the standardised W3C PROV formalism (Schröder et al., 2019). While a machine interpretable version in the form of turtle and other languages can be maintained, the information can easily be converted into a provenance graph that is understandable by researchers. A prior version of this formalism is known as OPM Provenance Model.

Archival and Publication

A reproducible and comprehensible documentation of the research investigations shall be achieved for a long period of months and years. To support this archival of the research, multiple services need to be employed. All research artefacts need to be archived under consideration of the FAIR principals (Wilkinson et al., 2016). For the archival the same requirements as already for the storage of the artefacts during the research process need to be satisfied. For the archival task we use multiple services within the CRC 1270 depending on the type of artefact. For source code artefacts, we use GitHub[23] in companion with Zenodo[24] as a public archive. Whereas GitHub is a platform for the management of GIT repositories, Zenodo is a public data storage repository for the archival. Zenodo provides DOIs for each version of a data set and can automatically import the resources from a GitHub release. When it comes to datasets the service DSpace[25] comes in and, finally, for the archival of publications, the RosDok[26] platform is used. DSpace is a repository software that enables storage of arbitrary datasets similar to the CKAN data storage, but with a richer customisable meta data set. Similar software includes Dataverse[27] and Invenio[28]. The RosDok service is a local variant of the mycore framework[29], which also is a repository software.

4 Discussion

Beside the service requirements of the VRE, other requirements exists that play an essential role. To those belong convenience, general technical and legal aspects as well as requirements from the particular research artefact types.

[23] see https://github.com/
[24] see https://zenodo.org/
[25] see https://duraspace.org/dspace/
[26] see http://rosdok.uni-rostock.de/
[27] see https://dataverse.org/
[28] see https://invenio-software.org/
[29] see http://www.mycore.de/

Research Artefact Types Not all research artefacts are digital, i. e., a multitude of different media are involved such as prototypes of 3d structures, materials, and sensor data. These different media have to be handled with their corresponding meta information including provenance and version information. Tracking this is essential in order to enable comprehensive understanding of the research experiment. We are still working on an efficient solution tracking real objects and their corresponding digital information.

Convenience Dimensions Usability and training effort are other dimensions that need to be considered. This is highly dependent on the prior knowledge of the re- searchers which are not only different between research fields but also within a re- search group. As such, often a compromise between ease of use, functionality, and training effort has to be found when selecting a service. Often, researchers have to learn many new methods and concepts, especially the concept of versioning is not always clear. Furthermore, the user interfaces are not standardised so that every ad- ditional service requires at least a minimal level learning. A thorough training concept is currently developed and already partly implemented for the researchers in the CRC 1270.

Technical Dimensions Despite the previously discussed aspects of the services, some additional general technical facets were considered: The most important issue is that tools support single-sign-on functionality so that researchers don't need to register for every service individually. Software maintenance costs but also installation and infrastructure maintenance effort are also influencing aspects. Furthermore, the possibility of extending the functionality of a tool is needed in order to adapt the solution to local requirements. A very important issue is security; the services have to enable fine-grained rights management and the possibility to create backups.

Legal Dimensions The legal dimensions for supporting research processes by a VRE are manifold. From a top-down-view, this starts with governmental and institutional laws and guidelines that regulate data handling. Furthermore, especially in biological and medical fields, there are often approvals of experimental procedures that also contain regulations on the handling of research artefacts. Then there are laws about supervising working habits of employees that is technically easy when digital systems are involved. But also the visibility and editability of investigations and research artefacts need to be controlled, especially when it comes to industrial cooperations and

data privacy. Data privacy is also crucial with respect to combining multiple sources of research artefacts which may lead to de-anonymisation of them. There are many more legal aspects involved which we, however, cannot analyse here completely.

5 Conclusion

In this paper we argued that a VREs can support researchers in multiple disciplinary teams at different locations during the entire data lifecycle. We presented the VRE of the CRC 1270 ELAINE. We discussed different services and showed how they can contribute to the individual phases of the data lifecyle: 1. study planning and data collection, 2. collaborative modelling and data analysis, and 3. reproducibility, provenance and archival. In favour of more flexibility, the VRE was built from many specialised services. The requirements were collected by conducting interviews with the participating researchers, other research data management experts and early adoption test phases. While the VRE is still in development, a lot of open issues have to be addressed in future work. This starts by systematically assessing information about acceptance and usability from the researchers but also includes better integration of the particular services into each other as well as into the workflow of the researchers. Also the support for automatic workflows from data recording to analysis, including an assessment of the data quality could be of interest. Finally, as the VRE is envisaged as central data and information storage, methods of automatic data or text mining and information retrieval could be employed in order to provide automatic documentation.

Acknowledgement

This research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the Collaborative Research Centre 1270 ELAINE.

Literatur

Carmelo Ardito, Maria Francesca Costabile, Giuseppe Desolda, Maristella Matera, and Paolo Buono. A meta-design approach to support information access and manipulation in virtual research environments. In *Lecture Notes in Computer Science*, pages 115–126. Springer International Publishing, 2016. doi: 10.1007/978-3-319-50070-6_9.

Michelle Barker, Silvia Delgado Olabarriaga, Nancy Wilkins-Diehr, Sandra Gesing, Daniel S. Katz, Shayan Shahand, Scott Henwood, Tristan Glatard, Keith Jeffery, Brian Corrie, Andrew Treloar, Helen Glaves, Lesley Wyborn, Neil P. Chue Hong, and Alessandro Costa. The global impact of science gateways, virtual research environ ments and virtual laboratories. *Future Generation Computer Systems*, 95:240–248, jun 2019. doi: 10.1016/j.future.2018.12.026.

Kristin Briney. *Data Management for Researchers: Organize, maintain and share your data for research success*. Pelagic Publishing Ltd, 2015.

Kai Budde, Julius Zimmermann, Elisa Neuhaus, Max Schröder, Adelinde Uhrmacher, and Ursula van Rienen. Requirements for documenting electrical cell stimulation experiments for replicability and numerical modeling. In *2019 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, jul 2019. To appear.

Leonardo Candela, Donatella Castelli, and Pasquale Pagano. Virtual research en vironments: An overview and a research agenda. *Data Science Journal*, 12(0): GRDI75–GRDI81, 2013. doi: 10.2481/dsj.grdi-013.

Laia Codó, Genís Bayarri, Javier Alvarez Cid-Fuentes, Javier Conejero, Romina Royo, Dmitry Repchevsky, Marco Pasi, Athina Meletiou, Mark D. McDowall, Fatima Re- ham, José A. Alcantara, Brian Jimenez-Garcia, Jürgen Walther, Ricard Illa, François Serra, Michael Goodstadt, David Castillo, Satish Sati, Diana Buitrago, Isabelle Brun- Heath, Juan Fernandez-Recio, Giacomo Cavalli, Marc Marti-Renom, Andrew Yates, Charles A. Laughton, Rosa M. Badia, Modesto Orozco, and Josep Ll. Gelp´ı and. MuGVRE. a virtual research environment for 3d/4d genomics. apr 2019. doi: 10.1101/602474.

Vladimir Dimitrov and Stiliyan Stoyanov. Solutions for data discovery service in a virtual research environment. *Scalable Computing: Practice and Experience*, 19(2):181–187, may 2018. doi: 10.12694/scpe.v19i2.1350.

B. Hanson, A. Sugden, and B. Alberts. Making data maximally available. *Science*, 331 (6018):649–649, feb 2011. doi: 10.1126/science.1203354.

Falk Herwig, Robert Andrassy, Nic Annau, Ondrea Clarkson, Benoit Côté, Aaron D'Sa, Sam Jones, Belaid Moa, Jericho O'Connell, David Porter, Christian Ritter, and Paul Woodward. Cyberhubs: Virtual research environments for astronomy. *The Astrophysical Journal Supplement Series*, 236(1):2, may 2018. doi: 10.3847/1538-4365/aab777.

Payam Emami Khoonsari, Pablo Moreno, Sven Bergmann, Joachim Burman, Marco Capuccini, Matteo Carone, Marta Cascante, Pedro de Atauri, Carles Foguet, Alejan- dra N Gonzalez-Beltran, Thomas Hankemeier, Kenneth Haug, Sijin He, Stephanie Herman, David Johnson, Namrata Kale, Anders Larsson, Steffen Neumann, Kristian Peters, Luca Pireddu, Philippe Rocca-Serra, Pierrick Roger, Rico Rueedi, Christoph Ruttkies, Noureddin Sadawi, Reza M Salek, Susanna-Assunta Sansone, Daniel
Schober, Vitaly Selivanov, Etienne A Thévenot, Michael van Vliet, Gianluigi Zanetti, Christoph Steinbeck, Kim Kultima, and Ola Spjuth. Interoperable and scalable data analysis with microservices: applications in metabolomics. *Bioinformatics*, mar 2019. doi: 10.1093/bioinformatics/btz160.

Frank Krüger and Sascha Spors. A questionnaire to estimate the needs for research data management. 2018. doi: 10.18453/rosdok_id00002290.

Fulvio Marelli, Mirko Albani, and Helen Glaves. Everest: a virtual research environment for earth sciences. In *EGU General Assembly Conference Abstracts*, volume 18, 2016.

Richard McClatchey. Data provenance tracking as the basis for a biomedical virtual research environment. *CoRR*, abs/1803.07433, 2018. URL http://arxiv.org/abs/ 1803.07433.

Carlo Meghini, Franco Niccolucci, Achille Felicetti, Paola Ronzino, Federico Nurra, Christos Papatheodorou, Dimitris Gavrilis, Maria Theodoridou, Martin Doerr, Dou- glas Tudhope, Ceri Binding, Roberto Scopigno, Andreas Vlachidis, Julian Richards, Holly Wright, Guntram Geser, Sebastian Cuy, Johan Fihn, Bruno Fanini, and Hella Hollander. ARIADNE. *Journal on Computing and Cultural Heritage*, 10(3):1–27, aug 2017. doi: 10.1145/3064527.

Chris Morris, Paolo Andreetto, Lucia Banci, Alexandre M.J.J. Bonvin, Grzegorz Cho- jnowski, Laura del Cano, José Marıa Carazo, Pablo Conesa, Susan Daenke, George Damaskos, Andrea Giachetti, Natalie E.C. Haley, Maarten L. Hekkelman, Philipp Heuser, Robbie P. Joosten, Daniel Kouřil, AlešKřenek, Tomáš Kulhánek, Victor S.Lamzin, Nurul Nadzirin, Anastassis Perrakis, Antonio Rosato, Fiona Sanderson, Joan Segura, Joerg Schaarschmidt, Egor Sobolev, Sergio Traldi, Mikael E. Trellet, Sameer Velankar, Marco Verlato, and Martyn Winn. West-life: A virtual research environment for structural biology. *Journal of Structural Biology: X*, page 100006, feb 2019. doi: 10.1016/j.yjsbx.2019.100006.

Christian Pietsch, Jochen Schirrwagen, and Vitali Peil. Conquaire: Coupling a local gitlab instance with an institutional repository for instant research data publications. 2019. doi: 10.5281/zenodo.2602714.

Max Schröder, Hendrikje Raben, Frank Krüger, Andreas Ruscheinski, Ursula van Rienen, Adelinde Uhrmacher, and Sascha Spors. Provenance patterns in numerical modelling and finite element simulation processes of bioelectric systems. In *2019 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, jul 2019. To appear.

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Apple- ton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard, Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna- Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Vel- terop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR guiding principles for scientific data management and stew ardship. *Scientific Data*, 3:160018, mar 2016. doi: 10.1038/sdata.2016.18

Laufender Betrieb

Integriertes Management und Publikation von wissenschaftlichen Artikeln, Software und Forschungsdaten am Helmholtz-Zentrum Dresden-Rossendorf (HZDR)

Edith Reschke[1], Dr. Uwe Konrad[2]

Helmholtz-Zentrum Dresden-Rossendorf

Zusammenfassung

Mit dem Ziel, das Publizieren von Artikeln, Forschungsdaten und wissenschaftlicher Software gemäß den FAIR-Prinzipien[3] zu unterstützen, wurde am HZDR ein integriertes Publikationsmanagement aufgebaut. Insbesondere Daten- und Softwarepublikationen erfordern die Entwicklung bedarfsgerechter organisatorischer und technischer Strukturen ergänzend zu bereits sehr gut funktionierenden Services im Publikationsmanagement. In der Zusammenarbeit mit Wissenschaftlern des HZDR und internationalen Partnern in ausgewählten Projekten wurde der Bedarf an Unterstützung im Forschungsdatenmanagement analysiert. Darauf aufbauend wurde schrittweise ein integriertes System von Infrastrukturen und Services entwickelt und bereitgestellt. In einer seit Mai 2018 gültigen Data Policy wurden die Rahmenbedingungen und Regelungen sowohl für wissenschaftliche Mitarbeiter als auch für externe Messgäste definiert.

Zusammenfassend werden unsere Erfahrungen im integrierten Publikationsmanagement für Artikel, Forschungsdaten und Forschungssoftware vorgestellt. Es wird ein Ausblick auf die nächsten Schritte und Aufgaben gegeben und Aspekte der Integration im Kontext der europäischen und nationalen Forschungsorganisationen herausgearbeitet.

Abstract

With the aim of supporting the publication of articles, research data and scientific software in accordance with the FAIR principles, an integrated publication management system was established at the HZDR. In particular, data and software publications require the development of needs-based organizational and technical structures in addition to already well-functioning services in publication management.

[1] https://orcid.org/0000-0002-9112-3326
[2] https://orcid.org/0000-0001-8167-9411
[3] https://www.go-fair.org/fair-principles/; 12.04.2019

In cooperation with HZDR scientists and international partners in selected projects, the demand for the support of research data management was analyzed. Building on this, an integrated system of infrastructures and services was gradually developed and made available. In a data policy valid since May 2018, the frameworks and regulations were defined for both scientific staff and external visitors.
In summary, our experiences in integrated publication management for articles, research data and research software are presented. It provides an outlook on the next steps and tasks and identifies aspects of integration in the context of European and national research organizations.

1 Rahmenbedingungen am HZDR – Ausgangssituation, Besonderheiten

Das Helmholtz-Zentrum Dresden-Rossendorf[4] (HZDR) gehört zur Helmholtz-Gemeinschaft Deutscher Forschungszentren und hat neben dem Hauptstandort Dresden weitere Standorte in Freiberg, Leipzig, Schenefeld (bei Hamburg) und Grenoble (Frankreich). Im HZDR wird Spitzenforschung in den Forschungsbereichen Energie, Gesundheit und Materie betrieben. Die Zusammenarbeit mit den Partnerinstitutionen und die Arbeit in internationalen Forschungsprojekten sind am HZDR ein wesentlicher Bestandteil wissenschaftlicher Forschung. In den Labors der acht Forschungsinstitute und an den sieben Forschungsgroßgeräten, zum Beispiel am ELBE-Zentrum für Hochleistungs-Strahlungsquellen und am Hochfeld-Magnetlabor Dresden, haben HZDR-Wissenschaftler[5] und externe Nutzer einzigartige Arbeitsbedingungen und Experimentiermöglichkeiten.

Seit 2011 haben wir am HZDR die Regelung zur „Sicherung guter wissenschaftlicher Praxis und Verfahren zum Umgang mit wissenschaftlichem Fehlverhalten“. Darin ist u.a. definiert, für welche Forschungsdaten die Archivierung Pflicht ist. Seitdem wird auch erfasst, wo und auf welche Art Forschungsdaten archiviert werden, die Grundlage einer Publikation sind. Wie in Abbildung 1 zu sehen ist, steigt der Anteil der zu archivierten Forschungsdaten seitdem kontinuierlich an und betrug zuletzt über 43%.

[4] https://www.hzdr.de/
[5] 2018 ca. 500 Wissenschaftler aus >60 Ländern

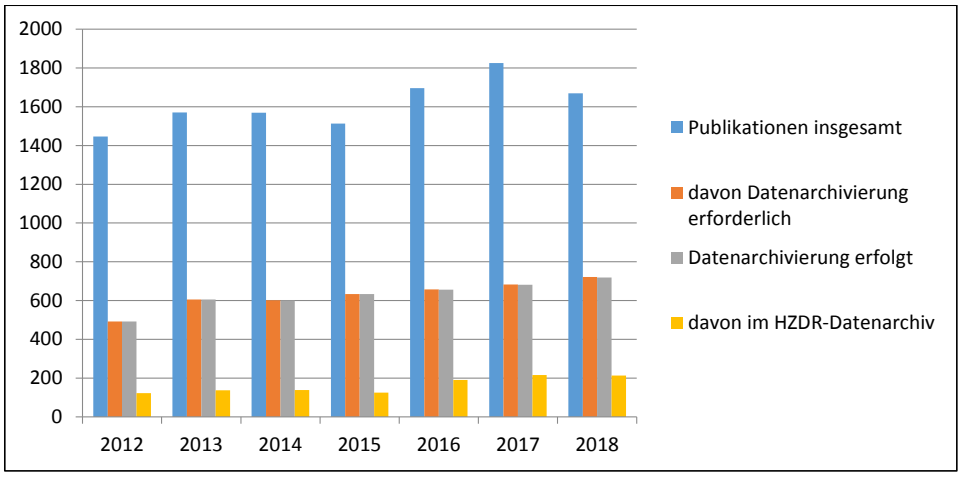


Abbildung 1: Forschungsdatenarchivierung am HZDR 2012-2018

Neben der Archivierung im HZDR-Datenarchiv wurden die Daten oftmals sehr individuell, in externen Repositorien und auch auf persönlichen „Lieblingsplätzen“ archiviert. Nicht immer waren die Langzeitverfügbarkeit und die Wiederauffindbarkeit gewährleistet. Ein wesentlicher Grund dafür waren u.a. auch das Fehlen eines Datenrepositoriums und einer Data Policy im HZDR.

Mit dem Ziel, Forschungsdaten und Forschungssoftware gem. den FAIR-Prinzipien und Open Access zu publizieren, zu archivieren, wieder auffindbar und nachnutzbar zu machen, haben wir unsere Publikationsinfrastruktur um das Datenrepositorium RODARE erweitert, welches auf dem am CERN entwickelten Invenio basiert.

2 Integriertes Publikationsmanagement am HZDR

2.1 Policies und Prozesse

Wissenschaftliche Reputation basiert zu einem bedeutenden Teil auf wissenschaftlichen Publikationen. Publiziert werden Artikel (Texte), Forschungsdaten und Forschungssoftware, wobei Forschungsdaten und Forschungssoftware erst in neuerer Zeit in den Publikationsfokus gerückt sind. Ein wichtiger Aspekt dabei sind die sich schnell verändernden rechtlichen Rahmenbedingungen, die mittelbar den Publikationsprozess beeinflussen. Dazu sind u.a. das Urheberrecht und der Datenschutz zu zählen. Die Internationalität und Vernetzung der Forschung und des Publikationsmarktes sind weitere Faktoren, die es zu berücksichtigen gibt.

Wissenschaftliches Publizieren gemäß den FAIR-Prinzipien erfordert ein effizientes Publikationsmanagement, das drei Komponenten integriert:

- die Bereitstellung der notwendigen technischen Infrastrukturen,
- Services für alle Publikationen (Text, Daten, Software) und
- eine passende Struktur für die Unterstützung und Beratung der Wissenschaftler.

Das integrierte Publikationssystem muss flexibel angepasst werden können und die Akzeptanz durch den Wissenschaftler inkl. der Gäste finden.

Jedes Publikationsmanagement unterliegt natürlich rechtlichen Rahmenbedingungen. Ergänzend zu den allgemein gültigen Gesetzen (Urheberrecht, Datenschutz, u.a.) wurden HZDR-internen Regelungen erstellt:

- „Sicherung guter wissenschaftlicher Praxis und Verfahren zum Umgang mit wissenschaftlichem Fehlverhalten“
- „Publikationsordnung“
- „Terms and Conditions for the Storage, Access and Curation of Research Data“ (HZDR Data Policy)
- “Terms and Conditions for User Access to the Experimental Facilities”

Diese Regelungen sind eine solide Grundlage für die Entwicklung eines rechtssicheren integrierten Publikationsmanagements.

Verantwortlich für die Entwicklung und den Aufbau eines Publikationsmanagements am HZDR ist die Zentralabteilung Informationsdienste und Computing, zu der auch die Bibliothek gehört. Beteiligt sind natürlich weitere Abteilungen wie die wissenschaftliche Infrastruktur, die Stabsabteilung Recht und Patente und die Stabsabteilung Programmplanung und Internationale Projekte.

Abbildung 2: Bereiche, die ins Publikationsmanagement integriert werden müssen

2.2 Integrierte Systemlandschaft

Die Publikationsdatenbank (ROBIS) am HZDR gibt es seit 1993. Bis Mitte 2018 war sie für folgende Services verfügbar:

- Nachweis aller Publikationen seit 1993
- Verwertungsprüfung
- Open Access Finanzierung
- Genehmigungsworkflow
- Erstellung von institutionellen und persönlichen Publikationslisten
- Statistik, wissenschaftliches Controlling
- Langzeitarchivierung der elektronischen Volltexte
- Verlinkung von Publikationen zur Datenarchivierung

ROBIS ist eine Eigenentwicklung, basierend auf der Nutzung einer Oracle-Datenbank mit Weboberfläche, die für Publikationen im Laufe der Zeit immer wieder an den lokalen Bedarf angepasst und weiterentwickelt wurde. Da die Anforderungen an ein Datenrepositorium insbesondere unter technischem Aspekt sich nicht mit einer

Erweiterung von ROBIS abdecken ließen, wurde entschieden, ein separates Datenrepositorium aufzubauen.

In Interaktion mit ROBIS sollen alle notwendigen Publikationsservices für Textpublikationen, Forschungsdaten und Forschungssoftware bereitgestellt werden.

Auf der Basis der Invenio[6] Frame work library wurde das Rossendorf Data Repository RODARE entwickelt. RODARE ist auf GitLab[7] eingestellt und steht unter der GNU General Public License by the Free Software Foundation (GPLv3) zur Nachnutzung zur Verfügung. Seit Mitte 2018 ist die Verlinkung der Workflows des Publikationsrepositorium ROBIS und des Forschungsdatenrepositoriums RODARE aktiv geschalten (siehe Abb. 3). Die Autoren können selbst entscheiden, in welchem Repositorium sie mit ihrer Metadatenerfassung und Datenspeicherung beginnen. Die beiden Repositorien übergeben die „Kern"- Metadaten automatisch an das jeweils andere Repositorium, so dass doppeltes Datenerfassen ausgeschlossen wird. Es bleibt die Notwendigkeit, dass die Autoren die Repositorien-spezifischen Einträge anschließend vornehmen müssen.

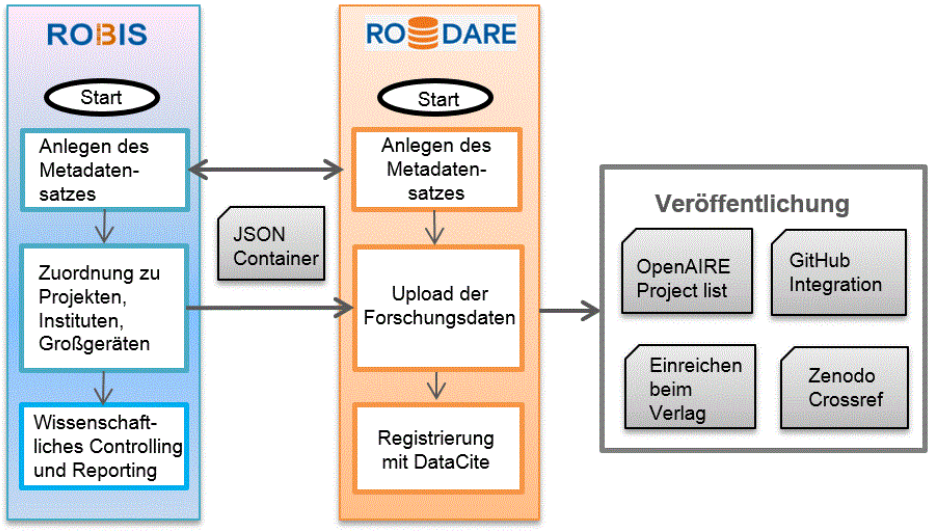


Abbildung 3: Interaktionen zwischen ROBIS und RODARE

[6] https://inveniosoftware.org/ - 09.04.2019
[7] https://gitlab.hzdr.de/rodare/rodare - 09.04.2019

2.3 Daten und Software

Daten und Software gehören seit jeher zum wissenschaftlichen Publikationsprozess. Nur führten sie lange Zeit ein Schattendasein. Mit der Forderung nach guter wissenschaftlicher Arbeit rückte das Publizieren von Forschungsdaten und wissenschaftlicher Software in seiner Bedeutung auf zu den wissenschaftlichen Textpublikationen. Forschungsdaten und Forschungssoftware sind sehr oft die Grundlage einer Publikation. Sie liegen daher viel eher vor als die Textpublikation oder auch als eigenständige Publikation. Deshalb war die Entwicklung eines Datenrepositoriums zusätzlich zu unserem Publikationsrepositorium notwendig. Forschungsdaten und Software bedingen spezifische Metadaten, die sich nur zum Teil mit den Metadaten in ROBIS decken. In RODARE wurden zusätzliche Funktionen gebraucht; zum Beispiel der Umgang mit großen bzw. komplexen Datenstrukturen, die automatische Registrierung einer DOI, APIs zu OpenAIRE, GitHub bzw. GitLab, die Zuordnung von Lizenzen u.a.m.

Um die im Punkt 2.2. dargestellten Funktionen zu erfüllen, ist es notwendig, dass alle Publikationen, also Textpublikationen, Forschungsdaten und Forschungssoftware im Publikationsrepositorium ROBIS nachgewiesen werden. Mit der „Verzahnung“ von ROBIS und RODARE haben wir es vermeiden können, dass die Autoren ihre Daten mehrfach eingeben müssen.

2.4 Publikationssupport aus einer Hand

Die Bereitstellung von Publikationsinfrastruktur und zugehörenden Services sind Bausteine im Gesamtkonzept „Wissenschaftliches Publizieren“. Akteure sind insbesondere die wissenschaftlichen Institute (als Betreiber der Forschungsgeräte), die Abteilung Informationsdienste und Computing und die Bibliothek. Die Koordinierung der publikationsspezifischen, abteilungsübergreifenden Aktivitäten aller Akteure ist die Kernaufgabe des Data Librarian in der Bibliothek. Der Data Librarian ist für folgende Services zuständig:

- Anwenderberatung zu ROBIS und RODARE
- Unterstützung für die Erstellung und Pflege der Daten-Management-Pläne
- Qualitätssicherung und Kuratierung der Metadaten
- Beratung zu Open Access und Lizensierungsoptionen.

Entscheidend, ob diese Angebote von den potentiellen Autoren angenommen werden oder nicht, sind die Qualität der Information, Beratung und Unterstützung, idealerweise projektbegleitend vom Projektstart bis zur Veröffentlichung der Publikationen.

2.5 Weiterentwicklung von Services im digitalen Forschungsprozess

Entscheidend für die Akzeptanz der Prozesse und Infrastrukturen durch die Nutzer ist eine optimale Einbettung in die tägliche Arbeit und die Vermeidung einer aufwändigen Eingabe und Pflege von Zusatzdaten. Dies trifft ganz besonders für den Lebenszyklus von Daten- und Softwarepublikationen zu, da diese oft in komplexen Strukturen vorliegen, in größeren Teams über einen langen Zeitraum bearbeitet werden und der Aufwand für Reproduzierbarkeit sowie Nachnutzbarkeit wesentlich größer als für Textpublikationen ist.

Deshalb ist es erforderlich, Informationen und Metadaten frühzeitig, kontinuierlich und systematisch in einem durchgängigen digitalen Forschungsprozess zu erfassen. Dieser beginnt bereits mit der Anmeldung eines Experimentes / Projektes für ein wissenschaftliches Experiment, z.B. an einem Forschungsgerät. Am HZDR erfolgt diese Anmeldung über das System GATE, über das der Nutzer per Webportal den Zugang beantragt und nach einem Reviewprozess im Erfolgsfall bekommt. Bereits in dieser frühen Phase wird die Data Policy verbindlich eingeführt und es werden viele Metadaten abgefragt, die für die Daten-Management-Pläne und schließlich für alle Publikationsarten von Interesse sind und automatisiert verarbeitet werden können.

Während der Zeit der wissenschaftlichen Experimente spielen digitale Laborbücher für Forschungsdaten eine wichtige Rolle. In diesen erfassen die Forscher Informationen und Bilder zu verwendeten Geräten, Proben, Einstellungen und besonderen Vorkommnissen. Werden diese Daten strukturiert erfasst und gespeichert, können sie später die grundlegenden Metadaten (z.B. nach DataCite Schema 4.x) ergänzen und sachlich erschließbar machen. Am HZDR wird dafür derzeit in Pilotprojekten das System openBIS (Quelle ETH Zürich) eingesetzt.

Für Forschungssoftware sind die Entwicklungsumgebungen die Plattform kontinuierlicher Änderungen und Tests. Am HZDR wird den Entwicklern dafür zukünftig das lokale System GitLab angeboten, das seinerseits mit der führenden Entwicklungsplattform GitHub (betrieben von Microsoft) vernetzt ist. Von diesen Systemen kann die Publikation direkt in das Publikationssystem RODARE erfolgen, so dass dem Softwareentwickler keine großen zusätzlichen Aufwände entstehen.

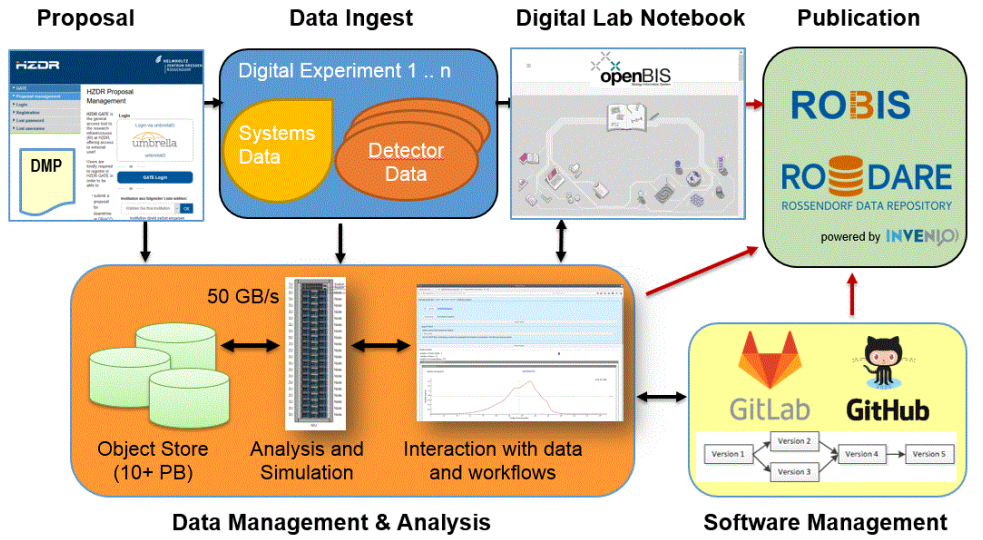


Abbildung 4: Daten und Software im Lebenszyklus am HZDR

In Abbildung 4 ist gut zu erkennen, dass man im Publikationsprozess den gesamten Lebenszyklus aller Arten von Publikationen mit allen beteiligten Partnern betrachten muss, um eine Qualität im Sinne von FAIR zu erreichen. Sobald man eine Komponente vernachlässigt hat der publizierende Wissenschaftler erhebliche Aufwände zur Eingabe und Pflege von Metadaten und wird Argumente finden, dies nicht zu tun. Man kann die Qualität von textuellen Veröffentlichungen noch vor der Publikation verbessern, bei Daten und Software ist es am Ende oft zu spät.

Einen großen Einfluss auf die Nachhaltigkeit der Ergebnisse haben auch Metadaten-Standards, die von Geräteherstellern, der Industrie und der Wissenschaft selbst entwickelt werden. An dieser Stelle spielen insbesondere die Research Data Alliance (RDA) und die geplanten community-spezifischen Instanzen der Nationalen Forschungsdaten-Infrastruktur (NFDI) in Deutschland eine besondere Rolle. Hier haben die führenden Wissenschaftler und großen Daten-Infrastrukturen der jeweiligen Communities eine große Verantwortung zur Vereinheitlichung von Prozessen und Metadaten sowie zur Aus- und Weiterbildung der Forscher.

Die Helmholtz-Gemeinschaft trägt dieser Verantwortung mit dem „Helmholtz Information & Data Science Framework“ Rechnung, in dem Konzepte zur gemeinschaftsweiten Zusammenarbeit entwickelt werden.

Beginnend 2019 werden 5 Serviceplattformen eingerichtet, die diese Prozesse in allen Zentren vorantreiben und unterstützen werden[8]. Das sind:

- die „Helmholtz Data Science Academy (HIDA)
- die “Helmholtz Federated IT Services (HIFIS)“
- das Helmholtz Metadata Center (HMC)
- die Helmholtz Artificial Intelligence Coordination Unit (HAICU)
- die Helmholtz Imaging Plattform (HIP)

Speziell die ersten drei genannten Plattformen sind darauf ausgerichtet, die Prozesse der Aus- und Weiterbildung, der Softwareentwicklung und der Metadatendefinition und -erfassung zu entwickeln und zu vereinheitlichen.

Fazit

Um wissenschaftliche Artikel, Forschungsdaten und Forschungssoftware nach den FAIR Prinzipien publizieren zu können, muss der digitale Forschungsprozess als Ganzes betrachtet und optimiert werden. Durch die nahtlose Integration der Teilprozesse und IT-Infrastrukturen vom Antrag eines Forschungsvorhabens bis zur Langzeitarchivierung kann der publizierende Wissenschaftler von aufwändigen Metadateneingaben entlastet und gleichzeitig die Qualität und Nachhaltigkeit der Ergebnisse gesichert werden.

Diese ganzheitliche Betrachtungsweise erfordert ein enges Zusammenwirken der beteiligten Servicebereiche; zum Teil müssen die klassischen Aufgabenzuordnungen überwunden werden. Am HZDR wurde dieser Prozess 2015 mit der Fusion des Rechenzentrums mit der Bibliothek begonnen und seitdem konsequent umgesetzt. Die technische Integration der Systeme (GATE, Invenio, openBIS, GitHub, GitLab u.a.) ist ein erheblicher Aufwand, bedeutet aber für den Nutzer einen spürbaren Mehrwert. Die positiven Rückmeldungen aus der Wissenschaft bestätigen uns, dass wir auf dem richtigen Weg sind.

[8] https://www.helmholtz.de/forschung/information_data_science/helmholtz_inkubator/ - 01.04.2019

Referenzen

Barillari, C. (2018): openBIS – an open resource for academic laboratories. Vortrag, https://os.helmholtz.de/fileadmin/user_upload/os.helmholtz.de/Workshops/eln18hzi_barillari.pdf - 15.04.2019

Konrad, U.; Görzig, H.; Juckeland, G. (2018): Forschungsdatenmanagement am Helmholtz-Zentrum Dresden-Rossendorf und am Helmholtz-Zentrum Berlin (RDM@DB). Konferenzbeitrag, DOI: 10.14278/rodare.62

Konrad, U. (2017): Research Data Management to increase research quality. Konferenzbeitrag, DOI: 10.5281/zenodo.1040289

Konrad, U. (2017): Wissenschaftliche Software – Anspruch und Realität im Forschungsprozess. Konferenzbeitrag, 10.5281/zenodo.1040289

Verstetigung zentraler Dienstleistungen zum Forschungsdatenmanagement – Das Kompetenzzentrum Forschungsdaten der Universität Bielefeld

Maik Stührenberg[2], Johanna Vompras[1], Nils Hachmeister[3], Edith Rimmert[1], Cord Wiljes[1], Jochen Schirrwagen[1], Dirk Pieper[1]

[1] Universitätsbibliothek Bielefeld
[2] Bielefelder IT-Servicezentrum (BITS)
[3] Bielefeld Center for Data Science (BiCDaS)

Zusammenfassung

Die Digitalisierung der Forschung und der damit verbundene Wandel im Umgang mit Forschungsdaten stellen die wissenschaftlichen Disziplinen sowie die institutionellen Serviceeinrichtungen der Hochschulen vor neuartige Herausforderungen. Neben der Verankerung in den einzelnen wissenschaftlichen Disziplinen ist eine institutionelle Grundversorgung, die Infrastruktur und Beratungsangebote umfasst, von zentraler Bedeutung für ein effektives Forschungsdatenmanagement. An der Universität Bielefeld wurde hierfür das *Kompetenzzentrum Forschungsdaten* eingerichtet, an dessen Beispiel in diesem Beitrag die Aufgaben, Angebote und zukünftige Entwicklung des institutionellen Forschungsdatenmanagements dargestellt und diskutiert werden sollen.

1 Forschungsdaten und Forschungsdatenmanagement

Forschungsdaten bilden die Grundlage wissenschaftlicher Erkenntnis. Die Erhebung von Forschungsdaten ist oftmals mit hohem technischem und personellem Aufwand verbunden. Die Digitalisierung der Forschung und das mit ihr verbundene exponentielle Wachstum von Forschungsdaten stellt die wissenschaftlichen Disziplinen sowie die institutionellen Serviceeinrichtungen wie Rechenzentren und Bibliotheken vor neuartige Herausforderungen. Darüber hinaus setzen

Forschungsförderer wie das BMBF[1], die DFG[2], die Europäische Kommission[3] oder der

European Research Council[4] zunehmend ein institutionell abgesichertes Forschungsdatenmanagement als ein Kriterium für die positive Begutachtung von Anträgen voraus. Neben der Verankerung in den einzelnen wissenschaftlichen Disziplinen ist die Grundversorgung, die Infrastruktur und Beratungsangebote umfasst, von zentraler Bedeutung für ein effektives Forschungsdatenmanagement. Zahlreiche Hochschulen haben daher in den vergangenen Jahren bereits Serviceeinrichtungen zum Forschungsdatenmanagement aufgebaut[5] oder richten diese derzeit ein. Im Folgenden soll die Geschichte des Kompetenzzentrums Forschungsdaten der Universität Bielefeld, die aktuellen Herausforderung und zukünftige Pläne vorgestellt werden.

2 Historie des Forschungsdatenmanagements an der Universität Bielefeld

Die Universität Bielefeld hat den Wert eines Forschungsdatenmanagements frühzeitig erkannt und entsprechende Weichen gestellt. Die wichtigsten Phasen und Meilensteine werden in Abbildung 1 dargestellt und im Folgenden beschrieben.

[1] Vgl. beispielhaft die 2017 erschienene „*Richtlinie zur Förderung von Forschung zu Digitalisierung im Bildungsbereich*“, in der als besondere Zuwendungsvoraussetzung die Umsetzung des Forschungsdatenmanagements im Antrag darzulegen ist und begutachtet wird: https://www.bmbf.de/foerderungen/bekanntmachung-1420.html

[2] „*DFG-Merkblatt Sonderforschungsbereiche*“: http://www.dfg.de/formulare/50_06/50_06_de.pdf

[3] “Guideline Open Access to publications and research data: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

[4] “*Guidelines on the Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020*”: http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

[5] Für ein Verzeichnis s. https://www.forschungsdaten.org/index.php/FDM-Kontakte_an_Hochschulen

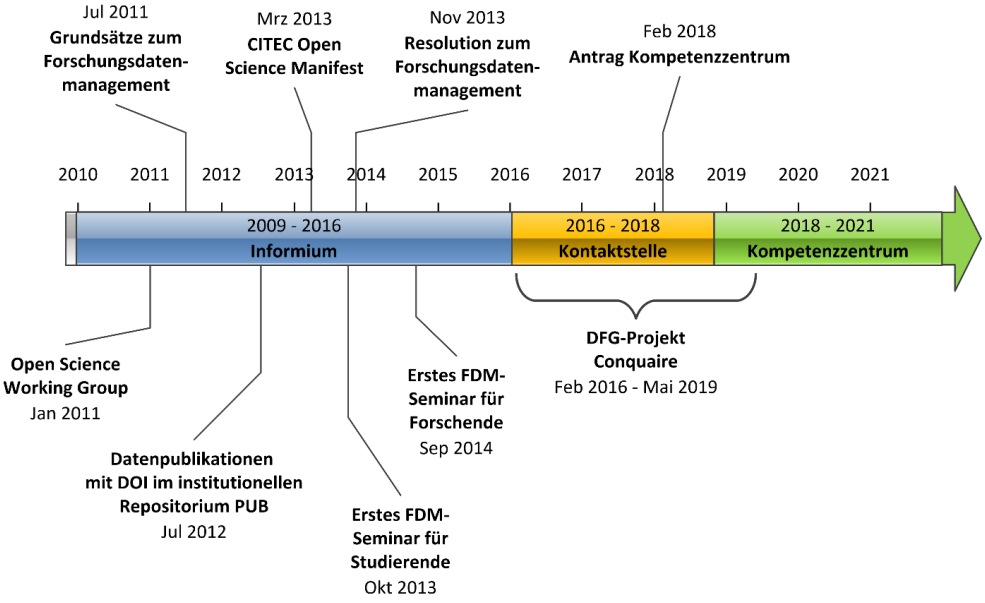


Abbildung 1: Zeitlicher Ablauf der Entwicklung des Kompetenzzentrums Forschungsdaten

Pilot-Initiative Informium (2009-2015)

Im Rahmen des von der Universität geförderten Pilotprojekts *Informium* wurden die Voraussetzungen für den Aufbau von hochschulweiten Services für das Forschungsdatenmanagement geschaffen (Vompras, 2011). Diese gründen sich auf den Pfeilern *Policy*, durch die Verabschiedung der „*Grundsätze zu Forschungsdaten an der Universität Bielefeld*“[6] (2011) und der bundesweit ersten „*Forschungsdatenresolution*“[7], *Support*, z.B. durch die Einrichtung einer „Kontaktstelle Forschungsdaten“ und *technischer Infrastruktur*, z.B. durch Datenpublikationen im institutionellen Repositorium[8]. Diese ersten Aktivitäten für die Errichtung eines nachhaltigen Service-Angebots wurden durch verschiedene Entwicklungen weiter gestärkt:

- Gründung der *Open Science Working Group*[9], die das Ziel verfolgt, den Zugang zu wissenschaftlichen Erkenntnissen an der Universität Bielefeld weiter zu verbessern.

[6] https://www.uni-bielefeld.de/forschungsdaten/fdm-bi/grundsaetze/
[7] https://www.uni-bielefeld.de/forschungsdaten/fdm-bi/resolution/
[8] https://pub.uni-bielefeld.de/data
[9] http://uni-bielefeld.de/open-science/oswg/

- Einrichtung einer Forschungsdatenmanagement-Stelle am *Cluster of Excellence „Cognitive Interaction Technology* (CITEC)[10] im Dezember 2011, sowie der Veröffentlichung des dem „*CITEC Open Science Manifests*“[11] im März 2013.
- Start der am *Center for Biotechnology* (CeBiTec) angesiedelte Koordination des „German Network for Bioinformatics Infrastructure – de.NBI“[12]
- Aufbau einer virtuellen Forschungsumgebung im Rahmen des Teilprojekts INF für den SFB 882 inklusive der Wahrnehmung zentraler Servicefunktionen für den FDM-Bereich (Friedhoff et al. 2013)

Kontaktstelle Forschungsdaten (2016-2018)

Im Januar 2016 konnte durch die Etablierung einer „*Kontaktstelle Forschungsdaten*“ (angesiedelt an der Universitätsbibliothek) eine erste Verstetigung des Engagements für ein hochschulweites Forschungsdatenmanagement erreicht werden.

Im Februar 2016 startete in Kooperation mit dem Exzellenzcluster CITEC das DFG-geförderte Projekt *Conquaire*[13] (*Continuous quality control for research data to ensure reproducibility*) mit dem Ziel, die Qualität von Forschungsdaten und die Reproduzierbarkeit von Forschungsergebnissen zu verbessern (Ayer 2017). Conquaire entwickelte einen webbasierten Service, der Forschende bei der Erstellung und Versionierung von Daten (einschließlich Code, z.B. für die Analyse verwendete Skripts) und bei der Wiederverwendung ihrer Daten unterstützt.

Die Kontaktstelle beteiligte sich darüber hinaus am Teilprojekt INF[14] des SFB 1288, „*Praktiken des Vergleichens*“ und vernetzte sich über landesweite (Expertengruppe FDM der Digitalen Hochschule NRW[15]), nationale (DINI/nestor AG Forschungsdaten[16]) und internationale Kooperationen (OpenAIRE[17], Research Data Alliance[18], Data Documentation Initiative (DDI)[19], Knowledge Exchange[20]).

[10] https://www.cit-ec.de/en/open-science/
[11] https://cit-ec.de/de/open-science-manifest
[12] https://www.denbi.de
[13] http://uni-bielefeld.de/conquaire/
[14] http://www.uni-bielefeld.de/sfb1288/projekte/inf.html
[15] https://www.dh-nrw.de/handlungsfelder/forschung/forschungsdatenmanagement/
[16] https://dini.de/ag/dininestor-ag-forschungsdaten/
[17] https://www.openaire.eu/
[18] https://rd-alliance.org/
[19] https://www.ddialliance.org/
[20] http://knowledge-exchange.info/projects/project/research-data

3 Von der Kontaktstelle zum Kompetenzzentrum

Auf Grund steigenden Beratungsbedarfes, der kontinuierlich zunehmenden Aufwände für die Weiterentwicklung der FDM-Angebote und dem Ziel, diese zu verstetigen, wurde deutlich, dass eine Aufstockung der personellen Ressourcen erforderlich sein würde. Die Erfahrungen aus der Arbeit der Kontaktstelle Forschungsdaten hatten gezeigt, dass Forschungsdatenmanagement eine zeitaufwändige Querschnittsaufgabe darstellt, die das Zusammenspiel verschiedener Akteurinnen und Akteure erfordert. Um dem wachsenden Bedarf an technischen Anforderungen der Forschenden besser gerecht werden zu können, wurde das Bielefelder IT Service-Zentrum (BITS)[21] mit in die Planung des neu zu gründenden Kompetenzzentrums eingebunden. Ebenso beteiligt wurde der Geschäftsführer des *Bielefeld Center for Data Science* (BiCDaS)[22], um mögliche inhaltliche Überschneidungen entsprechend würdigen zu können.

Um den gestiegenen Bedarf quantitativ zu untersuchen und damit eine mögliche Ressourcenplanung zu erleichtern, wurden verschiedene Bedarfserhebungen an der Universität Bielefeld durchgeführt und auf bereits vorhandene Untersuchungen an anderen Einrichtungen zurückgegriffen:

So wurde in der Fakultät für Psychologie und Sportwissenschaft in enger Abstimmung mit den Forschenden 2016/2017 ein Anforderungsprofil für die Erstellung eines FDM-Servers für neurophysiologische Daten (EEG, fMRT) angefertigt.

Auf Basis dieses Profils konnten beispielhafte Anforderungen an das FDM aus Sicht einer konkreten Fachdisziplin an der Universität Bielefeld formuliert werden. Zu den dort formulierten Anforderungen gehörten eine Rechte-Verwaltung unter Einbindung des zentralen Identity-Managements der Universität mit der Möglichkeit, Forschungsdaten rollenbasiert einzelnen Nutzenden und Gruppen zugänglich zu machen, die Beschreibung von Datensätzen mittels standardisierter Metadaten, die Aggregation von Forschungsdatensätzen in größeren Sammlungen sowie deren Versionierung, und die automatisierte (oder workflow-gesteuerte) Konvertierung von Forschungsdatensätzen in verschiedenen Formaten.
Darüber hinaus fand unter den OWL-Hochschulen Hochschule für Musik (Detmold), Technische Hochschule Ostwestfalen-Lippe, FH Bielefeld, Universität Bielefeld sowie

[21] https://www.uni-bielefeld.de/bits/
[22] https://www.uni-bielefeld.de/datascience/

der Universität Paderborn im Jahr 2017 eine Umfrage statt, die deutlich machte, wie heterogen die Erstellung, Verwendung und Publikation von Forschungsdaten in den jeweiligen Disziplinen ist. Dabei wurde deutlich, dass ein universitätsweit zu nutzendes Angebot an Beratung und weiteren Dienstleistungen zumindest grundlegend mit einer Vielzahl von Datenformaten und Arbeitsweisen in den jeweiligen Fachrichtungen vertraut sein muss, um eine erkennbare Unterstützungsleistung bieten zu können. Die Vernetzung einer zentralen Anlaufstelle mit Ansprechpersonen in den einzelnen wissenschaftlichen Disziplinen erscheint hier als ein viel versprechender Weg.

Die Analyse der zum Zeitpunkt der Antragstellung bereits im institutionellen Repositorium PUB[23] gespeicherten Forschungsdaten bestätigte die Befunde. Bei den im Dezember 2017 untersuchten 205 Forschungsdatensätzen[24] (mit insgesamt 1393 Dateien) reichte die Bandbreite der untersuchten MIME-Types von klassischen Office-Dateiformaten wie Word, Excel, CSV und PDF über Formate zur Speicherung von Audio- und Video-Inhalten bis hin zu Archivformaten und einer vollständigen Virtual Appliance – insgesamt wurden 35 verschiedene MIME-Types verwendet. Ein ähnliches Bild ergab die Auswertung der Dateigrößen: von einer nur 89 Byte großen Datei bis hin zu über 14 GB war das gesamte Spektrum vertreten.

Neben der Berücksichtigung konkreter Anforderungen und real vorliegender Forschungsdaten an der Universität Bielefeld wurden auch bereits vorhandene externe Umfragen ausgewertet, um die Bedarfe zur Einrichtung eines Kompetenzzentrums Forschungsdaten zu ermitteln. Hierzu wurden die Ergebnisse einer Umfrage unter 123 Professorinnen und Professoren sowie 267 wissenschaftlicher Mitarbeiterinnen und Mitarbeiter an der Humboldt-Universität zu Berlin aus dem Oktober 2013 (Simukovic et al. 2013) sowie einer analogen Erhebung aus dem Jahr 2014 an der Christian-Albrechts-Universität Kiel (Stüve et al. 2014) herangezogen. Auch hier zeigte sich die bereits beobachtete Heterogenität in Fragen der verwendeten Dateiformate (mit Unterschieden bzgl. strukturierter Daten) sowie der Dateigrößen (bis hin in den TB-Bereich). Als konkrete Handlungsanweisung ließen sich die Aussagen der Forschenden über den Speicherort ihrer Forschungsdaten auffassen: ein nicht unerheblicher Anteil an Forschungsergebnissen wird auf privaten oder dienstlichen Rechnern gespeichert, zentrale Server (inkl. einer üblicherweise vorhandenen

[23] https://pub.uni-bielefeld.de
[24] Zum Stand April 2019 sind 254 Publikationen in PUB als Forschungsdaten gekennzeichnet. Eine aktualisierte Auswertung ist in Planung.

Backup-Strategie) wurden erst an dritter Stelle genannt. Oftmals fehlten Kenntnisse über vorhandene Datenrepositorien, teilweise mangelte es den Befragten aber auch an Vertrauen diesen gegenüber. Die Einrichtung eines zentralen Dienstleistungszentrums muss also – neben der Bereitstellung von Infrastrukturen und Beratungen – auch für die Bekanntheit entsprechender Angebote sorgen.

Im Ergebnis Anforderungsanalyse wurde ein Konzept *Einrichtung eines Kompetenzzentrums* Forschungsdaten erarbeitet. Dieser Antrag wurde im Mai 2018 durch das Rektorat der Universität positiv Bielefeld positiv beschieden, so dass das Kompetenzzentrum im November 2018 seine Arbeit aufnehmen konnte.

4 Das Kompetenzzentrum Forschungsdaten

Die Mitarbeitenden des Kompetenzzentrums setzen sich zusammen aus Kolleginnen und Kollegen der Universitätsbibliothek sowie des Bielefelder IT-Servicezentrums. Personell wurde darauf geachtet, eine Kontinuität zu bestehenden Aktivitäten im Bereich des Forschungsdatenmanagements zu erhalten, was dadurch gewährleistet werden konnte, dass neben der bisherigen Stelleninhaberin der Kontaktstelle Forschungsdaten auch der ehemalige Forschungsdatenmanager des CITEC für die Aufgabe gewonnen werden konnte.

Zum Start des Kompetenzzentrums Forschungsdaten wurden auf einem Kick-Off Workshop mit Stakeholdern an der Universität die Aufgaben und Ziele des Kompetenzzentrums definiert, dessen Ergebnis in Abbildung 2 dargestellt ist und im Folgenden beschrieben werden.

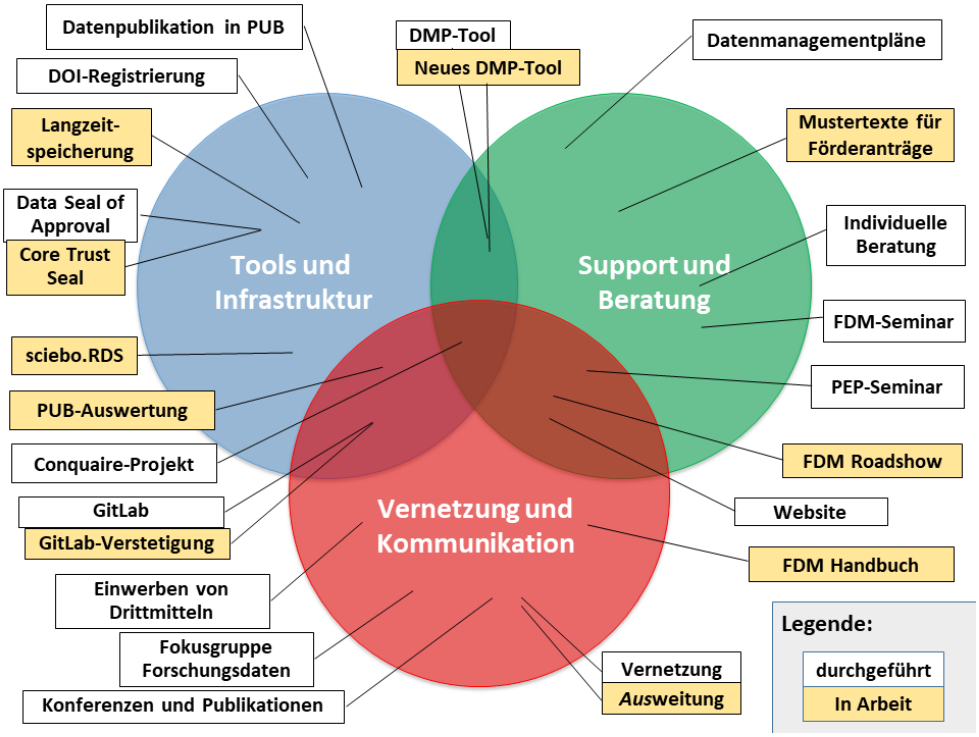


Abbildung 2: Übersicht über vorhandene und geplante Aktivitäten des Kompezenzzentrums Forschungsdaten

4.1 Support und Beratung

Das Bratungsangebot des Kompetenzzentrums Forschungsdaten umfasst alle Fragen zum Forschungsdatenmanagement, inbesondere:

- *Erstellen von Datenmanagementplänen*: In Kooperation mit dem Dezernat FFT (Forschungsförderung & Transfer) der Universität werden Antragstellende und Projektverantwortliche über Empfehlungen von Förderern und Wissenschaftsorganisationen zu Aspekten im Bereich Forschungsdaten informiert und bei der Erstellung von Forschungsdatenmanagementplänen unterstützt

- *Rechtliche Aspekte des Forschungsdatenmanagements*: Gerade bei der qualitativen Forschung werden Beratungen im Bereich Datenschutz und Persönlichkeitsrecht der Befragten von Forschenden nachgefragt. Dabei sind sowohl Fragen der rechtlich einwandfreien Erhebung aber auch der Nachnutzung bestehender Forschungsdaten relevant. Diese werden in Abstimmung mit dem Justitiariat bzw. der Datenschutzbeauftragten der Universität bearbeitet.

- *Forschungsdatenpublikation*: Mit dem institutionellen Repositorium PUB besitzt die Universität Bielefeld ein etabliertes System zur Verzeichnung, Archivierung, Verlinkung und Präsentation von Publikations- und Forschungsdaten. Forschende werden bei der Publikation und Registrierung von Forschungsdaten in PUB (oder in disziplinären Repositorien) unterstützt.

- *Auswahl technischer Infrastruktur*: Forschung findet immer häufiger über verschiedene Arbeitsgruppen (und damit Standorte) verteilt statt. Daher werden Angebote zum sicheren Austausch und ggf. der kollaborativen Bearbeitung von Forschungsdaten benötigt. Aktuell verwendet die Universität Bielefeld die NRW-weite Sync-&-Share-Lösung sciebo zu diesem Zweck und beteiligt sich an der Weiterentwicklung sciebo.RDS[25], um das Werkzeug noch besser auf die Bedürfnisse der Forschenden abzustimmen. Abgesehen davon spielen aber auch Fragen nach Backup-Lösungen oder der Langzeitsicherung eine große Rolle in Beratungsgesprächen. Hierzu wurden exemplarische Prozesse im BITS konzipiert, um Forschenden eine transparente Speicherung und Langzeitsicherung im Zuge der Nachweispflicht anbieten zu können.
- *Schulungen und Fortbildungen*: Regelmäßige Schulungen für Wissenschaftlerinnen und Wissenschaftler sowie Informationen über sonstige Aus- und Weiterbildungsmaßnahmen im Bereich FDM sollen das Bewusstsein für die Thematik stärken.
    - *„Einführung in das Forschungsdatenmanagement“ (für Forschende)*
      Das Seminar vermittelt einen kompakten Überblick über Anforderungen und Angebote des Forschungsdatenmanagements. Es richtet sich an Forschende, die sich über den qualitätsbewussten Umgang mit Forschungsdaten und über die Forschungsdaten-Angebote der Universität Bielefeld informieren möchten.
    - *„Erstellung eines individuellen Forschungsdatenmanagementplanes“*
      Ziel des Workshops ist es, dass die Teilnehmerinnen und Teilnehmer unter Anleitung einen individuellen Datenmanagementplan für ein eigenes, aktuelles Forschungsprojekt verfassen. Hierfür wird, nach einer thematischen Einleitung, das DMP-Tool RDMO genutzt.
    - *„Forschungsdaten und Software unter Kontrolle mit GitLab"*
      Dieses Seminar bietet eine praktische Einführung in die Arbeit mit GitLab und der zugrundeliegenden Software Git.
    - *Seminar "Research Data Management"* (für Studierende)
      Das Seminar bietet eine Einführung in Motivation, Herausforderungen und Lösungen des Forschungsdatenmanagements (Wiljes 2019). Studierende werden mit den Grundlagen des Forschungsdatenmanagements und seiner Bedeutung für gute wissenschaftliche Praxis vertraut gemacht und erwerben einen Überblick über die organisatorischen, technischen und rechtlichen Aspekte des Forschungsdatenmanagements. Das Seminar vermittelt Strategien und Werkzeuge zur effizienten Dokumentation, Sicherung, langfristigen Archivierung, Publikation und Erschließung von Forschungsdaten.
    - *Roadshow Forschungsdatenmanagement (für Arbeitsgruppen)*
      Angebote im Bereich Forschungsdaten werden in den AG-Kolloquien der Arbeitsgruppen vorgestellt. Die inhaltliche Ausgestaltung erfolgt abgestimmt auf die thematischen Bedarfe der jeweiligen Arbeitsgruppe.

Abbildung 3 und Abbildung 4 stellen die zeitlichen Aufwände des Kompetenzzentrums für Aktivitäten in Bereich Support und Beratung im Zeitraum November 2018 bis Mai 2019 dar.

[25] http://www.research-data-services.org

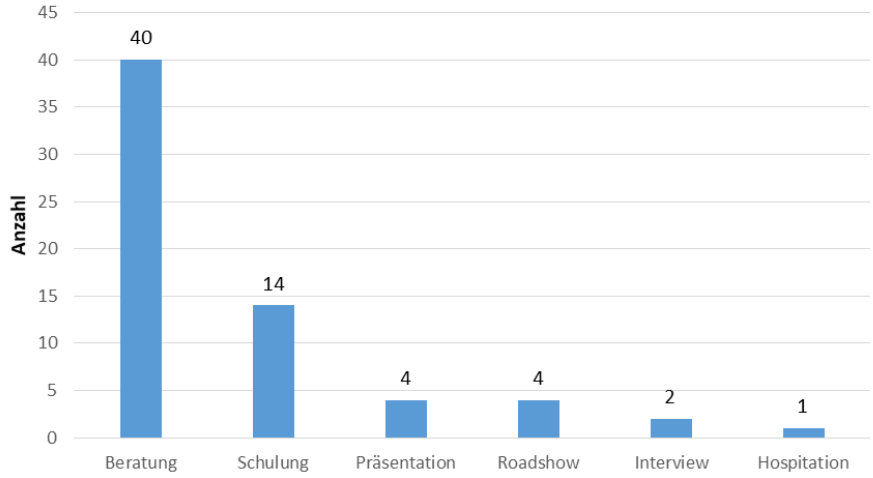


Abbildung 3: Anzahl Support-Aktivitäten des Kompetenzzentrums Forschungsdaten im Zeitraum 11/18-03/19

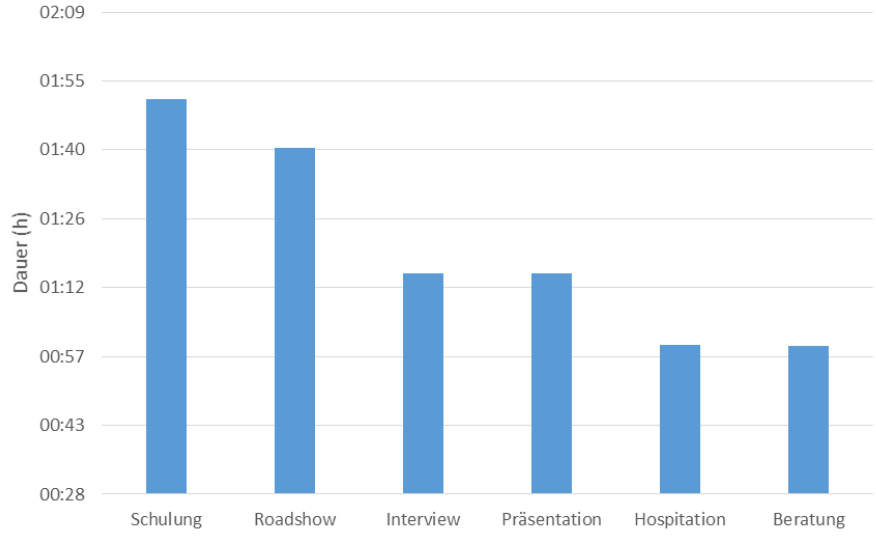


Abbildung 4: Durchschnittliche Dauer der Support-Aktivitäten des Kompetenzzentrums Forschungsdaten im Zeitraum 11/18-03/19

4.2 Tools und Infrastruktur

Das Beratungsangebot wird ergänzt durch eine Reihe von Tools, die universitätsweit Forschenden zur Verfügung gestellt werden, um sie bei der Arbeit mit Forschungsdaten zu unterstützen.

- *Institutionelles Forschungsdaten-Repositorium PUB*
  Das institutionelle Repositorium PUB[26] ermöglicht Forschenden der Universität Bielefeld seit Juli

[26] https://pub.uni-bielefeld.de/data/

2012 die Publikation von Forschungsdaten unter Vergabe von DOIs (Digital Object Identifier). Seitdem wurden mehr als 250 Datenpublikationen auf PUB veröffentlicht. PUB war bereits nach den Kriterien des Data Seal of Approval[27] als vertrauenswürdiges Forschungsdatenarchiv zertifiziert. Da dieses mit Ende 2018 durch das Core Trust Seal[28] abgelöst wurde, wird derzeit eine Zertifizierung unter den erweiterten Kriterien angestrebt.

- *GitLab*
  Die Universitätsbibliothek nutzt seit 2013 für eigene Softwareprojekte eine selbstgehostete Instanz des webbasierten Versionierungstools GitLab[29], die nun allen Mitarbeitenden der Universität zur Verfügung steht. Angebunden an das zentrale Identity Management können Forschende mit ihren gewohnten Accounts nicht nur Programm-Quelltexte sondern auch weitere Forschungsdaten unter einer leicht zugänglichen Oberfläche versioniert verwalten und darüber hinaus die vorhandenen Werkzeuge zur Projektverwaltung wie Wiki oder Issue Tracker nutzen, um untereinander Aufgaben zuzuweisen oder offene Punkte besser nachverfolgen zu können. Es ist geplant, die aktuelle GitLab-Installation in Kooperation mit dem BITS noch weiter auszubauen und die UB vom Betrieb zu entlasten. Verstetigung der GitLab-Instanz. Aktuell wird die GitLab-Instanz von der Universitätsbibliothek betrieben. Neben der Übernahme des technischen Betriebs durch das BITS bestehen aktuell Bestrebungen, den Dienst auch für eine Reihe weiterer Universitäten zur Verfügung zu stellen. In einem ersten Schritt denkbar ist die Anbindung bereits vorhandener dezentraler GitLab-Instanzen und die institutsübergreifende Verknüpfung von hier gespeicherten Ressourcen. Ebenso ist der Austausch von Knowhow bezüglich des Betriebs sowie Best-Practices für die Anwendung angedacht.
- *Anforderungserhebung für sciebo.RDS*
  Die Universitäten Münster und Duisburg Essen wollen im Projekt sciebo.RDS („sciebo Research Data Services – Forschungsdatenmagementdienste und -werkzeuge für Wissenschaftler“) die bestehende Sync-&-Share-Lösung sciebo um Werkzeuge erweitern, die das Arbeiten mit Forschungsdaten unter einer einheitlichen Oberfläche erleichtern. Die Universität Bielefeld hat sich in einem Letter of Intent dazu bereit erklärt, Anforderungen (geistes-)wissenschaftlicher Disziplinen an eine solches Werkzeug zu erheben und in den Entwicklungsprozess mit einfließen zu lassen. Darüber hinaus soll die Anbindung an das Repositorium PUB sowie GitLab vorangetrieben werden, um Forschenden eine einheitliche Oberfläche anbieten zu können.
- *Datenmanagemenplan-Tool*
  Auf Basis des Content Management Systems Drupal wurde ein DMP-Tool entwickelt, über das Forschende online Datenmanagementpläne für Forschungsprojekte und Förderanträge erstellen können. Dieses wird derzeit durch das von der DFG geförderte Open Source-Werkzeug RDMO[30] (*„Research Data Management Organiser“)* ersetzt, das es Wissenschaftlerinnen und

  Wissenschaftlern mittels einer webbasierten Oberfläche die Erstellung von Datenmanagementplänen erlaubt und das bereits von einer Reihe von Forschungseinrichtungen

[27] Vgl. https://www.datasealofapproval.org/
[28] Vgl. https://www.coretrustseal.org/about/
[29] Vgl. https://about.gitlab.com/
[30] Vgl. https://rdmorganiser.github.io/

- deutschlandweit eingesetzt wird. An der Universität Bielefeld soll RDMO das bisher eingesetzte und selbst entwickelte DMP-Werkzeug ersetzen, um von der zentralen Entwicklung zu profitieren und Forschenden ein einheitliches Werkzeug an die Hand geben zu können. Auch hier soll die Anbindung an das zentrale Identity Management für einen möglichst leichten Einstieg sorgen.

4.3 Vernetzung und Kommunikation

Mit dem Relaunch der Website[31] inkl. Umzug auf das zentral eingesetzte CMS der Universität Bielefeld im Frühjahr 2019 wurde das *Kompetenzzentrum Forschungsdaten* universitätsweit einer breiteren Öffentlichkeit bekannt gemacht.

Forschungsdatenmanagement als Querschnittsaufgabe berührt viele Bereiche einer Universität. Aus diesem Grund arbeitet das Kompetenzzentrum Forschungsdaten eng mit verschiedenen lokalen Akteuren zusammen, dazu zählen das Dezernat FFT (Forschungsförderung & Transfer) für die Beratung bei Drittmittelanträgen, sowie das Justiziariat und die Datenschutzbeauftragte bei rechtlichen Fragen.

Etablierung von AnsprechpartnerInnen in den Fakultäten

Auf Grund der Vielzahl an Forschungsdisziplinen an der Universität Bielefeld ist darüber hinaus auch eine Vernetzung mit interessierten PartnerInnen in den jeweiligen wissenschaftlichen Fachrichtungen angestrebt. Damit soll das Bewusstsein für das Management von Forschungsdaten nicht nur stärker in den jeweiligen Disziplinen verankert werden sondern auch auf Besonderheiten des jeweiligen Faches besser eingegangen werden können.

Vernetzung mit externen Akteuren

Darüber hinaus wird eine Vernetzung auf regionaler, nationaler und ggf. auch internationaler Ebene mit anderen Dienstleistungszentren im Bereich Forschungsdatenmanagement angestrebt. Erste Kontakte bestehen noch aus der Arbeit der Stelleinhaberin der Kontaktstelle Forschungsdaten und werden weiterhin gepflegt und ausgebaut.

Knowledge Base/FAQ für Forschende

Es soll eine Knowledge Base erstellt werden, in der wiederkehrende Fragen rund um das Forschungsdatenmanagement in kurzen Textabschnitten oder in Form einer FAQ

[31] http://uni-bielefeld.de/forschungsdaten/

behandelt werden. Diese soll nicht die Beratungssituation ersetzen sondern kann der Vorbereitung auf Gespräche dienen. In einem späteren Schritt ist die Ausarbeitung zu einem FDM-Handbuch (in Form eines „lebendigen Dokuments“) avisiert.

5 Fazit und Ausblick

Die Etablierung eines dauerhaften institutionellen Angebots zum Forschungsdatenmanagement erfordert eine Vielzahl an Maßnahmen auf verschiedenen Ebenen: eine individuelle Beratung von Forschenden, neue technische Dienste sowie eine umfassende Kommunikation, die erfolgreich neue Angebote bekannt macht und zuverlässig die Bedarfe der Forschenden ermittelt. Daher ist eine enge Kooperation der bestehenden Abteilungen, insbesondere Bibliothek und Rechenzentrum sowie bestehenden Beratungsangeboten z.B. zu Forschungsförderung essenziell.

Mit steigenden Anforderungen und zunehmender Bekanntheit und Akzeptanz des Angebots bei den Forschenden wachsen auch die Aufwände. Ein besonders hoher Beratungsbedarf, der zudem hohe Steigerungsraten aufweist, besteht bei Fragen zum Datenschutz. Welche personellen Ressourcen von den Institutionen langfristig für das Forschungsdatenmanagement aufzuwenden sind, ist daher derzeit erst begrenzt abzuschätzen.

Forschungsdatenmanagement ist ein noch recht junges Thema das sich extrem dynamisch entwickelt. Daher bieten sich agile Methoden des Projektmanagements an, die eine kontinuierliche Anpassung an neue Erfordernisse und Erkenntnisse ermöglichen.

Dienstleistungen im Forschungsdatenmanagement sind dann besonders erfolgreich, wenn es gelingt, forschungsnah und individuell zu beraten. Hierfür spielt insbesondere der Aufbau von Kompetenzen im Forschungsdatenmanagement durch Fortbildungen und Train-The-Trainer Angebote, z.B. FDMentor (Dolzycka 2019) eine wichtige Rolle.

Über die Beratung von Forschenden in der eigenen Institution hinaus gewinnt die institutionsübergreifende Vernetzung der ForschungsdatenmanagerInnen zunehmend an Bedeutung. Durch den Austausch mit KollegInnen über Erfahrungen und Lösungen an andern Institutionen, bilden sich gemeinsame Standards heraus und werden Synergien ermöglicht.

Von den Forschungsförderern wird die Schaffung einer gemeinsamen Infrastruktur, insbesondere der nationalen Forschungsdateninfrastruktur (NFDI)[32] und der European Open Science Cloud (EOSC)[33] vorangetrieben, deren erfolgreicher Auf- und Ausbauelementar auf eine enge Verbindung zu den Forschenden und ihren konkreten Bedarfen angewiesen. Hierbei übernehmen die ForschungsdatenmanagerInnen vor Ort an den institutionellen Beratungseinrichtungen eine entscheidende Brückenfunktion, indem sie Bedarfe der Forschenden aggregieren sowie neue, institutionsübergreifende Angebote bekannt machen und vermitteln.

Literatur

Vompras, J., Schirrwagen, J., & Horstmann, W. (2011). Die Bibliothek als Dienstleister für den Umgang mit Forschungsdaten. In S. Schomburg, C. Leggewie, H. Lobin, & C. Puschmann (Eds.), *Digitale Wissenschaft: Stand und Entwicklung digital vernetzter Forschung in Deutschland* (pp. 101-106). Köln: hbz. http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/Tagung_Digitale_Wissenschaft.pdf

Friedhoff, S., Meier zu Verl, C., Pietsch, C., Meyer, C., Vompras, J., & Liebig, S. (2013). *Social Research Data: Documentation, Management, and Technical Implementation within the SFB 882* (SFB 882 Working Paper Series, 16). Bielefeld: DFG Research Center (SFB) 882 From Heterogeneities to Inequalities. https://pub.uni-bielefeld.de/publication/2560035

Paul-Stüve, Thilo, Rasch, Georg, Lorenz, Sören. Ergebnisse der Umfrage zum Umgang mit digitalen Forschungsdaten an der Christian-Albrechts-Universität zu Kiel (2014). Zenodo, 2015: http://dx.doi.org/10.5281/zenodo.32582

Simukovic, Elena; Kindling, Maxi; Schirmbacher, Peter (2013): Forschungsdaten an der Humboldt-Universität zu Berlin. Bericht über die Ergebnisse der Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin: https://edoc.hu-berlin.de/bitstream/handle/18452/14220/22YavRAszVauc.pdf?sequence=1

Schirrwagen, J., Cimiano, P., Ayer, V., Pietsch, C., Wiljes, C., Vompras, J., & Pieper, D. (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

Ayer, V., Pietsch, C., Vompras, J., Schirrwagen, J., Wiljes, C., Jahn, N., & Cimiano, P. (2017). Conquaire: Towards an architecture supporting continuous quality control to ensure reproducibility of research. *D-Lib Magazine*, *23*(1/2). doi:10.1045/january2017-ayer

Wiljes, C., & Cimiano, P. (Accepted, 2019). Teaching Research Data Management for Students. *Data Science Journal.* https://pub.uni-bielefeld.de/record/2934474

Dolzycka, D., Biernacka, K., Helbig, K., & Buchholz, P. (2019). Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement (Version 2.0). Zenodo. doi:10.5281/zenodo.2581292

[32] https://www.dfg.de/foerderung/programme/nfdi/
[33] https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

Der Umgang mit heterogenen (Forschungs-)daten an einer wissenschaftlichen Bibliothek – Use Cases und Erfahrungen aus technischer und nichttechnischer Sicht an der Universität Wien

Susanne Blumesberger, Raman Ganguly

ZID Universität Wien

Zusammenfassung

Forschungsdatenmanagement an wissenschaftlichen Bibliotheken setzt sich aus mehreren Aspekten zusammen. Neben der reinen administrativen Arbeit, der Bereitstellung einer Möglichkeit, die Daten kurz, -mittel oder langfristig verfügbar zu machen, sind auch technische und juristische Expertisen notwendig. Der Beitrag schildert anhand des Beispiels der Universität Wien worauf beim Umgang mit Forschungsdaten geachtet werden sollte.

Abstract:

Research data management at scientific libraries consists of several aspects. In addition to pure administrative work, the provision of a way to make the data short, medium or long-term available, also technical and legal expertise is necessary. The article uses the example of the University of Vienna to describe what should be considered when dealing with research data.

Das Repositorium als Ausgangspunkt für Forschungsdatenmanagement

Gemeinsam mit dem Zentralen Informatikdienst der Universität Wien startete die Universitätsbibliothek Wien 2007 mit dem Aufbau eines universitätsweiten Repositories, das auf der Open Source Software Fedora [1] basiert. Unter Einbeziehung mehrerer Universitätsinstitute wurden Anforderungen aus unterschiedlichen Wissenschaftsbereichen erhoben und in die Planung und Ausarbeitung miteinbezogen. Der Fokus für das Repositorium war schon bereits bei der Konzeption nicht nur ausschließlich auf die Forschung gerichtet sondern auch auf die Lehre. Als Metadatenschemata wurden das LOM-Schema (Learning Object Metadata) und Dublin Core gewählt. Bei der Entwicklung der Software arbeiten die drei Abteilungen: Universitätsbibliothek, Zentraler Informatikdienst und Center for Teaching und Learning eng zusammen.

Im April 2008 ging PHAIDRA (ein Akronym für Permanent Hosting, Archiving and Indexing of Digital Resources and Assets) online. Von Beginn an stand das Repositorium allen MitarbeiterInnen und Studierenden der Universität Wien offen, zunächst mit begrenztem Speicherplatz, später ohne Limitierung. Zusätzlich können Zugänge für externe Forschende vergeben werden, was ein kollaboratives Arbeiten mit dem System möglich macht. Ergänzt wurde PHAIDRA schon bald durch u:scholar, den Institutionellen Repositorium der Universität Wien, das mit einer eigenen Upload - und Suchmöglichkeit für wissenschaftliche Publikationen ausgestattet sind. Archiviert werden auch diese Objekte in PHAiDRA. Auf die rechtliche Sicherheit wurde von Beginn an geachtet, so stehen unter anderem den Usern CC-Lizenzen zur Verfügung. Objekte ohne Kennzeichnung des Urheberrechtes sind nicht zulässig. Allen NutzerInnen von PHAIDRA wurden bei der ersten Benutzung die Nutzungsbedingungen zur Kenntnis gebracht, die vorsehen, dass nur Objekte archiviert werden dürfen, die aus urheberrechtlicher Sicht auch in das Repositorium abgelegt werden dürfen und gegen die auch sonst keine juristischen oder ethischen Gründe vorliegen.

In PHAIDRA können alle Arten von Objekten archiviert werden, unter anderem Texte, Bilder, Video- und Audiofiles sowie Daten. Derzeit sind neben Artikeln, Bilder, die aus der Forschung entstanden sind und/oder der Lehre dienen, Karten, Bücher, Ausdio- oder Videomittschnitte von Vorlesungen, Präsentationen, Grafiken, Statistiken, alle Arten von Sammlungsobjekten und vieles mehr archiviert. Die Owner, jene Personen, die Objekte hochladen, können die Beschreibungen, Metadaten, immer wieder ergänzen und verändern. Die Objekte selbst können, da dies der Policy der Langzeitverfügbarkeit widersprechen würde, nicht mehr gelöscht werden. Allerdings ist eine Versionierung möglich, die neue Version erhält einen neuen Permalink und wenn gewünscht auch eine eigene Beschreibung. Wichtig für die Forschenden ist auch, dass der Zugang zu den Objekten zunächst zwar - entsprechend der Open Access Policy der Universität Wien - weltweit geöffnet ist, durch den Owner jedoch sofort auf die gesamte Universität Wien, auf Institute, Departments, Personengruppen und Einzelpersonen gesperrt werden kann. Diese Sperre kann mit einem Embargo versehen werden und jederzeit verändert werden. Damit ist es zum Beispiel möglich innerhalb eines Forschungsprojekts zunächst unter Ausschluss der Öffentlichkeit zu arbeiten und nach Projektende Objekte automatisch freizugeben. Diese Funktion

macht es auch möglich, Artikel in PHAIDRA zu archivieren, die vom Verlag aus noch unter einer Embargozeit stehen, also erst nach einer gewissen Zeitspanne verbreitet werden dürfen. Zu den fast seit Beginn bestehenden Services zählt auch die Möglichkeit mittels eines PHAIDRA-Importers Collections und eigene Bücher zu erstellen und im Bookviewer anzeigen zu lassen. Die Funktion der Collections ermöglicht die gemeinsame Darstellung von unterschiedlichen Objekten. Collections erhalten einen eigenen Permalink und eine eigene Beschreibung. Jede Collections kann wiederum Collections enthalten, was beispielsweise die Darstellung von Zeitschriften nach Jahrgängen bis auf Artikelebene ermöglicht. Das Besondere daran ist, dass Collections auch von Objekten gebildet werden kann, von denen man nicht der Owner ist. Damit wird es beispielsweise möglich per Link einen ganzen Satz an Objekten für eine Lehrveranstaltung an Studierende zu schicken.
Genutzt wird PHAIDRA derzeit von nahezu allen Disziplinen, vor allem aber von den Geisteswissenschaften [2], was unter anderem zeigt, dass diese Disziplinen bisher keine geeigneten Tools hatten.

Neue Anforderungen der User

Im Laufe der Zeit veränderten sich die Anforderungen der NutzerInnen. Stand am Anfang vor allem die sichere Langzeitverfügbarkeit im Mittelpunkt, kamen neue Bedürfnisse hinzu, wie etwa die Frage nach kurz-, bzw. mittelfristigen Speichermöglichkeiten. Neben der unterschiedlichen Dauer der Archivierung wird vermehrt der Bedarf an Archivierung von Datenbanken und Software erkennbar.

Aufgrund der neuen Herausforderungen wurde die Architektur von einem Repositorium hin zu einem Ökosystem aus Repositorien und Daten-Services geändert. So wurde zum Beispiel die Landschaft durch ein Git basierendes Repositorium [3] ergänzt um Software Code und dynamische Daten abzulegen. An diesen Anforderungen lässt sich euch ein verändertes Forschungsverhalten ablesen.

Als Services für die Daten werden automatisch handle IDs [4] für jedes Objekt mitgeliefert, zusätzlich werden DOI's für die Objekte auf Nachfrage über ein eingerichtetes DOI-Service vor oder nach der Archivierung vergeben.

Neben den Services für die Zitierung von Daten, werden die Visualisierung der Forschungsdaten und die Darstellung der Ergebnisse auf Projekt bezogen Webseiten immer wichtiger. So können Daten von Phaidra über Schnittstellen an in andere Webauftritte integriert werden. Dabei orientiert sich die technischen Umsetzung an den Empfehlung von der Vereinigung Confederation of Open Access Repositories (COAR), die im Rahmen der Interest Group Next Generation of Repositories [5]. Über diese Möglichkeit haben wir bereits Forschungsprojekt in ihrer Darstellung in der Öffentlichkeit unterstützt.

Auch unsere Projektpartner nutzen diese Möglichkeit um Applikationen für den Ingest und den Re-use an PHAIDRA anzubinden. Grundgesetzlich ist unser Designparadigma als ein System in der Mitte, in dem auch Maschinen, nicht nur Menschen die Daten ablegen und wiederverwenden können. Es ist ein gut organisierte Lagerhalle von Daten, nicht deren Produktionsort und auch nicht der Ort an dem sie visualisiert werden.

Dennoch bieten wird im Rahmen unserer Services auch Darstellung für die spezielle Datentypen an. So können Videofiles von PHAIDRA aus gestreamt werden und für Bilder habe wir einen Image Server. Mithilfe dieses Servers können hochauflösende Bilder über das Web zoombar dargestellt werden und die Bilder können über den IIIF Standard (International Image Interoperability Framework) [6] angesprochen werden.

Parallel zu den Forderungen nach einer sicheren Archivierung der Daten wuchs die Nachfrage von Beratung vor allem im technischen und im juristischen Bereich. Das hängt sicher auch damit zusammen, dass einige Fördergeber bereits Datenmanagementpläne verlangen und einen bewussten und nachhaltigen Umgang mit Forschungsdaten einfordern. Der FWF (Fonds zur Förderung der wissenschaftlichen Forschung) verlangt zum Beispiel bei aktuellen Einreichungen von Forschungsprojekten in seiner DMP-Vorlage Angaben über Datennutzungs- und Speicherungsstrategien sowie Hinweise auf den juristischen und ethischen Umgang mit Daten.

Datenmanagement anhand von Datenmanagementplänen

Verpflichtende Datenmanagementpläne haben dazu geführt, dass Forschende zunehmend Unterstützung bei der Antragstellung benötigen. Diese Unterstützung kann nur bis zu einem gewissen Grad durch Schulungen und Präsentationen abgedeckt werden, denn Forschungsdaten sind sehr unterschiedlich, betreffend die Struktur, die Menge aber auch was juristische und ethische Fragen betrifft. Dies wiederum verlangt den Einsatz unterschiedlicher Tools und Strategien. Während auf der einen Seite Open Science immer stärker eingefordert wird, vor allem von den Fördergebern, zeigt der langjährige Kontakt mit Forscherinnen und Forschern aus verschiedenen Fachbereichen, dass in einigen Bereichen sehr sensibel mit Daten umgegangen werden muss. Dazu zählen nicht nur Ergebnisse aus medizinischen Projekten sondern beispielsweise auch aus sozialwissenschaftlichen und geisteswissenschaftlichen Zusammenhängen. So kann beispielsweise ein vor 20 Jahren geführtes Interview im Rahmen der Kultur- und Sozialanthropologie mit einer kleinen Personengruppe, diese heute aufgrund veränderter politischer Gegebenheiten gefährden., Aber auch unklare rechtliche Besitzverhältnisse von Daten beeinträchtigen die Vision einer offenen Wissenschaft. Zu diesen extern bedingten Einschränkungen kommen auch Unsicherheiten und Widerstände bei den Forscherinnen und Forschern hinzu, die sich vor einer allzu freien Preisgabe ihrer Daten scheuen. Die Gründe dafür sind vielfältig. Bei hochkompetitiven Fachbereichen ist die Angst vor Ideenklau vorherrschend, in anderen Bereichen überwiegt wiederum die Sorge, dass Daten missbräuchlich verwendet werden könnten. Man denke dabei nur an Forschungen über den Nationalsozialismus.

Neue Herausforderungen

Die Herausforderungen für TechnikerInnen und BibliothekarInnen besteht nun darin, die Forschenden in einer Art und Weise zu unterstützen, dass ihnen hinsichtlich Verfügbarmachung ihrer Daten möglichst viele Optionen offenstehen. Das bedeutet, sie benötigen auf der technischen Seite einerseits offene Repositorien, in die sie Daten hochladen und weltweit mit einer DOI versehen zugänglich machen können, andererseits aber auch geschlossene Systeme, die es ihnen erlauben, ihre Ergebnisse sicher zu verwahren und vor unberechtigten Zugriffen zu schützen. Das betrifft zum Teil nicht nur die Daten sondern auch die Metadaten. Für manche Daten ist auch die Löschbarkeit unbedingt erforderlich, vor allem im medizinischen Bereich. Auf der

nichttechnischen Seite wiederum sind Strukturen notwendig, die entlang des gesamten Forschungsprozesses Beratungen zu juristischen oder ethischen Fragestellungen anbieten, bzw. zum Beispiel hinsichtlich Metadatenvergabe unterstützend wirken.

Eine weitere Herausforderung ist, wie oben bereits erwähnt wurde, dass Forscherinnen und Forscher nicht nur Daten in Form von Dateien archivieren wollen, sondern auch Datenbanken und Software. Bei diesen beiden Datenklassen versagt die herkömmliche Archivierung in einem Repositorium, da der Kontext und die Umgebung für die Lauffähigkeit ein wesentlicher Bestandteil der Archivierung sein muss. Ohne diese beiden Teile kann der Zustand in der Zeit zwar eingefroren werden, es kann aber nicht mehr von einer langfristigen Zurverfügungstellung von Daten gesprochen werden.

Datenmanagement als Teamwork

An der Universität Wien blicken wir nun auf eine mehr als zehnjährige Erfahrung im Bereich Forschungsdatenmanagement zurück. Der Zentrale Informatikdienst der Universität Wien arbeitet dabei sehr eng mit der Universitätsbibliothek zusammen, betreibt nicht nur gemeinsam das universitätsweites Repositorium PHAIDRA, sondern führt vor allem Beratungsgespräche immer im Team durch um auf die unterschiedlichen Anforderungen rasch eingehen zu können. Durch die 19 Partneruniversitäten von PHAIDRA hat sich ein großes Netzwerk gebildet, das sich in technischen und auch in nichttechnischen Belangen gegenseitig stützt. Das über 90 Personen umfassende RepositorienmanagerInnennetzwerk (RepManNet) tauscht zusätzlich systemübergreifend Erfahrungen aus arbeitet in diversen Arbeitsgruppen an der Konzeption von zukünftigen Repositorien und an Guidelines. Den Forschenden an der Universität Wien steht zusätzlich das Netzwerk “code4research” zur Verfügung, das dem Erfahrungsaustausch der WissenschaftlerInnen bzgl. der langfristigen Verfügbarkeit von komplexeren Daten wie Datenbanken und Software dient.

Zusammenfassend lässt sich sagen, dass Forschungsdatenmanagement nur im Zusammenspiel durch mehrere Personen mit unterschiedlichen Expertisen funktioniert, dass eine enge und von Vertrauen geprägte Zusammenarbeit mit den Forschenden erfolgen muss und dass die beteiligten Personen sich ständig auf neue

Herausforderungen einlassen müssen. Dabei muss auch mit anderen Stellen der Universität, zum Beispiel mit dem Forschungsservice, aber auch mit den Fördergebern gut zusammengearbeitet werden. Diese Zusammenarbeit führt nicht nur zu einem Ausbau der Kompetenzen auf allen Seiten, sondern auch zu konkreten Verbesserungen der Systeme. Für PHAIDRA hat dies beispielsweise im letzten Jahr bedeutet einen neuen Objekttyp, einen Container mit komplexen Features, zu entwickeln.

Fußnoten

[1] Die Software hieß zu den damaligen Zeitpunkt noch Fedora Commons. Die Umbenennung erfolgte im Zuge der Zusammenführung der Entwicklungen der beiden großen Open Source Repositorien Fedora und DSpace. Für beide ist nun die Organisation DuraSpace verantwortlich. Weiterführende Informationen sind online zu finden unter: https://duraspace.org/

[2] Bei der letzten statistischen Erhebung von den Objekte in PHAIDRA im Jänner 2019 machen Objekte die der Digital Humanities zugeordnet werden können 65% aus.

[3] Zur Zeit verwendet die Universität Wien GitHub Enterprise Git Repositorium.

[4] Handle ist die technische Basis von DOI und ist Open Source. Weiter Information sind online zu finden unter: https://www.handle.net

[5] „Our vision is to position repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication, on top of which layers of value added services will be deployed, thereby transforming the system, making it more research-centric, open to and supportive of innovation, while also collectively managed by the scholarly community. However, in order to leverage the value of the repository network, we need to equip it with a wider array of roles and functionalities, which can be enabled through new levels of web-centric interoperability.“ ist das Ziel der Arbeitsgruppen Next Generation Repository von COAR, die auf internationalen Experten der Bibliothek und der IT besteht. Online zu finden unter: https://www.coar-repositories.org/activities/advocacy-leadership/working-group-next-generation-repositories/

[6] Weitere Informationen zu diesem Framework könne online unter https://iiif.io/ abgerufen werden.

Literatur

Blumesberger, Susanne: Neue Anforderungen – viele offene Fragen. Zu den vielfältigen Rollen von Repositorien am Beispiel der UB Wien. In: Mitteilungen der VÖB

Eberhard, Igor; Kraus, Wolfgang: Der Elefant im Raum. Ethnographisches Forschungsdatenmanagement als Herausforderung für Repositorien. In: Mitteilungen der VÖB

Ganguly, Raman: Nachhaltige Softwareentwicklung: code4research. In: Mitteilungen der VÖB

Ganguly, Raman, Paolo Budroni, and Barbara Sánchez Solís. "Living Digital Ecosystems for Data Preservation: An Austrian Use Case Towards the European Open Science Cloud." In Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices, 203-210. ELPUB. Limassol, Cyprus, 2017; https://elpub.architexturez.net/doc/10-3233/978-1-61499-769-6-203

Hagmann, Dominik: Überlegungen zur Nutzung von PHAIDRA als Repositorium für digitale archäologische Daten. In: Mitteilungen der VÖB

Kulturpool: Österreichs Portal zu Kunst, Kultur und Bildung (30. April 2018): Phaidra, die Strahlende. Interview mit Susanne Blumesberger.

Torggler, Andrea; Andrae, Magdalena: Aus dem Leben einer/s Repman – ein Bericht aus dem österreichischen „Netzwerk für RepositorienmanagerInnen“. In: Mitteilungen der VÖB

Forschungsdatenmanagement in einer wissenschaftlichen Spezialbibliothek – Chancen und Herausforderungen in einem interdisziplinären Forschungsinstitut

Harald Kaluza

Deutsches Institut für Erwachsenenbildung – Leibniz-Zentrum für Lebenslanges Lernen e.V. (DIE)

Zusammenfassung

Der Erfahrungsbericht schildert die Herausforderungen und Potenziale beim Aufbau und Betrieb eines institutionellen Forschungsdatenmanagements am Deutschen Institut für Erwachsenenbildung (DIE), welches interdisziplinäre Forschungsleistungen in den Bereichen des Lernens und Lehrens Erwachsener erbringt. Mit sowohl qualitativen Daten (z.B. Interviewtranskripte) als auch quantitativen Daten (z.B. Leistungsmessungen) wird eine sehr heterogene Datenmenge erzeugt. Das Dienstleistungsangebot der wissenschaftlichen Spezialbibliothek am DIE wurde im Juni 2017 um ein institutionalisiertes Forschungsdatenmanagement sowie eine Forschungsdaten-Policy erweitert. Der Service umfasst die Begleitung eines jeden Forschungsprojekts über seinen gesamten Lebenszyklus beginnend von der Forschungsidee bis zur Dissemination der Projektergebnisse. Vorausgegangen ist ein umfassender Lern- und Implementierungsprozess bei allen am Forschungsdatenmanagement beteiligten Akteuren. Neben der Schaffung einer institutionellen Stelle des Forschungsdatenmanagers wurde frühzeitig die Zielgruppe der Wissenschaftler sensibilisiert, sich als Datengeber aktiv an diesem Prozess zu beteiligen. Dabei spielt die persönliche Interaktion der Forschungsdatenmanager mit Wissenschaftlern, die sich neuen Fachkulturen und Forschungsparadigmen stellen müssen, eine entscheidende Rolle. Das DIE partizipiert aktiv am Verbund Forschungsdaten Bildung und nutzt dabei entsprechende Vernetzungspotenziale in Bezug auf die Standardisierung von Prozessen sowie die Beratung und Schulung von Akteuren. Zunehmend verlangen potenzielle Drittmittelgeber die Beschreibung eines umfassenden Forschungsdatenmanagements (inklusive der Möglichkeit der Datennachnutzung) in Forschungsanträgen. Dies macht aus Sicht des Forschungsdatenmanagements eine Standardisierung der Prozesse notwendig. So bietet das DIE Formulierungen für Drittmittelanträge an, die in individuellen Beratungen jeweils angepasst werden. Daraus resultieren im Idealfall standardisierte Datenmanagementpläne. Management von Forschungsdaten kann schlussendlich die

Stellung der Bibliothek innerhalb eines Instituts stärken und diesem innovative Wege aufzeigen.

1. Einführung und Ausgangslage

Das Deutsche Institut für Erwachsenenbildung – Leibniz-Zentrum für Lebenslanges Lernen (DIE) mit Sitz in Bonn[1] ist die zentrale Einrichtung für Weiterbildungsforschung in Deutschland. Im Mittelpunkt der Forschung stehen Lernprozesse von Erwachsenen, Weiterbildungsangebote und deren Programme, Lehrkräfte in der Weiterbildung, die Organisation von Weiterbildungseinrichtungen sowie das Weiterbildungssystem in Deutschland und Europa. Adressaten der Forschung sind neben der Wissenschaft auch Lehrkräfte und Beratende in der Weiterbildung, Leitungspersonal und Programmplanende in den Weiterbildungseinrichtungen und -organisationen sowie Akteure aus Politik und den Verbänden. Organisatorisch teilt sich das DIE hierfür in die beiden Bereiche „Forschung“ und „Infrastruktur“ auf. In der Abteilung „Forschungsinfrastrukturen“ des letzteren Bereichs ist auch die wissenschaftliche Spezialbibliothek des DIE angesiedelt. Neben Literaturdokumentation, Medien- und Informationsversorgung sowie der Bereitstellung forschungsunterstützender Dienstleistungen und den Archiven zur Geschichte der Erwachsenenbildung ist in der Bibliothek auch das institutionelle Forschungsdatenmanagement angesiedelt. Dieser neuen Aufgabe liegt eine bewusste strategische Entscheidung zur Profilschärfung der Bibliothek als Teil des Forschungsinstituts zu Grunde, die im Zuge dieser Ausführungen noch näher beschrieben wird. Im Forschungsbereich des Instituts werden die genannten Themen der Weiterbildungsforschung in vier Abteilungen von Wissenschaftlern unterschiedlichster Fachrichtungen bearbeitet. So finden sich neben einer Vielzahl an (Erwachsenen-)Pädagogen vor Allem auch Vertreter verwandter Disziplinen wie der Soziologie und der Psychologie. Hinzu kommen einzelne Forscher aus ökonomischen oder geisteswissenschaftlichen Disziplinen.

Dieser interdisziplinäre Forschungsansatz spiegelt sich in den vielfältigen Methoden und Erhebungsverfahren der durchgeführten Projekte wider. Neben experimentellen Verfahren und psychologischen Tests finden sich verschiedenste Beobachtungsverfahren (teilnehmend, nicht teilnehmend, videogestützt) aber auch Längsschnittuntersuchungen wie die Volkshochschulstatistik und unterschiedlichste Formen standardisierter Befragungen. Hinzu kommen Akten- und Dokumentanalysen

[1] Vgl.: https://www.die-bonn.de/institut/default.aspx

und die Auswertungen prozessgenerierter Daten, beispielsweise in der Analyse von Volkshochschulprogrammen. Erzeugt wird eine Vielzahl sowohl qualitativer als auch quantitativer Daten in unterschiedlichster typologischer Ausprägung, z.B. reiner Text, Audio- und Videodaten oder statistische Daten. Heterogen sind dementsprechend auch die verarbeiteten Datenformate, die neben gängigen offenen und proprietären Formaten auch spezielle Typen, beispielsweise aus Textcodierungsprogrammen, beinhalten. Seit dem Jahr 2000 können in der Forschungslandkarte des DIE[2], die einen Überblick über aktuelle und abgeschlossene Forschungsprojekte zur Erwachsenenbildung/Weiterbildung an deutschen Hochschulen und am DIE ermöglicht, etwa 150 Projekte mit direkter institutioneller Beteiligung des DIE nachgewiesen werden. Ein entsprechender Handlungsbedarf, anfallende (Forschungs-)Daten professionell hinsichtlich einer Langzeitarchivierung, Replizierbarkeit und Nachnutzbarkeit zu behandeln, ist somit hinreichend gegeben. Als Leibniz-Institut wird das DIE regelmäßig (spätestens alle sieben Jahre) vom Senat der Leibniz Gemeinschaft evaluiert, indem die Arbeiten in Wissenschaft, Forschung, Beratung, Dienstleistungen und anderen Aufgabenfeldern bewertet werden[3]. Der wichtigste Impuls für das Institut, sich dem Themenfeld des Forschungsdatenmanagements anzunehmen erfolgte aus der Beurteilung des Instituts im Jahre 2012, in dem ein institutionelles Forschungsdatenmanagement empfohlen wurde. Diese Voraussetzungen führen zu den grundsätzlichen Fragen und Herausforderungen, mit denen sich das DIE konfrontiert sah: Wie gelingt es, ein institutionelles Forschungsdatenmanagement an einem vergleichbar kleinen, interdisziplinär arbeitenden Forschungsinstitut zu implementieren, etablieren und weiterzuentwickeln? Welche strategischen Möglichkeiten ergeben sich durch die Einbindung des Forschungsdatenmanagements in die Organisationsstruktur der Bibliothek, um deren Position und Stellenwert im Institut zu stärken? Zudem steht das Forschungsdatenmanagement in einem Spannungsfeld aus drei Einflussfaktoren: Zum ersten den Wissenschaftlern, die sich kritisch neuen Fachkulturen und Forschungsparadigmen stellen müssen. Zum zweiten den Erwartungen und Vorgaben der Institutsleitung bzw. des Trägers sowie zum dritten den dynamischen Entwicklungen im Bereich der Wissenschaftspolitik und Forschungsförderung.

[2] Vgl. https://www.die-bonn.de/weiterbildung/forschungslandkarte/default.aspx
[3] Vgl. https://www.leibniz-gemeinschaft.de/ueber-uns/evaluierung/

2. Implementierungsphase des Forschungsdatenmanagements

Dem Impuls aus der Evaluation von 2012 folgte im Jahr 2014 die Einstellung eines „Referenten für Forschungsinfrastruktur und Fachinformation“, der neben anderen Aufgaben als Forschungsdatenmanager für die Implementierung des institutionellen Forschungsdatenmanagements am DIE sorgen sollte. Die initiale Arbeit bestand dabei im Wesentlichen aus mehreren parallelen Arbeitsschwerpunkten: Dem thematischen Kompetenzaufbau, der Evaluation der institutionellen Strukturen, Daten und Arbeitsweisen sowie der Vernetzung mit anderen, disziplinarisch verwandten Einrichtungen. Basierend auf den gewonnenen Erkenntnissen wurden eine generelle Strategie und einzelne Prozesse für das Forschungsdatenmanagement entwickelt und implementiert.

2.1 Kompetenzaufbau und Strategieentwicklung

Das Thema Forschungsdatenmanagement war im Jahr 2014 im informationsinfrastrukturellen Diskurs längst verankert, hatte aber noch nicht die Dynamik entwickelt, die es in den folgenden Jahren bekommen sollte. Als wichtiger Impulsgeber ist für die Leibniz-Gemeinschaft der bereits seit 2009 bestehende Arbeitskreis Forschungsdaten[4] zu nennen. Aus den Nestor Aktivitäten zur digitalen Langzeitarchivierung entstand die DINI-Nestor AG Forschungsdaten (2014)[5]. Disziplinübergreifende Strukturen für deutschsprachige Forschungsinfrastrukturen wurden im BMBF-geförderten WissGrid-Projekt (2009-2012)[6] erörtert, die Langzeitarchivierung von Forschungsdaten war hier ein elementarer Bestandteil. Der hieraus entstandene „Leitfaden zum Forschungsdatenmanagement“[7] gilt als eine der ersten umfassenden deutschsprachigen Einführungen und Planungsinstrumente zum Thema. Aus supranational forschungspolitischer Sicht muss Horizon 2020[8] als großes europäisches Förderprogramm genannt werden. Es rückte das Forschungsdatenmanagement auf breiter internationaler Basis in den Fokus, und manifestierte den Diskurs somit auch auf politischer Ebene. In den folgenden Jahren entwickelten alle relevanten Stakeholder aus Politik, Wissenschaftsförderung und wissenschaftlicher Praxis eine Vielzahl an Initiativen, (disziplinären) Empfehlungen und Planungsinstrumenten, deren einzelne Nennungen an dieser Stelle über den

[4] Vgl. https://www.leibniz-gemeinschaft.de/ueber-uns/organisation/arbeitskreise/arbeitskreis-forschungsdaten/
[5] Vgl. https://dini.de/ag/dininestor-ag-forschungsdaten/
[6] Vgl. https://www.sub.uni-goettingen.de/projekte-forschung/projektdetails/projekt/wissgrid/
[7] Vgl. Ludwig 2013
[8] Vgl. https://ec.europa.eu/programmes/horizon2020/what-horizon-2020

Rahmen des Textes hinausgehen würden. Festzuhalten bleibt, dass das Forschungsdatenmanagement auf theoretischer Ebene bereits im Jahr 2014 in den wichtigsten Grundzügen anhand des „Datenlebenszyklus“[9] normiert worden war und die zentralen Aufgaben über einen Datenmanagementplan (DMP) standardisiert wurden. Es fehlte jedoch vor Allem an Erfahrungsberichten aus der Praxis, da das Thema zu diesem Zeitpunkunkt in der Breite noch nicht angewandt wurde.

Als hilfreich für das DIE-Forschungsdatenmanagement erwiesen sich an dieser Stelle Hospitationen und Beratungen bei disziplinverwandten Einrichtungen, die bereits ein Forschungsdatenzentrum (FDZ) oder Datenarchiv betrieben. Zu nennen sind hier u.a. das FDZ des Bundesinstituts für Berufsbildung (BIBB)[10] und das Datenarchiv des Leibniz-Instituts für Sozialwissenschaften (GESIS)[11]. Neben fachlichen Impulsen konnten hier vor allem auch Vernetzungsoptionen und -strategien erörtert werden. Eine frühe Erkenntnis war dabei, dass es wenig zielführend und gemessen an den vorhandenen Ressourcen auch nicht möglich war, ein institutionelles Forschungsdatenmanagement aufzubauen, welches sämtliche anfallenden Datentypen komplett von der Erhebung bis zur etwaigen Archivierung und Präsentation in einem eigenen institutionellen Angebot mit einbezieht. Es ist sinnvoll, bereits bestehende Infrastrukturen für die eigene Arbeit zu nutzen, bzw. diese komplementär an ihnen auszurichten, damit parallele Strukturen beispielsweise zu o.g. vermieden werden.

2.2 Aufbau interner Strukturen

Neben der Kompetenzentwicklung musste das Thema Forschungsdatenmanagement am Institut etabliert werden, um den Aufbau strukturierter Prozesse zu fördern. Unbestritten ist, dass Forschungsdatenmanagement für die beteiligten Wissenschaftler als Produzenten der Forschungsdaten, bezogen auf etablierte Arbeitsweisen und bestehende Methoden einen Kulturwandel bedeuten muss. Die Idee des Forschungsdatenmanagements, Forschungsabläufe planmäßig auf die etwaige spätere Veröffentlichung der Daten abzustimmen war und ist für einzelne Wissenschaftler neu und ungewohnt, und wird in der Regel mit einem unnötigen Mehraufwand verbunden.[12] Um aber ein funktionierendes

[9] Vgl. z.B. Ludwig 2013, S.14 f.
[10] Vgl. https://www.bibb.de/de/53.php
[11] Vgl. https://www.gesis.org/institut/abteilungen/datenarchiv-fuer-sozialwissenschaften/
[12] Vgl. Borgmann 2012

Forschungsdatenmanagement an einem Institut einzuführen und zu etablieren, sollten die Wissenschaftler in den Implementierungsprozess mit einbezogen werden. Der „Arbeitskreis Forschungsdaten“ des DIE setzte sich neben dem Forschungsdatenmanager und einem Bibliothekar aus je einem Wissenschaftler der vier wissenschaftlichen Abteilungen zusammen. Zum einen sollte hier das theoretische Wissen zum Forschungsdatenmanagement über Multiplikatoren in das Institut getragen werden. Zum anderen sollten Anregungen der späteren Datengeber bei den Planungen für das Forschungsdatenmanagement mit einfließen. Zu einzelnen Themen (Recht, IT) wurden bei Bedarf entsprechende Kollegen, z.B. der Datenschutzbeauftragte des DIE, hinzugezogen. Gleichzeitig konnten von Seiten des Forschungsdatenmanagers wichtige Einblicke in die verschiedenen Methoden und Datentypen im Institut gewonnen werden. Erfolgreich geplant und durchgeführt wurde durch den Arbeitskreis ein einführender, institutsweiter Workshop zum Forschungsdatenmanagement, der durch einen externen Experten aufgewertet wurde. Beides führte dazu, dass das Thema am Institut weiter verankern konnte. Eine weitere entscheidende positive Maßnahme war die Erarbeitung der institutionellen Forschungsdaten-Policy durch den Arbeitskreis. Als Herausforderung stellten sich jedoch immer Aspekte dar, die sich auf eine konkrete praktische Erprobung einzelner Bereiche des Forschungsdatenmanagements bezogen. Das Ziel, an einem Projekt prototypisch die Umsetzung eines DMP zu erproben, um wichtige Erkenntnisse für den Transfer aus der Theorie in die Praxis zu gewinnen, wurde nicht erreicht. An diesem Punkt wurde in der Regel auf mangelnde Ressourcen verwiesen.

Es zeigte sich immer wieder, dass trotz stetiger interner Lobbyarbeit eine Implementierung der Prozesse ohne verpflichtende Richtlinien „bottom up“ nicht zu bewerkstelligen war. Hier bestätigte sich, dass der genannte Kulturwandel zwar eng mit den Wissenschaftlern kommuniziert werden muss, die Rahmenbedingungen für eine konkrete Umsetzung z.B. in Form einer Forschungsdaten-Policy aber „top down“ seitens der Institutsleitung über die einzelnen Abteilungsleiter mitgetragen und durchgesetzt werden müssen. Im Juni 2017 wurde am DIE die „Forschungsdaten-Policy für Forschende des DIE“[13] verabschiedet und durch den wissenschaftlichen Direktor gesondert vorgestellt. Diese beinhaltet im Wesentlichen die folgenden Aspekte: Die Minimalanforderungen an ein Forschungsprojekt des DIE bestehen darin,

[13] Vgl. https://www.die-bonn.de/forschungsdaten-policy/default.aspx

die Daten nach den DFG-Regeln guter wissenschaftlicher Praxis[14] mindestens für 10 Jahre zu archivieren. Eine Veröffentlichung der Daten soll nach Möglichkeit angestrebt werden, basiert aber vor allem auf deren Rechtssicherheit. Die grundsätzliche Entscheidung zur Veröffentlichung der Daten liegt immer bei den einzelnen Projektleitungen bzw. Datengebern und dem Institutsvorstand. Eine negative Entscheidung zur Veröffentlichung ist zu begründen. An dieser Stelle trägt die Policy dem Umstand Rechnung, dass eine pauschale Verordnung zur Veröffentlichung der Daten weder praktikabel noch realistisch ist und sich mithin nicht an einer klassischen Version von „Open-Data“, sondern an den „FAIR-Data“-Prinzipien orientiert.[15] Insbesondere die Gewährleistung der Rechtssicherheit von Daten ist in der empirischen Bildungsforschung nicht immer möglich. Mitunter ist damit ein enormer Arbeitsaufwand verbunden, der nicht in allen Projekten im Verhältnis zum Gesamtaufwand steht, so dass eine individuelle Abwägung zur Veröffentlichung der Daten pro Projekt zielführender erscheint. Mit der Verabschiedung der Policy startete der Regelbetrieb des Forschungsdatenmanagements am DIE.

3. Regelbetrieb

3.1 Basis

Der Regelbetrieb des DIE-Forschungsdatenmanagements ruht auf zwei Säulen: Dem dauerhaften Beratungsangebot durch den institutionellen Forschungsdatenmanager sowie dem „Datenmanagement-Handbuch für Forschende des DIE“. Das Handbuch dient als Leitfaden für die Wissenschaftler, der neben den theoretischen Hintergründen und Empfehlungen zum Forschungsdatenmanagement sämtliche Informationen zur praktischen Umsetzung aggregiert. Hauptbestandteil ist der DMP, welcher idealerweise während der kompletten Laufzeit des Projektes gepflegt werden soll. Zu allen Punkten der Checkliste des DMP finden sich detaillierte Erklärungen, Beispiele, Literaturhinweise und Links zu weiteren Ressourcen, sowie hausinterne Ansprechpartner für bestimmte Themenbereiche. Insbesondere für die Beantragung von Drittmitteln und Fragen zum Datenschutz finden sich Vorlagen oder Textbausteine, etwa für informierte Einwilligungserklärungen. Der Forschungsdatenmanager sollte als Ansprechpartner in allen Phasen des Projektes zur Beratung hinzugezogen werden. Ziel ist es, den Forschungsprozess bzw. ein Forschungsprojekt möglichst von der initialen Projektidee bzw. der Planungsphase bis

[14] Vgl. Deutsche Forschungsgemeinschaft 2013
[15] Vgl. Wilkinson 2016

zu dessen Ende dauerhaft und regelmäßig zu begleiten, um so eine professionell kuratierte Datenbasis zu erreichen.[16]

3.2 Erfolge und Herausforderungen

Nach etwa zwei Jahren Regelbetrieb können erste Erfahrungen ausgewertet werden. Positiv anzumerken ist, dass erste Projekte, die seit 2017 gestartet sind, Prinzipien des Forschungsdatenmanagements umsetzen. Zwei Projekte haben dezidiert und erfolgreich Ressourcen für Forschungsdaten in Drittmittelanträgen akquirieren können. Auch die Anzahl der laufenden Beratungen und Anfragen zum Forschungsdatenmanagement steigt stetig. Ein weiterer wichtiger Aspekt ist die stetige Rückmeldung zum Datenmanagement-Handbuch, die den Transfer der theoretischen Empfehlungen in die Praxis evaluieren und zurückspiegeln. Wie erwähnt konnten vor der Implementierung keine praktischen Erfahrungen in die Konzeption der Empfehlungen mit einfließen. Die Rückmeldungen und Erfahrungen der Wissenschaftler sind somit ein elementarer Bestandteil der Arbeit und führen zu einer laufenden Anpassung der Empfehlungen. So hat sich beispielsweise schnell herausgestellt, dass nicht alle Projekte vertiefende Informationen zu allen Aspekten des Datenmanagements benötigen, diese etwa im Falle von zu nutzenden Metadatenstandards gar abschreckend wirken können. An solchen Stellen muss genau zwischen den allgemeinen relevanten Informationen und speziellen, vertiefenden Informationen unterschieden werden, die situationsbezogen angewendet werden können. Gerade bei diesen Fällen muss immer wieder betont werden, dass das Forschungsdatenmanagement viele dieser Aspekte im Verbund mit den Wissenschaftler stark unterstützen oder gar übernehmen kann. Die Praxis zeigt ebenfalls, dass zwischen Projektidee und Drittmittelbeantragung nicht immer ein entsprechendes Zeitfenster für ein bedachtes und durchdachtes Datenmanagement liegen kann, da beispielsweise oft kurzfristig auf noch kurzfristiger erschienene Ausschreibungen reagiert werden muss. Hier gilt es, trotzdem schnell und pointiert Hilfestellungen zur Ressourcenkalkulation über eine Analyse der entstehenden Daten zu liefern. Erste Datenpakete, die archiviert und zur Nachnutzung bereitgestellt werden können sind mittelfristig nach Abschluss der Projekte am DIE zu erwarten.

Diesen positiven Aspekten muss jedoch entgegengesetzt werden, dass die Akzeptanz des Forschungsdatenmanagements weiterhin auszuweiten ist. Längst nicht alle

[16] Vgl. Grafik Ablauf FDM am DIE: https://www.die-bonn.de/img/FDM-Ablauf_v02-01_web.png

Projekte wenden sich proaktiv in der Planungsphase an das Forschungsdatenmanagement, oft wird ein DMP als Mehrarbeit ohne Nutzen für das eigene Projekt empfunden. Stetige Informationsarbeit seitens des Forschungsdatenmanagers ist weiterhin nötig. So werden alle neuen Kollegen mit den Maßnahmen des Forschungsdatenmanagements vertraut gemacht, ferner Abteilungen und deren Leitungen regelmäßig mit Informationen versorgt. Ebenfalls nimmt der Forschungsdatenmanager bei Bedarf an Abteilungssitzungen teil. Obligatorisch soll die Anbindung an die hauseigene Drittmittelstelle werden, bei der eine Beratung zum Forschungsdatenmanagement zur internen Bewilligung vorzuweisen ist. Auch wird die Arbeit der 2019 am DIE neu gegründeten lokalen Ethikkommission mit der Beratung zum Forschungsdatenmanagement verknüpft. Der wichtigste Aspekt bleibt jedoch, dass Anforderungen und Bedingungen, die im DMP oder in der Drittmittelakquise zum Ausdruck gebracht werden, stetig umgesetzt und somit dauerhaft vom Forschungsdatenmanager begleitet werden müssen. Datenmanagementpläne müssen mit Leben gefüllt werden.

3.3 Anreize und Türöffner

Zu den ersten Erfolgen bei der Umsetzung des Forschungsdatenmanagements haben letztlich zwei „harte“ Faktoren beigetragen, die das generell vorhandene Verständnis für Open Science und FAIR-Data als Anreiz zur aktiven Durchführung deutlich überlagert haben. Die erfolgreiche Akquise von Drittmitteln ist hierbei primär zu nennen. Nachdem ein erstes Projekt erfolgreich Mittel mit Hilfe des Forschungsdatenmanagements einwerben konnte, wurde eine deutlich höhere Bereitschaft anderer Projekte registriert, das Forschungsdatenmanagement in die Planungen zu integrieren. Ein weiterer Faktor war die Einführung der Europäischen Datenschutz-Grundverordnung (DSGVO) im Mai 2018, die in Bezug auf den Umgang mit personengebundenen Daten, die in der empirischen Bildungsforschung eine tragende Rolle spielen, eine neue, vertiefte Beschäftigung mit dem Thema erforderte. Das DIE-Forschungsdatenmanagement konnte hier frühzeitig in Verbindung mit dem institutionellen Datenschutzbeauftragten für das Thema sensibilisieren, Beratungen und Lösungen anbieten, die sehr gut in Anspruch genommen wurden. Mit der lokalen Ethikkommission bietet sich zudem eine Möglichkeit, eng zu kooperieren. Ein weiterer Vorteil ist die Tatsache, dass es sich beim DIE um ein vergleichsweise kleines Institut handelt. Um die stetige Benutzung und Implementierung eines DMP zu gewährleisten, ist ein enger persönlicher Kontakt zu den Datengebern von immenser Bedeutung.

Verglichen etwa mit den Strukturen einer Universität mit verschiedenen Fakultäten und Instituten kann dies deutlich einfacher gehandhabt und beispielsweise auch über kurze, persönliche informelle Kontaktaufnahme durchgeführt werden.[17] Dieser eher niederschwellige Zugang kann zudem einem erzwungenen Charakter des Forschungsdatenmanagements entgegenwirken, der dem angesprochenen Kulturwandel nicht zu Gute käme. Gerade hier muss ein Gleichgewicht zwischen den Interessen der Wissenschaftler und den Anforderungen einer Forschungsinfrastruktur gefunden werden.

3.4 Optimierung der Prozessorganisation

Das Forschungsdatenmanagement des DIE befindet sich in einem kontinuierlichen Entwicklungsprozess. Das Hauptaugenmerk bei der Entwicklung der einzelnen Prozesse am DIE wurde im bisherigen Verlauf insbesondere auf die Begleitung der ersten Phasen eines Forschungsprojekts gelegt. Die Weiterentwicklung findet zeitgleich mit dem Fortschreiten der einzelnen laufenden Projekte statt, indem Rückmeldungen verarbeitet und aktuelle Themen erprobt werden. So wurden Fragen zum Metadatenmanagement, der dauerhaften Archivierung und der Publikation der Daten zwar schon bedacht, werden aber zum entsprechenden Zeitpunkt intensiver behandelt werden. Grundsätzlich strebt das DIE eine stetige Optimierung der vorhandenen Instrumente und Prozesse an. So ist beispielsweise an standardisierte und automatisierte Tools zum Erstellen und Pflegen von DMP[18] zu denken, oder an eine andere Zugangsart zum Datenmanagement-Handbuch, etwa in Form eines Wikis. Auch gemessen an den vorhandenen Ressourcen handelt es sich zudem um eine realistische Option, die dauerhafte Archivierung der Daten und deren Publikation bzw. deren Nutzungsmanagement an größere, zentrale (disziplinäre) Datenarchive abzugeben. Das DIE könnte die Metadaten auf diese Weise immer noch in einem angedachten eigenen Datenportal präsentieren und auf den Datenzugang verweisen. Weiterhin sollen alle Prozesse anhand aktueller oder entstehender Referenzmodelle für das Forschungsdatenmanagement analysiert werden, sei es, um weitere Kompetenzbedarfe[19] zu ermitteln oder um die Gesamtstrategie anzupassen oder neu auszurichten[20]. Dies dient der internen Qualitätssicherung.

[17] Vgl. Cremer et al. 2015
[18] Vgl. z.B. https://rdmorganiser.github.io/
[19] Vgl. Projekt PODMAN: https://fdm.uni-trier.de/projektbeschreibung/
[20] Vgl. Teilprojekt RISE-DE im Projekt FDMentor: https://www.forschungsdaten.org/index.php/FDMentor

4. Strategische Optionen

4.1 Externe strategische Optionen

Eine institutionelle und persönliche Vernetzung mit anderen Akteuren des Forschungsdatenmanagements ist unabdingbar. Auf der Ebene der Daten ist es überaus sinnvoll, bestehende Strukturen, seien es vorhandene Archive und Repositorien oder bereits entwickelte Instrumente und Standards, zu nutzen. Netzwerke dienen dem Kompetenzerwerb und bieten zudem die Möglichkeit, eigene Initiativen in einen breiteren Diskurs mit einzubringen. Seit 2016 ist das DIE Netzwerkpartner im Verbund Forschungsdaten Bildung[21]. Der von den Antragsstellern DIPF, GESIS und IQB im Rahmen einer BMBF-Förderung im Jahr 2013 initiierte Verbund hat das Ziel, Daten der empirischen Bildungsforschung zentralisiert nachzuweisen und in den passenden Datenarchiven zu archivieren. Zudem werden die gebündelten Kompetenzen genutzt, um Schulungen und Materialien für das Forschungsdatenmanagement im Kontext der empirischen Bildungsforschung zu produzieren und der Community zur Verfügung zu stellen. Als Netzwerkpartner profitiert das DIE von den angebotenen Dienstleistungen, erhöht als Datengeber die Sichtbarkeit seiner Daten und kann an den Weiterentwicklungen und stattfindenden Diskursen aktiv teilhaben. Zur erhöhten Sichtbarkeit des Instituts tragen auch externe Beratungen bei, die das DIE bei Projekten mit disziplinärem Bezug an anderen Einrichtungen durchgeführt hat. Das DIE möchte sich zudem aktiv an Weiterentwicklungen und Innovationen im Bereich des Forschungsdatenmanagements beteiligen. So wird es als Partner im 3. Quartal beginnenden BMBF-geförderten Verbundprojekt „DDP-Bildung“ teilnehmen. Es hat eine Entwicklung von standardisierten und somit vergleich- und bewertbaren DMP zum Ziel. Weitere strategische Optionen bestehen in einer anvisierten Teilnahme am GO FAIR-Implementation Network EcoSoc-IN[22]sowie perspektivisch an einer Partizipation an einem Konsortium der NFDI[23]. Hierfür wird die strategische Option, mittelfristig als ein vom RatSWD zertifiziertes „Forschungsdatenzentrum Erwachsenenbildung“ zu firmieren geprüft[24]. Sämtliche Vernetzungsoptionen spiegeln auch den Leibniz-Auftrag nach nachhaltiger Forschung und Entwicklung wider, und werden als wichtige strategische Optionen für das Institut gesehen.

[21] Vgl. https://www.forschungsdaten-bildung.de/
[22] Vgl. https://www.go-fair.org/implementation-networks/overview/ecosoc-in/
[23] Vgl. https://www.ratswd.de/pressemitteilung/11022019
[24] Vgl. https://www.ratswd.de/forschungsdaten/fdz

4.2 Interne strategische Optionen

Zu Beginn der Implementation war das Forschungsdatenmanagement organisatorisch zwar im großen Bereich der Forschungsinfrastrukturen verortet, hatte zunächst aber keinen direkten inhaltlichen bzw. organisatorischen Bezug zu anderen Abteilungen. Im Laufe der Implementierung wurde deutlich, dass solche Bezüge notwendig sind, um den angestrebten Prozessen das nötige Gewicht zu verleihen und interne Ressourcen mit nutzen zu können. Die Bibliothek des DIE startete im Jahr 2016 einen noch andauernden Prozess der strategischen Neuausrichtung und Profilschärfung. Es bot sich an, die thematisch ohnehin verwandten Themenbereiche zu integrieren. Das Forschungsdatenmanagement wurde ein Bestandteil der neuen „Datenstrategie“ der Bibliothek, welches u.a. auch Bibliothekare als „Datenmanager“ positionieren sollte. Diese Strategie beruht auf der Schnittmenge dreier Bereiche, die im Zusammenspiel die erweiterten Aufgaben der Bibliothek widergeben. Dies zeigt die folgende Abbildung.

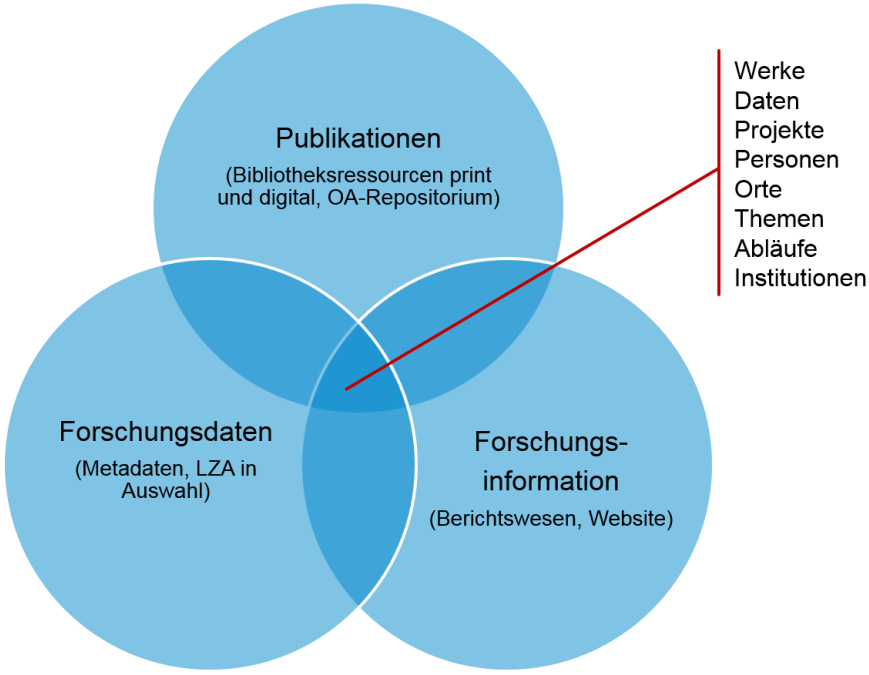


Abb.1: Datenstrategie der DIE-Bibliothek; Quelle: eigene Erstellung

1. „Publikationen“ beinhalten den „klassischen Teil der Bibliotheksarbeit mit dem Zugang und der Dokumentation von digitalen und Printpublikationen.
2. „Forschungsinformation“ beinhaltet das Berichtswesen über Projekte, Personen und verbundene Publikationen, deren Meldungen und Darstellung (z.B. auf den persönlichen Institutswebseiten der Wissenschaftler).
3. Forschungsdaten.

Allen drei Themen gemein ist die Möglichkeit bzw. Notwendigkeit, diese mit den gleichen Entitätstypen bzw. deren standardisierten Metadaten zu dokumentieren und dadurch miteinander zu verknüpfen. Neben einer angestrebten Reduzierung der Dokumentationsvorgänge an unterschiedlichen organisatorischen und technischen Stellen, kann der Zugang zu den Daten vereinfacht und vereinheitlicht werden. Zudem können inhaltliche Zusammenhänge besser aufgezeigt werden. Die Arbeit der Bibliothek wird im Institut deutlich sichtbarer und immens aufgewertet. Ein neues Subteam „Datenredaktion“ innerhalb der Bibliothek betreut die Daten dokumentarisch und entwickelt ein integriertes Datenhaltungsangebot in enger Zusammenarbeit mit der Haus-IT weiter.

5. Forschungsdatenmanager im Spannungsfeld

Es wird deutlich, dass bei einem institutionellen Forschungsdatenmanagement vielschichtige, heterogene und mitunter diametral zueinander stehende Aufgaben anfallen. Der Forschungsdatenmanager wird demnach einem stetigen Spannungsfeld ausgesetzt. Zum einen gilt es, einen wissenschaftlichen Kulturwandel, der den Wissenschaftlern eine Öffnung ihrer Arbeit abverlangt, zu moderieren. Hierbei ist ein Beharren auf neuen Konventionen und Vorgaben wenig zielführend und kontraproduktiv. Es gilt, weiterhin berechtigte Interessen der Forscher und deren gesetzlich verankerte Freiheit mit der ebenso berechtigten Forderung nach offenen und „fairen“ Daten behutsam in Einklang zu bringen. Gleichzeitig müssen aber auch die strategischen Ziele eines Leibniz-Institutes berücksichtigt werden. Ein Institut wird an der Leistung der Forschung und deren Infrastrukturen gemessen. Neben den durchaus diskussionswürdigen Bewertungsindikatoren für Infrastruktureinrichtungen muss der Forderung nach quantifizierbaren Ergebnissen und Sichtbarkeit in der externen Wahrnehmung nachgekommen werden. Es gilt, vor dem Hintergrund begrenzter Ressourcen realistische Entwicklungsstrategien und Vernetzungsoptionen zu entwickeln. Dabei ist das Feld des Forschungsdatenmanagements eines der dynamischsten Gebiete im Bereich der Forschungsinfrastrukturen. Die Vielzahl an

neuen Initiativen, Netzwerken, Förderlinien und Stakeholdern muss beobachtet und bewertet werden. Man muss entscheiden, welche Möglichkeiten sich für das eigene Institut ergeben und wie diese mit der Institutsstrategie, den vorhandenen Ressourcen und den Wünschen der Wissenschaftler in Einklang zu bringen sind. Forschungsdatenmanagement ist eine hochdynamische Querschnittsaufgabe, die entsprechende Behutsamkeit, Beharrlichkeit und Weitsicht erfordert.

6. Fazit

Nach zwei Jahren operativen Forschungsdatenmanagements am DIE lässt sich anhand des vorliegenden Erfahrungsberichts ein positives Fazit ziehen. Forschungsdatenmanagement ist auch an einem vergleichsweise kleinen Forschungsinstitut implementier- und durchführbar. Zudem erlangt Institutsbibliothek durch die Integration des Forschungsdatenmanagements in ihre Organisationsstruktur einen deutlich verbesserten Stellenwert und erhöhte Sichtbarkeit. Forschungsdatenmanagement erfordert im Spannungsfeld aus Wissenschaftlern, Institutspolitik und einem dynamischen Umfeld aus Wissenschaftsförderung und Wissenspolitik den Mut, schrittweise zu handeln. Die Implementation ist ein Prozess, der von der Institutsleitung mitgetragen werden muss. Der Kulturwandel in der Wissenschaft kann nicht nur von Seiten der Infrastrukturen initiiert werden. Es ist eine enge und dauerhafte Begleitung des Forschungsprozesses durch den Forschungsdatenmanager im Zusammenspiel mit den Wissenschaftlern unabdingbar. Vernetzung mit anderen Akteuren unterstützt dieses Vorhaben sowohl auf inhaltlicher, operativer als auch auf strategischer Ebene. Vor allem bedarf es jedoch einer großen Beharrlichkeit und Geduld, um den langfristigen Kulturwandel hin zu „fairen“ Daten mitzugestalten.

Literatur:

Letztes Abrufdatum aller in den Fußnoten angegebenen Internetquellen: 29.04.2019

Borgman, C. L. (2012). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology 63,* 1059–1078.

Cremer, F.; Engelhardt, C. & Neuroth, H. (2015). Embedded Data Manager – Integriertes Forschungsdatenmanagement: Praxis, Perspektiven und Potentiale. *Bibliothek Forschung und Praxis 39* (1), 13–31. https://doi.org/10.1515/bfp-2015-0006.

Deutsche Forschungsgemeinschaft. (2013). *Vorschläge zur Sicherung guter Wissenschaftlicher Praxis: Denkschrift.* (ergänzte Aufl.). Weinheim: Wiley-VCH.

Ludwig, J. & Enke, H. (Hrsg.). (2013). *Leitfaden zum Forschungsdatenmanagement – Handreichungen aus dem WissGrid-Projekt.* Glückstadt: vwh-Verlag.

Wilkinson, M. D; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018 EP. https://doi.org/10.1038/sdata.2016.18.

Schulung und Weiterbildung

Gemeinsam für Open Science – Forschungsbegleitende FDM-Beratung an der Universität Trier

Marina Lemaire

SeS Uni Trier

Zusammenfassung

FAIRe Daten stellen viele Forschende vor große Herausforderungen. Während für *Findable* und *Accessible* Infrastruktureinrichtungen Forschenden Lösungen bereitstellen und dafür die Verantwortung übernehmen, obliegen die Aspekte *Interoperable* und *Reusable* in erster Linie den Forschenden selbst. Sie müssen zum großen Teil zunächst die dafür notwendigen Kompetenzen erwerben und vor allen Dingen ihren Forschungsprozess umstellen, um diese Anforderungen zu erfüllen. Da dies mit einem vermeintlich hohen Arbeits- und Zeitaufwand verbunden ist, benötigen sie von ihren Forschungseinrichtungen adäquate Unterstützung. Die Universität Trier setzt hier auf ein forschungsprojektbegleitendes Beratungsprogramm, dass es den Wissenschaftler*innen ermöglicht an ihren eigenen Forschungsgegenständen Forschungsdatenmanagementqualifikation zu erwerben und die positiven Effekte des Forschungsdatenmanagements zu erleben.

Abstract

FAIR Data pose major challenges to many researchers. With regard to *Findable* and *Accessible*, the infrastructures provide solutions for researchers and take responsibility for them, whereas *Interoperability* and *Reusability* lie primarily in the researchers' own responsibility. To a large extent, they must firstly acquire the necessary competencies and, above all, adapt their research process in order to meet these requirements. Since this is associated with a seemingly high work- and time expenditure, they require support from their research institutions. Here, the Trier University relies on a research-project-accompanying consulting program that enables scientists to acquire research data management qualifications in their own objects of research and to experience the positive effects of research data management.

Problemstellung

Viele Forschende befürworten die Intentionen, die hinter Konzepten von FAIR-Data, Open Science und Open Data stehen. Sie sehen sich jedoch vor große Herausforderungen gestellt, die in diesen Termini formulierten Ziele mit einem angemessenen Arbeits- und Zeitaufwand zu erreichen. Denn während es für die Anforderung der Auffindbarkeit (*Findable*) und der Zugänglichkeit (*Accessible*) von Forschungsdaten bereits zahlreiche technische Angebote von Infrastruktureinrichtungen gibt, bleibt es den Forschenden weitestgehend selbst überlassen, wie sie die Interoperabilität (*Interoperable*) und die Nachnutzbarkeit (*Reusable*) ihrer Forschungsdaten sicherstellen. Denn diesen Ansprüchen können Forschungsdaten nur dann genügen, wenn bereits vor und während des Forschungsprozesses ein professionelles Datenmanagement durchgeführt wird. Nur auf diese Weise kann die Nachnutzung von Forschungsdaten durch andere gewährleistet werden. Für diese Ziele müssen viele Forschende ihre von der analogen Welt geprägten Arbeitsprozesse an die digitale Umwelt anpassen. Davor schrecken sie jedoch zurück, weil sie der Auffassung sind, nicht über die dafür notwendigen Kompetenzen zu verfügen. Diese Einschätzung mag zum Teil richtig sein, z.B. bezüglich der Kenntnisse über fach-/datenspezifische Metadaten- oder Dokumentationsstandards, geeignete Softwaretools bzw. die sie betreffenden rechtlichen Rahmenbedingungen. Hinzu kommt, dass das Datenmanagement oftmals als ein nicht integraler Bestandteil des Forschungsprozesses verstanden und daher als Mehrbelastung wahrgenommen wird. Dies ist jedoch ein Irrtum, wenn man sich vergegenwärtigt, dass das angewandte Datenmanagement den Forschungsprozess beeinflusst und umgekehrt. Denn einerseits haben die Forschungsdaten und die Fragestellung bzw. Methoden einen Einfluss auf die Auswahl der Metadaten-/Dokumentationsstandards und gegebenenfalls der Software. Andererseits können die in der Software implementierten Workflows die Anpassung des Forschungsprozesses und der Forschungsmethoden mit sich bringen. Diese Interferenzen gilt es bei der Entwicklung eines digitalen Forschungskonzeptes zu identifizieren und angemessen darauf zu reagieren. Dieses Bewusstsein ist jedoch noch nicht verbreitet.[1]

[1] Vgl. Lemaire 2018.

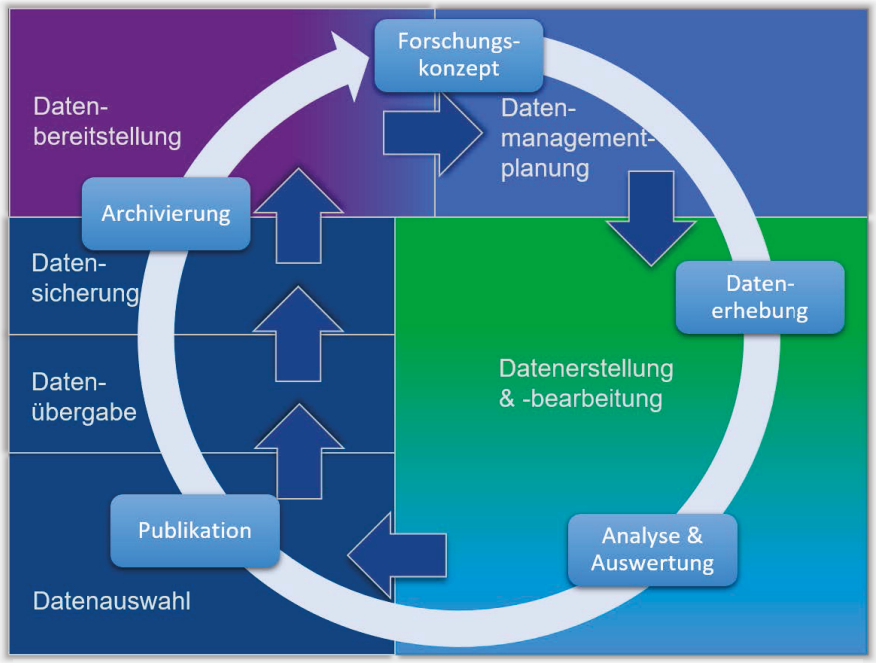


Abb. 1 Forschungsprozess und Datenmanagement sind zwei miteinander interagierende Prozesse

Der derzeitige Gemütszustand der Wissenschaftler*innen in Bezug auf das Forschungsdatenmanagement (nachfolgend FDM) kann recht gut mit dem Akronym VUCA beschrieben werden. VUCA steht für Volatilität, Unsicherheit, Komplexität und Ambiguität: Mit *Volatilität* ist die Geschwindigkeit der Veränderungen in der digitalen Welt gemeint, die zu Instabilität führen kann und einen hohen Anpassungsdruck ausübt. *Unsicherheit* entsteht durch das Fehlen von Informationen, um Konsequenzen von Entscheidungen und Handlungen besser abschätzen zu können. Sie ist auch eine Ursache der *Komplexität*, die durch die netzwerkartigen Strukturen von Informationen und Prozessen entsteht und daher schwer zu überblicken ist. Dem schließt sich die *Ambiguität* an durch fehlende Klarheit bezüglich der Beurteilung von Ursachen und Auswirkungen eines Ereignisses. Diese VUCA-Effekte lassen sich in sämtlichen von der Digitalisierung betroffenen Lebensbereichen feststellen. Solche Überforderungen führen bei Menschen, die vor eine Entscheidung gestellt werden, oft dazu, im Gewohnten zu verharren und keine neuen Wege einzuschlagen.[2]

Derzeitig ist ein Ungleichgewicht festzustellen zwischen den angebotenen Infrastrukturen und Services für die Unterstützung des FDM einerseits, insbesondere

[2] Vgl. Welpe 2018: 24–28.

für die langfristige Verfügbarkeit und Nachnutzbarkeit, sowie andererseits der Forschenden, denen die Kompetenzen fehlen, die angebotenen Services zu nutzen und ihre Forschungsdaten entsprechend der Anforderungen aufzubereiten. Um diesem Ungleichgewicht entgegenzuwirken und den Anforderungen gerecht zu werden, ist es für Forschungseinrichtungen notwendig, ihren Forschenden adäquate Hilfestellungen bei der Bewältigung dieser Herausforderungen zu leisten.

Rahmenbedingungen für Forschungsdatenmanagement und -infrastrukturen an der Universität Trier

Die Universität Trier hat seit mehr als 20 Jahren[3] systematisch die Integration moderner Informationstechnologien und informationswissenschaftlicher Forschungsansätze in den Geisteswissenschaften gefördert und zukunftsweisende profilbildende Impulse für die Entwicklung der digitalen Geisteswissenschaften gesetzt. Sie entwickelt Informationsinfrastrukturen und Dienstleistungen, die das FDM und die Durchführung IT-gestützter Forschungsvorhaben unterstützen. Ausgangspunkt für die intensivere Beschäftigung mit Daten und deren Aufbereitung und Bereitstellung für die Forschung war die Beobachtung, dass mit dem Abschluss des Sonderforschungsbereich 235[4] Forschungsdaten, die über einen Zeitraum von 15 Jahren erhoben worden waren, drohten in Vergessenheit zu geraten. Im „Informationsnetzwerk zur Geschichte des Rhein-Maas-Raumes“ (RM.net | rmnet.uni-trier.de), dessen Aufbau im DFG-Schwerpunkt „Themenorientierte Informationsnetzwerke“ gefördert wurde, konnten ausgewählte Datenbestände publiziert werden. Ausgehend von diesen SFB-Erfahrungen wurde im nachfolgenden SFB 600 „Fremdheit und Armut“[5] von Beginn an eine Plattform für das FDM aufgebaut. Im Zeitraum von 2004 bis 2012 entstand die virtuelle Forschungsumgebung FuD (VFU FuD | fud.uni-trier.de). Sie wird aktuell in über 35 (z.T. bereits abgeschlossenen) nationalen wie internationalen Forschungsvorhaben an Universitäten, außeruniversitären Forschungseinrichtungen und Bibliotheken eingesetzt. Inzwischen führt die Universität Trier die FuD-Software erfolgreich im Regelbetrieb[6] fort.

[3] So wurde z. B. 1998 das Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften (heute TCDH | Trier Center for Digital Humanities) gegründet. <https://kompetenzzentrum.uni-trier.de>.

[4] SFB 235 „Zwischen Maas und Rhein: Beziehungen, Begegnungen und Konflikte in einem europäischen Kernraum von der Spätantike bis zum 19. Jahrhundert“ (Förderung von 1987 bis 2002) <http://gepris.dfg.de/gepris/projekt/5474491>.

[5] SFB 600 „Fremdheit und Armut. Wandel von Inklusions- und Exklusionsformen von der Antike bis zur Gegenwart“ (Förderung von 2002 bis 2012) <http://gepris.dfg.de/gepris/projekt/5485009>.

[6] Vgl. Minn/Burch 2016.

Zwischen 2011 und 2013 wurde außerdem das Virtuelle Forschungsdatenrepositorium ViDa (ViDa | vida.uni-trier.de) für die Sicherung und Bereitstellung ausgewählter Forschungsdaten aus dem SFB 600 prototypisch aufgebaut. Aktuell wird es zu einer universitätsweiten Archivierungs- und Repositoriumssoftware weiterentwickelt, um die diversen disziplinenspezifischen Bedarfe für die Forschungsdatenarchivierung und -bereitstellung abzudecken. Auch ViDa wird über ein Geschäftsmodell in den Regelbetrieb überführt werden, um das System anderen wissenschaftlichen Einrichtungen zur Verfügung stellen und kooperativ weiterentwickeln zu können.

Basierend auf den Erfahrungen zur Implementierung von Informationsinfrastrukturen für die geisteswissenschaftliche Forschung hat die Universität Trier begonnen, eine Strategie zur universitätsweiten Integration eines nachhaltigen FDM zu entwickeln. Mit der Einrichtung des Servicezentrums eSciences (SeS | esciences.uni-trier.de) (gegründet 2015) wurden erste Schritte unternommen, bestehende universitäre Organisationsstrukturen zu überprüfen und innovative Servicestrukturen für das FDM und die Unterstützung von eSciences-Vorhaben zu schaffen. Dabei stehen die Bedarfe der Forschenden[7] im Fokus, um die notwendigen FDM-Services bereitzustellen. In diesem Zusammenhang erarbeitet das PODMAN-Projekt[8] ein FDM-Referenzmodell und ein zugehöriges prozessorientiertes Benchmarking-Verfahren. Über diese Instrumente soll Universitäten und außeruniversitären Forschungseinrichtungen ein Orientierungsrahmen bereitgestellt werden, den sie flexibel zur Umsetzung einer eigenen FDM-Strategie nutzen können.

Die Philosophie des Servicezentrums eSciences für die Bereitstellung von Forschungsinfrastrukturen, insbesondere der VFU FuD und ViDa, sehen und sahen stets ein ganzheitliches Begleitungs- und Beratungskonzept für die Anwender*innen vor: von der Planung über die Projektdurchführung bis zum Projektabschluss. Dies ergibt sich zum einen aus dem Anspruch Softwaresysteme (weiter) zu entwickeln, die den gesamten Forschungsprozess unterstützen und zudem von nachfolgenden Projekten/Einrichtungen nachgenutzt werden können. Zum anderen fällt die Transformation der Forschungsarbeit ins Digitale vielen Forschenden noch schwer

[7] Für die Bedarfserhebung an der Universität Trier wurde 2015 eine Umfrage durchgeführt. Vgl. Lemaire / Rommelfanger u. a. 2016.
[8] BMBF-Projekt: „Prozessorientierte Entwicklung von Managementinstrumenten für Forschungsdaten im Lebenszyklus“ (PODMAN) <www.fdm.uni-trier.de>. Vgl. Blask / Förster u. a. 2019.

und zum Teil wird Expertenwissen im Umgang mit neuen Technologien benötigt, das durch die intensive Zusammenarbeit mit den Fachkoordinator*innen des Servicezentrum eSciences deutlichleichter ins Forschungsteam integriert werden kann. Dies entlastet die eigentliche Forschungsarbeit, weil sowohl die Einstiegshürde für den Einsatz digitaler Werkzeuge und Methoden geringer, als auch die Lernkurve für die Wissenschaftler*innen flacher sind und somit mehr Zeit für die eigentliche Forschung bleibt. Zudem hat sich in Trier vielfach die Zusammenarbeit in interdisziplinären Teams (Informatiker*innen, Fachwissenschaftler*innen, Informationswissenschaftler*innen und Hardware-Expert*innen) als ein erfolgreicher Ansatz bewährt, da die für die Projektumsetzung notwendigen Kompetenzen gebündelt und so effizienter und zielführender eingesetzt werden können (als dies eine Einzelperson vermag). Digitalisierung gleich in welchem Lebensbereich ist eine Gemeinschaftsaufgabe und bedarf der Kooperation verschiedener Expert*innen.[9]

Um nachhaltige, nachnutzbare Forschungsdateninfrastrukturen wie FuD und ViDa zu entwickeln, müssen und mussten die Entwickler*innen die verschiedenen Phasen des Forschungsprozesses in unterschiedlichen Projekten analysieren, dabei die einzelnen Arbeitsschritte vom projektspezifischen Kontext abstrahieren und zu einer allgemeinen technischen Funktion modellieren. Bei der Analyse einzelner Forschungsvorhaben bilden die klassischen Fragen eines Datenmanagementplans[10] das Grundgerüst und werden im Gespräch mit den Forschenden dahingehend vertieft und verfeinert, dass am Ende logisch aufeinander aufbauende Arbeitsschritte (Workflows) definiert werden können, die den Forschungsprozess repräsentieren. Der Abgleich von Daten- und Workflowlogiken mehrerer Projekte, die zunächst als Einzelfälle in ihrem jeweiligen Forschungskontext betrachtet werden, münden zusammengenommen in ein nachhaltiges adaptives Softwarekonzept, das durch modularen Aufbau, Schnittstellen- und Konfigurationsoptionen die individuellen Bedarfe im konkreten Anwendungsfall abdeckt. Auf diese Weise entstehen nachnutzbare Forschungsinfrastrukturen, die nicht nur den gesamten Datenlebenszyklus unterstützen, sondern damit auch wichtige Eigenschaften für einen nachhaltigen Software-Betrieb besitzen.[11] So hat sich in Trier ein großer Erfahrungsschatz nicht nur für die Entwicklung von Forschungssoftware, sondern ebenso für die FDM-Unterstützung von Forschungsprojekten angesammelt,

[9] Vgl. u.a. Fournier 2017; Schircks 2017: 18.
[10] Vgl. Ludwig/Enke 2013; Minn / Lemaire 2017.
[11] Vgl. Buddenbohm/Enke u.a. 2014: 17–19; Katerbow/Feulner u.a. 2018.

von dem die Forschenden insbesondere der Geistes- und Sozialwissenschaften auch unabhängig vom Einsatz der bereitgestellten Forschungsdateninfrastrukturen profitieren. Davon ausgehend ist das forschungsbegleitende FDM-Beratungskonzept entwickelt worden.

FDM-Beratungskonzept der Universität Trier

Um den oben skizzierten Problemlagen zu begegnen, bietet das Servicezentrum eSciences ein forschungsprojektbegleitendes Beratungsprogramm an. Es zielt darauf ab, Forschende bei der Entwicklung und Umsetzung ihrer FDM-Strategie während des gesamten Forschungsprozesses zu begleiten. Dies sieht vor, zu Projektbeginn gemeinsam mit den Forschenden ein digitales Forschungskonzept zu entwickeln unter Berücksichtigung der Ziele, der Fragestellung, der Forschungsdatenarten sowie der verwendeten Methoden und Werkzeuge sowie die Auswahl eines geeigneten Repositoriums zu unterstützen. Hierbei bildet der Projektantrag - insbesondere die Abschnitte Ziele, Fragestellung, Methoden und Arbeitsprogramm - die Ausgangsbasis und wird in einem Einzelgespräch vertieft. Je nach Forschungsdesign wird gemeinsam ein Konzept für die praktische Umsetzung der FDM-Ziele erarbeitet: zum Beispiel die Auswahl eines Software Tools, wie man mit diesem einen vorgegebenen Metadatenstandard erfüllen oder seine Analysemethode anwenden kann.

Die Forschenden haben die Möglichkeit, zweimal während der Projektdurchführung gemeinsam mit den Fachkoordinator*innen die Praktikabilität und den Erfolg ihres digitalen Forschungskonzeptes zu evaluieren und je nach Ergebnis ihre FDM-Strategie gegebenenfalls anzupassen. Hierbei schauen sich die Fachkoordinator*innen bei Bedarf die erstellten Daten an und beurteilen mit, ob sie den Anforderungen der gewählten FDM-Strategie genügen und diskutieren gegebenenfalls Lösungsmöglichkeiten, wie sie erreicht werden können. Hierbei können auch Entwicklungen berücksichtigt werden, z.B. wenn sich Änderungen bezüglich von Fach oder datenspezifischen Leitlinien ergeben haben, die ausgewählten Standards weiterentwickelt wurden oder sich in dem anvisierten Forschungsdatenarchiv/-repositorium Änderung für die Datenübergabe ergeben haben.

Am Ende eines Projektes werden die Wissenschaftler*innen bei der Datenübergabe in ein Forschungsdatenarchiv/-repositorium unterstützt. Dies kann beispielsweise die Hilfestellung bei der Schließung eines Datenübergabevertrages, die Auswahl einer geeigneten Lizenz oder die Bedienung des Ingest-Tools umfassen.

Praxisbeispiel Forschungsdatenmanagementberatung

Bei der Erarbeitung eines digitalen Forschungskonzeptes orientieren wir uns zunächst an den üblichen Fragen, die in einem Datenmanagementplan wiederzufinden sind. Hierzu haben wir einen auf die Universität Trier angepassten Datenmanagementplan für die geistes- und sozialwissenschaftliche Forschung zusammen mit dem Forschungszentrum Europa entwickelt.[12] Der erste Einstieg und damit die Vorbereitung des Beratungsgesprächs erfolgt über das Exposé des Projektes, dem im besten Fall die Fragestellung und Ziele, die Untersuchungsmaterialien und Methoden sowie die geplanten Arbeitsabläufe zu entnehmen sind. Hierzu notieren wir uns FDM-spezifische Besonderheiten, die uns auffallen. Dies können beispielsweise größere Speicherbedarfe sein, spezifische Software, die die Daten nur in einem proprietären Format ausgibt, zu erfüllende Metadatenstandards oder rechtliche Aspekte. Im Gespräch selbst wird die Wissenschaftler*in zunächst aufgefordert, ihre Perspektive auf die sich stellenden Probleme bei der Umsetzung einer FDM-Strategie zu schildern und an welchen Stellen sie sich in erster Linie Unterstützung wünscht. Auf diese Weise gewinnt die Berater*in einen ersten Eindruck über die bereits vorhandenen FDM-Kompetenzen und kann zunächst auf die drängendsten Probleme aus der Sicht des Forschenden eingehen, um so eine Vertrauensbasis zu schaffen. Im weiteren Verlauf des Gespräches geht die Berater*in auf die zuvor identifizierten FDM-spezifischen Aspekte ein und diskutiert diese mit der Wissenschaftler*in. Gemeinsam wird der Datenmanagementplan entworfen oder fortgeschrieben, der die FDM-Anforderungen beschreibt und die Strategien für deren Erfüllung beinhaltet.

Werden in einem Forschungsprojekt z.B. Befragungen oder Interviews durchgeführt, so gehen wir bei Bedarf gemeinsam die Vorlage für die informierte Einwilligung durch und überlegen, welche Passagen für dieses Forschungsprojekt relevant sind. Ebenso sprechen wir darüber, wie die Anforderungen der Ablage personengeschützter Daten unter den Rahmenbedingungen des Projektes an der Universität Trier umgesetzt werden können.

Besteht für ein Forschungsprojekt insbesondere Bedarf bei der Datenerhebung, kann es bei der Auswahl der geeigneten Software Unterstützung erhalten. Hierbei unterstützen wir die Evaluation und achten insbesondere darauf, ob die in der Software implementierten Workflows mit den geplanten Arbeitsprozessen des Projektes

[12] Vgl. Minn / Lemaire 2017.

zusammenpassen; ob die zu erfüllenden Datenauszeichnungsstandards damit umzusetzen sind; ob ein offenes Ausgabeformat für die spätere Archivierung möglich ist; ob die Lizenzbedingungen besondere Auswirkungen haben u.v.m. Je nachdem für welches Softwareprodukt sich die Wissenschaftler*in entscheidet und sofern dies gewünscht ist, besprechen wir, wie sie ihre Daten damit aufbereiten kann. Oftmals möchten die Forschenden die ihnen vertrauten Softwareprodukte der Text- und Tabellenverarbeitung verwenden. Dann ist es uns besonders wichtig, mit ihnen gemeinsam die Tücken der nicht von einem Datenbanksystem unterstützten strukturierten Datenerfassung zu besprechen, um typische Fehler zu vermeiden. Hierbei erörtern wir auch Problemlagen, für die wir möglicherweise aktuell noch keine zufriedenstellende technische Lösung finden, worauf die Wissenschaftler*in in ihrer Forschungsarbeit jedoch achten sollte, um zu einem späteren Zeitpunkt einen adäquaten Lösungsansatz zu finden. Diese Aspekte werden dann in den beiden Evaluationsphasen während der Projektlaufzeit nochmals aufgegriffen.

Wenn für die Evaluationsphasen vereinbart wurde, dass die Fachkoordinator*in die bereits erhobenen Daten begutachtet, die beispielsweise in einer Exceltabelle erfasst wurden, dann wird die Tabellenstruktur und damit die Datenmodellierung beurteilt sowie ob die Daten konsistent erscheinen. Hier achten wir z.B. darauf, ob ein einheitliches Datumsformat verwendet wird, ob Personennamen nach einem nachvollziehbaren Schema konsistent erfasst sind bzw. ob die eindeutige Identifizierung gegeben ist. Was genau geprüft wird, ist selbst verständlich vom jeweiligen Forschungsprojekt abhängig.

Resümee und Ausblick

Die Rückmeldungen der Forschenden, die eine solche Beratung in Anspruch nehmen, zeigen uns, dass es richtig ist den Wissenschaftler*innen durch die Einzelberatung die Möglichkeit zu geben, sich an ihren eigenen Forschungsgegenständen die notwendigen FDM-Kompetenzfelder praxisnah zu erschließen und die positiven Effekte unmittelbar in ihrer Forschungsarbeit zu erleben. Von großem Vorteil ist dabei, dass die Fachkoordinator*innen des Servicezentrums eSciences eigene Erfahrungen in der disziplinübergreifenden geisteswissenschaftlichen Forschung mitbringen, bereits eine Vielzahl an geistes- und sozialwissenschaftlichen Forschungsprojekten mit ihren diversen Methoden begleitet haben und sich dadurch intensiv mit dem jeweiligen Forschungskonzept auseinandersetzen können, um gemeinsam mit den Forschenden auf Augenhöhe ein digitales Forschungskonzept zu entwickeln, das

möglichst nah an den gewohnten Arbeitsprozessen bleibt. Auf diese Weise erleben die Wissenschaftler*innen wie der Forschungsprozess und das Datenmanagement ineinandergreifen. Ebenso stellen wir immer wieder fest, dass sich der Zeitaufwand, den eine 1:1-Beratung mit sich bringt, lohnt, weil es bereits Beratenen gelingt, die gewonnenen Kompetenzen auf neue Problemstellungen bzw. Projekte zu übertragen und eigene Lösungsstrategien zu entwickeln. Auch wenn den Forschenden zunächst eine steile Lernkurve abverlangt wird, da es häufig schwer fällt, den Forschungsgegenstand/ die -methode auf ein Datenmodell zu abstrahieren und/oder der Ein- und Umstieg in digitale Arbeitsweisen durch die Komplexität des Forschungsgegenstandes erschwert wird, zahlt sich der Aufwand im Projektverlauf bzw. bei Nachfolgeprojekten dennoch aus.

Auch wenn wir in der Beratung versuchen, den Forschenden die unserer Ansicht nach optimaleren Lösungswege aufzuzeigen, um möglichst nah an die Open Science Anforderungen heranzukommen, so versuchen wir, den von der Wissenschaftler*in gewählten Weg so gut wie möglich zu unterstützen und mitzutragen, denn wir verstehen uns als Assistent*innen, die den Weg zu FAIRen Daten ebnen und die Akzeptanz erhöhen möchten. Dies ist unseres Erachtens nur mit einer sensiblen auf die Bedürfnisse und Befindlichkeiten der Forschenden eingehenden Beratung möglich, die auf Augenhöhe erfolgen muss, damit der bestehenden Verunsicherung in Bezug auf FDM entgegengewirkt werden kann.

Wir sind uns bewusst, dass wir nur aufgrund der oben beschriebenen Rahmenbedingungen an der Universität Trier in der Lage sind, unseren Geistes- und Sozialwissenschaftler*innen diesen Service zu bieten, weil wir über Jahre hinweg für diese Fachdisziplinen die entsprechenden Kompetenzen aufbauen konnten. Da die Universität Trier einen geistes- und sozialwissenschaftlichen Schwerpunkt hat, gelingt es uns mit diesem Angebot einen großen Teil unserer Forschenden fachspezifische FDM-Unterstützung anbieten zu können. Dennoch ist es für uns essentiell, auch für die anderen Fachdisziplinen qualitativ hochwertige fachspezifische FDM-Angebote machen zu können. Für die Psychologie an der Universität Trier haben wir glücklicherweise das Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID | www.leibniz-psychology.org) vor Ort. Aber beispielsweise für unsere Wissenschaftler*innen der Raum- und Umweltwissenschaften müssen wir noch Lösungen finden, die eine adäquate, bedarfsorientierte FDM-Unterstützung gewährleisten. Hierbei werden uns die im oben genannten PODMAN-Projekt

entwickelten Instrumente des DIAMANT-Modells[13] unterstützen, insbesondere um zunächst die konkreten fachspezifischen Bedarfe und Kompetenzen zu ermitteln.[14]

Literatur

Blask, Katarina / Förster, André u. a. (2019): Wege Zur Optimierung Des Forschungsdatenmanagements – Die Forschungsperspektive Des PODMAN-Projekts | Request PDF. In: Bilbiothek. Forschung Und Praxis 43 (1), S. 61–67.

Blask, Katarina / Förster, André u. a. (2018a): Anforderungskataloge Für Fachspezifische FDM-Services. <https://doi.org/10.25353/UBTR-3301-5402-77XX>.

Blask, Katarina / Förster, André u. a. (2018b): Forschungsprozessspezifische Kompetenzmatrix für die Einführung des Forschungsdatenmanagements (FDM). <https://doi.org/10.25353/ubtr-2751-5387-20xx>.

Blask, Katarina / Förster, André u. a. (2018c): Qualifizierungskonzept für das Forschungsdatenmanagement an Hochschulen und außeruniversitären Forschungseinrichtungen. <https://doi.org/10.25353/ubtr-8061-5402-08xx>.

Buddenbohm, Stefan / Enke, Harry u. a. (2014): Erfolgskriterien für den Aufbau und nachhaltigen Betrieb Virtueller Forschungsumgebungen (DARIAH - DE Working Papers, 7). Göttingen. <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2014-5-4>.

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. In: o-bib. Das offene Bibliotheksjournal 2017/3, S. 88–93.

Katerbow, Matthias / Feulner, Georg u. a. (2018): Handreichung zum Umgang mit Forschungssoftware. Hg. v. Arbeitsgruppe Forschungssoftware im Rahmen der Schwerpunktinitiative Digitale Information der Allianz der deutschen Wissenschaftsorganisationen Allianz-AG Forschungssoftware. [Online]. <http://doi.org/10.5281/zenodo.1172970>.

Lemaire, Marina (2018): Vereinbarkeit von Forschungsprozess und Datenmanagement. Forschungsdatenmanagement nüchtern betrachtet. In: o-bib. Das offene Bibliotheksjournal 5 (4), S. 237–247.

Lemaire, Marina / Rommelfanger, Yvonne u. a. (2016): Umgang mit Forschungsdaten und deren Archivierung. Bericht zur Online-Bedarfserhebung an der Universität Trier (eSciences Working Papers, 02). Trier. <http://nbn-resolving.de/urn:nbn:de:hbz:385-10156>.

Ludwig, Jens / Enke, Harry (Hg.) (2013): Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Glückstadt. <http://nbn-resolving.de/urn:nbn:de:gbv:7-isbn-978-3-86488-032-2-8>.

Minn, Gisela / Burch, Thomas u. a. (2016): FuD2015 – Eine virtuelle Forschungsumgebung für die Geistes- und Sozialwissenschaften auf dem Weg in den Regelbetrieb (eSciences Working Papers, 01). Trier. <http://nbn-resolving.de/urn:nbn:de:hbz:385-10103>.

[13] Vgl. Blask / Förster u. a. 2018b; Blask / Förster u. a. 2018a; Blask / Förster u. a. 2018c.

[14] Das DIAMANT-Modell wird in dem Vortrag „Das DIAMANT-Model – Die Einführung eines multiperspektivischen Referenzmodells für die Implementierung von Forschungsdatenmanagement-Services und -Infrastrukturen“ im Panel „Strategien“ vorgestellt.

Minn, Gisela / Lemaire, Marina (2017): Forschungsdatenmanagement in den Geisteswissenschaften. Eine Planungshilfe für die Erarbeitung eines digitalen Forschungskonzepts und die Erstellung eines Datenmanagementplans (eSciences Working Papers, 03). Trier. <http://nbn-resolving.de/urn:nbn:de:hbz:385-10715>.

Schircks, Arnulf D. (2017): Die Arbeitswelt 4.0 kompetent gestalten. In: Strategie für Industrie 4.0: Praxiswissen für Mensch und Organisation in der digitalen Transformation. Hg. v. Arnulf D. Schircks, Randy Drenth u. Roland Schneider. Wiesbaden, S. 1–34. <https://doi.org/10.1007/978-3-658-16752-3>.

Welpe, Isabell M. / Brosi, Prisca u. a. (2018): Digital Work Design: Die Big Five für Arbeit, Führung und Organisation im digitalen Zeitalter, plus E-Book inside. 1. Aufl. Frankfurt New York.

Data Literacy Education – Kooperative Vermittlung von Kompetenzen für Digitales Datenmanagement am Beispiel des neuen Masterstudiengangs Digitales Datenmanagement der HU Berlin und FH Potsdam

Maxi Kindling[1], Laura Rothfritz[2]

[1]Institut für Bibliotheks- und Informationswissenschaft, Humboldt-Universität zu Berlin
[2]Fachbereich Informationswissenschaften, Fachhochschule Potsdam

Einleitung

Am Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin und dem Fachbereich Informationswissenschaften der Fachhochschule Potsdam wird ein neuer weiterbildender Masterstudiengang Digitales Datenmanagement (DDM) entwickelt, der zum Sommersemester 2020 starten soll. Er greift ein Desiderat an Aus- und Weiterbildung für die Data Literacy Education auf, das nicht nur seitens der Informationsinfrastruktureinrichtungen im Bereich Wissenschaft und Forschung formuliert wird. Der kompetente Umgang mit digitalen Daten in verschiedenen Szenarien wird in allen Domänen des digitalen Lebens und Arbeitens gebraucht. Der Studiengang DDM legt den Fokus auf Wissenschaft und Forschung und basiert auf Vorarbeiten an beiden Einrichtungen, die sich seit vielen Jahren mit diesen Themen befassen. Außerhalb der bibliotheks- und informationswissenschaftlichen Lehre und Forschung (im Folgenden abgekürzt: LIS, Library and Information Science) konnten diese Angebote bislang aber kaum Sichtbarkeit erreichen. Der Studiengang fokussiert auf Querschnittsthemen wie sie auch in Referenzrahmen formuliert werden, die außerhalb der Informationsinfrastruktur in Wissenschaft und Forschung international Beachtung finden. Die Studieninhalte von DDM bilden einen umfassenden Ansatz für Data Literacy Education ab. Sie vermitteln für Interessierte mit fachlichem Hintergrund anderer Domänen oder beruflichen Zielen außerhalb von Wissenschaft und Forschung unter anderem wichtige Rahmenbedingungen im Umgang mit digitalen Daten. Ergänzend wird die kritische Auseinandersetzung des Umgangs mit Daten unter Bezugnahme auf aktuelle Diskurse integriert.

Data Literacy Education und Datenmanagement: Bedarf

Der digitale Wandel der Wissenschaft und seine Implikationen für forschungsunterstützende Services sind bereits seit vielen Jahren Schwerpunkte in

der LIS. Traditionell werden Absolventinnen[1] dieser Disziplin in Informationsinfrastruktureinrichtungen gebraucht. Kompetenzen im Umgang mit Daten, die aus Digitalisierung und Vernetzung resultieren, werden heute in allen Wissenschaftsdisziplinen, in der Wissenschaftsverwaltung und den Domänen Kultur, Wirtschaft und öffentliche Verwaltung benötigt. Der Stifterverband für die Deutsche Wissenschaft etwa bezeichnet die "grundlegende Kompetenz, um in der digitalen Welt in Wissenschaft, Arbeitswelt und Gesellschaft bestehen und teilhaben zu können" als "Data Literacy". Gemeint ist "die Fähigkeit, planvoll mit Daten umzugehen und sie im jeweiligen Kontext bewusst einsetzen und hinterfragen zu können. Dazu gehört: Daten zu erfassen, erkunden, managen, kuratieren, analysieren, visualisieren, interpretieren, kontextualisieren, beurteilen und anzuwenden. Data Literacy gestaltet die Digitalisierung und die globale Wissensgesellschaft in allen Sektoren und Disziplinen. Gleichzeitig müssen Hochschulabsolventinnen aller Fächer über fachspezifische Datenkompetenzen für die Wissenschaft und für die Arbeitswelt verfügen."[2] Daraus wird der Begriff "Data Literacy Education" abgeleitet.

Der Stifterverband schlägt vor, die digitale Transformation als Bezugspunkt für die Entwicklung neuer Curricula (sogenannte Curricula 4.0) zu betrachten (vgl. Michel et al., 2018, S.3). Hierfür sollen technologische und gesellschaftliche Trends für die Kompetenzentwicklung sowie die Rückwirkungen dieser Trends auf Lebens- und Arbeitswelten der Zukunft berücksichtigt werden.(vgl. Michel et al., 2018, S.5 These 1)

Verschiedene Institutionen unterstreichen regelmäßig die Notwendigkeit einer Entwicklung entsprechender Aus- und Weiterbildungsangebote für die Wissenschaft. So formulierte beispielsweise der Rat für Informationsinfrastrukturen (RfII) 2016 die Anforderung, neue Studiengänge und Berufsbilder für das Management von Forschungsdaten zu entwickeln. (Vgl. RFII, 2016, S. 49ff.) Auch die Schwerpunktinitiative „Digitale Information" der Allianz der deutschen Wissenschaftsorganisationen empfiehlt die Ausbildung von Spezialistinnen für das Datenmanagement. (Vgl. Allianz, 2018, S. 3) Der Bedarf an Kompetenzvermittlung für Forschungsdaten im deutschsprachigen Raum wurde zuletzt im Rahmen der Podiumsdiskussion der Konferenz der deutschen Community der Research Data Alliance im Februar 2019 (RDA-DE) ausdrücklich hervorgehoben.

[1] Aus Gründen der besseren Lesbarkeit wird im Folgenden das Femininum verwendet. Selbstverständlich sind Personen jeden Geschlechts angesprochen.
[2] URL: https://www.stifterverband.org/data-literacy-education

Für die Aus- und Weiterbildung sind verschiedene Ansätze für unterschiedliche Zielgruppen vorstellbar und sinnvoll, die mit verschiedenen Qualifizierungszielen verknüpft sind. (Vgl. Neuroth et al., erscheint) Der Wissenschaftsrat erkannte jüngst in seinen "*Empfehlungen zu hochschulischer Weiterbildung als Teil des lebenslangen Lernen"* die Vermittlung von Kompetenzen für das digitale Arbeiten als eine lebenslange Aufgabe an, bei der die Weiterbildung an Hochschulen eine wichtige Rolle spielt. (Vgl. Wissenschaftsrat, 2019) Das Datenmanagement ist in diesem Kontext ein wichtiges Berufsfeld, das hochgradig dynamisch und in unterschiedlichen Domänen relevant ist.

In Deutschland gibt es bislang vergleichsweise wenige Weiterbildungsangebote an Hochschulen. (Vgl. Wissenschaftsrat, 2019) Diese allgemein auf Weiterbildungsangebote bezogene Aussage konnte auch mit dem Fokus auf Weiterbildung im Bereich Datenmanagement durch eine Erhebung bestätigt werden, die 2017 an der FH Potsdam durchgeführt wurde. Sie erfasste Studiengänge, die sich mit digitalen Daten im weitesten Sinne beschäftigen und zielte darauf ab, Angebote mit einer inhaltlichen Ausrichtung auf das Datenmanagement zu identifizieren: "Die Analyse ergab, dass es zurzeit kein einziges Studienangebot in Deutschland mit dem Schwerpunkt Datenmanagement gibt, weder berufsbegleitend noch konsekutiv." (Neuroth et al., erscheint) Es gibt aber zumindest zwei Hochschulen in Deutschland, die das Thema in Studienschwerpunkten, Modulen oder Teilmodulen von entsprechenden Studiengängen aufgreifen wie der Bachelorstudiengang Data and Information Science (B.Sc.) an der TH Köln[3] und der berufsbegleitende Masterstudiengang Bibliotheksinformatik (M.Sc.) an der TH Wildau.[4]

Der Studiengang Digitales Datenmanagement

Der weiterbildende Master Digitales Datenmanagement (120 ECTS) wird kooperativ von FH Potsdam und HU Berlin nach dem Konzept des Blended Learning innerhalb von vier Semestern angeboten. Der Studienstart ist für das Sommersemester 2020 geplant.[5] In den ersten drei Semestern sind Präsenztermine wahrzunehmen, die durch Phasen des Selbststudiums ergänzt werden. Diese finden zum Teil in Berlin und zum

[3] https://www.th-koeln.de/studium/data-and-information-science-bachelor_52793.php
[4] https://www.th-wildau.de/studieren-weiterbilden/studiengaenge/bibliotheksinformatik-msc-berufsbegleitendes-studium/
[5] Alle hier gemachten Angaben zum Studiengang gelten vorbehaltlich der Einrichtungsgenehmigungen durch die Hochschulleitungen und die Ministerien der beiden Bundesländer.

Teil in Potsdam statt. Die Modulkurse können auch einzeln absolviert und mit einem Zertifikat abgeschlossen werden.

Das Curriculum ist anhand der Themenblöcke Rahmenbedingungen, Technologien und Methoden strukturiert, die jeweils ein Modul bilden. Sie werden in den ersten drei Semestern durchlaufen und jeweils mit einer schriftlichen Ausarbeitung abgeschlossen. Die Pflichtmodule umfassen folgende Inhalte:

Modul	Modulkurse	Inhalte
Rahmenbedingungen des Datenmanagements	Theoretische Grundlagen Datenmanagement und Data Literacy	● Begriffsbestimmung Tätigkeitsbereiche im Datenmanagement ● Forschungsprozesse ● Informationssysteme zur Unterstützung der Forschung und zum Forschungsmonitoring (FIS)
	Forschungs- und Informationsinfrastrukturen	● Kulturpolitische, organisatorische und technische Dimensionen von Forschungs- und Informationsinfrastrukturen ● Nationale und Internationale Förderstrukturen ● Wissenschaftssysteme
	Open Access, Open Data, Open Science	● Nationale und internationale Entwicklungen ● Rechtliche Aspekte des Datenmanagements (Urheberrecht, Datenschutzrecht) ● Lizenzierungsmodelle
	Metadaten, Standards, Interoperabilität	● Etablierte disziplinäre Metadatenstandards ● Anforderungen an Metadaten (z. B. FAIR-Prinzipien) ● Technische, syntaktische und semantische Interoperabilität
Technologien des Datenmanagements	Informationstechnologische Grundlagen: Internet- und Webtechnologien	● Protokolle und Internetstandards ● Client-Server-Konzept ● Virtuelle Maschinen
	Informationstechnologische Grundlagen: Datenmanagementsysteme	● Datenmodelle und Datenrepräsentation (z. B. RDB, XML-DB, TripleStore, NoSQL) ● Abfragesprachen (z. B. SQL, XQuery, SPARQL)

		● Datentransformation, -mapping und Datenintegration
	Einführung in Algorithmen und Datenstrukturen	● Abstrakte Datenstrukturen und ihre Repräsentation (z. B. XML, JSON, RDF) ● Verfahren zur Verarbeitung abstrakter Datenstrukturen (z. B. SAX-Parser, XSLT) ● Anwendungsbeispiele (z. B. Indexstrukturen in Datenbanken)
	Digitale Repositorien	● Kategorien und Grundanforderungen an digitale Repositorien ● Qualitätsmerkmale und Zertifikate für Repositorien ● Persistent Identifier Systeme
Methoden des Datenmanagements	Forschungsdaten-management	● Lebenszyklus von Forschungsdaten ● Workflows und Tools im Forschungsdatenmanagement ● Grundlagen der digitalen Langzeitarchivierung
	Datenmanagementpläne	● Anforderungen und Beispiel für DMPs ● Tools zur Erstellung von DMPs ● Verbreitung und Nutzen von DMPs
	Statistische Methoden in der Datenaufarbeitung und –auswertung	● Quantitative Forschungsmethoden ● Auswertung mit geeigneten Anwendungen: deskriptive Statistik ● Testverfahren, schließende Statistik, multivariate Verfahren
	Datenanalyse und Datenvisualisierung	● Knowledge Discovery in Databases ● Big Data/Smart Data ● Methoden und Technik der Datenanalyse und -visualisierung

Im vierten Semester erfolgt die Erarbeitung und Verteidigung der Masterarbeit. Die Verbindung zwischen Wissenschaft und Praxis ist ein wichtiger Bestandteil des Studiums. Sie wird durch im Studienverlauf fest integrierte Projektarbeiten und

Reallabore[6] realisiert. Der Studiengang wird ausführlich beschrieben in Petras et al. (2019).

Kompetenzbereiche internationaler Frameworks

Neben der didaktischen Herausforderung, Kompetenzvermittlung für Personen aus verschiedenen Domänen (Wissenschaft, Kultur, Verwaltung, Wirtschaft) zu realisieren, besteht eine weitere Herausforderung in der Gestaltung der Studieninhalte im hochgradig dynamischen Umfeld des Datenmanagements. Ein wichtiger Ansatz ist hier die Orientierung an internationalen Entwicklungen, wie sie beispielsweise in Referenzrahmen (im Folgenden: Frameworks) zur Data Literacy dargestellt werden. Die Frameworks greifen als Referenzwerke aktuelle und als relevant identifizierte Kompetenzbereiche auf, bilden diese systematisch ab und beschreiben Qualifikationsziele. Zugleich liefern sie Ansatzpunkte für die terminologische Auseinandersetzung. In den letzten Jahren setzten sich vor allem im angloamerikanischen Raum Bibliotheken sehr differenziert mit dem Bereich der Kompetenzvermittlung für den Umgang mit Daten auseinander. Durch die Herausbildung des neuen Berufsprofils “Data Scientist” und die Entwicklung der European Open Science Cloud (EOSC) zeigte sich das Desiderat an konkreten Kompetenz- und Aufgabenbeschreibungen für das digitale Datenmanagement noch einmal spezifischer. Im Folgenden werden drei einschlägige Kompetenzframeworks vorgestellt, die es ermöglichen, die Kompetenzen, die in DDM vermittelt werden sollen, vor dem Hintergrund internationaler Entwicklungen einzuordnen.

Data Information Literacy (DIL)

Das Konzept der Data Information Literacy (DIL) wurde in einem Forschungsprojekt der Bibliotheken der Purdue University, der University of Minnesota, der University of Oregon und der Cornell University entwickelt. Hintergrund des Projektes war die Erkenntnis, dass im Zuge der Digitalisierung der Forschung zur "e-Science" neue Kompetenzen und Fähigkeiten im Bereich des Datenmanagements und der Datenkuratierung von den Forschenden erwartet werden. Diese sollen im Kontext der langjährigen Tradition der Vermittlung von Informationskompetenz (engl: Information Literacy) von Bibliotheken an Forschende vermittelt werden. Ziele des zweijährigen Projektes waren der Aufbau einer Infrastruktur in Bibliotheken für die Vermittlung von

[6] Reallabore sind Präsenzveranstaltungen und dienen dazu, in den Seminaren erlangte Kenntnisse praxisorientiert zu vertiefen.

DIL, Forschenden die Möglichkeit zu geben, Kompetenzen passend zu ihren disziplinären Kontexten zu erlangen und die Entwicklung eines Prozesses für Bibliotheken, disziplinspezifische Curricula anzubieten und zu vermitteln. An jeder Bibliothek wurden sowohl literaturbasiert als auch mit Hilfe von Interviews und Beobachtungen unterschiedliche Formen der Vermittlung von Data Information Literacy entwickelt und getestet. (Vgl. Carlson und Johnston, 2015)

Data Information Literacy wurde in diesem Zusammenhang als neue Art der Literacy definiert, die von anderen Literacies, wie der Data Literacy, der Statistical Literacy und der Information Literacy abzugrenzen ist, jedoch auch auf ihnen aufbaut. Während Data Literacy als die Fähigkeit beschrieben wird, Datensätze zu lesen und zu verstehen, bezieht sich die Statistical Literacy auf die Fähigkeit, auf Datensätzen basierende Statistiken kritisch zu analysieren und zu interpretieren. (Vgl Carlson und Johnston, 2015, S. 15f) Information Literacy (Informationskompetenz) wird als die Gesamtheit aller Fähigkeiten und Fertigkeiten definiert, die erforderlich sind, um Informationsbedarfe festzustellen, Informationen zu beschaffen, und effizient zu verwenden. (Vgl. auch ALA, 2015). Information Literacy befähigt dazu, Konzepte, Theorien und Argumente mit Hilfe von Informationen kritisch betrachten und zu interpretieren. (Vgl. Schield, 2004)

Data Information Literacy baut auf allen diesen Literacies auf, insbesondere auf der Informationskompetenz, stellt jedoch neben dem Auffinden und Verarbeiten von Informationen gerade die Generierung von Daten in den Vordergrund. DIL bezieht sich demnach auf Kompetenzen, die über den gesamten Forschungszyklus und den Umgang mit Daten reichen.

Während Literacy den Zustand der Informiertheit und des Besitzens von einer Reihe von Kompetenzen beschreibt, werden in Kompetenzframeworks die jeweiligen Fähigkeiten beschrieben, die der Literacy zugeordnet werden können. DIL beinhaltet insgesamt zwölf Kompetenzbereiche, welche sich jeweils in mehrere konkret anwendbare Fähigkeiten untergliedern. Folgende Kompetenzbereiche werden aufgeführt:

- Disziplinspezifische Forschungskulturen
- Datenumwandlung und Interoperabilität
- Datenkuration und -nachnutzung

- Datenmanagement und -organisation
- Datenarchivierung
- Datenprozession und -analyse
- Qualität und Dokumentation von Daten
- Datenvisualisierung und Repräsentation
- Datenbanken und Datenformate
- Auffinden und Nutzen von Daten
- Ethische Aspekte der Datennutzung
- Metadaten und Datenbeschreibung

Diese Kompetenzbereiche sind für Forschende definiert, die befähigt werden sollen, ihren Forschungsprozess im Bezug auf eine datengetriebene Forschung hin zu optimieren. Es wird davon ausgegangen, dass innerhalb der Bibliotheken diese Kompetenzen bereits bestehen und somit auch vermittelt werden können.

Sapp Nelson (2017) entwickelte aufbauend auf dem DIL-Kompetenzframework eine Matrix, in der für jeden Kompetenzbereich (kognitives) Wissen, (psychomotorische) Fähigkeiten und (affektive) Einstellungen definiert und in drei Level unterteilt wurden. Die Level bezeichnen die jeweiligen Qualifizierungsgrade: Undergraduate Education (etwa: Bachelor-Niveau), Graduate Education (etwa: Master bzw. PhD-Niveau) und Data Steward (hier angesiedelt im Bereich Post-Doc oder Projektleitung). “Data Stewards” werden hier als Datenmanager auf einem Research Enterprise Level verstanden, die den Umgang mit Daten für eine große Anzahl Forschender bzw. Forscherinnengruppen organisiert. (Vgl. Sapp Nelson, 2017, S. 4) Alle Fähigkeiten bauen mit steigendem Niveau aufeinander auf, wobei das Level “Data Steward” sich insbesondere auf die Fähigkeit bezieht, die entsprechenden Kompetenzen zu vermitteln.

EDISON

Während DIL eher vor dem Hintergrund der Informationskompetenz Fähigkeiten für das Management von (Forschungs-)Daten beschreibt, wurden im Projekt EDISON eine detaillierte Kompetenzmatrix (EDISON Data Science Framework (EDSF)) sowie ein Modell-Curriculum und Berufsprofile für den Bereich der Data Science entwickelt.[7] Data Scientists werden innerhalb des Projekts als Anwender definiert, welche über ausreichendes Wissen und Expertise in den Bereichen „Business Needs“, Domänenwissen, Analytische Fähigkeiten, Programmierung und Systems Engineering

[7] URL: http://edison-project.eu/

verfügen. So können sie das Management des gesamten Forschungsprozesses durchgängig über alle Stufen des (Big-)Data-Lifecycles durchführen, um den wissenschaftlichen und wirtschaftlichen Erwartungswert einer Organisation oder eines Projekts zu erfüllen. (Vgl. Demchenko 2017, S. 12, übersetzt).

Das EDSF unterteilt Data Science in fünf Kompetenzbereiche: Data Science Analytics, Data Management, Data Science Engineering, Research/Scientific Methods und Data Science Domain Knowledge (Business Processes). Das Datenmanagement wird in EDISON durch das Berufsprofil data handling/management repräsentiert, durchgeführt von Personen mit den Berufsbezeichnungen Data Stewards, Digital Data Curators, Digital Librarians oder Data Archivists. Data Stewards werden definiert als ,,data handling and management professional whose responsibilities include planning, implementing and managing (research) data input, storage, search, and presentation. Data Stewards create data models for domain specific data, support and advice domain scientists/researchers during the whole research cycle and data management lifecycle." (Demchenko, 2017, S. 17) Data Stewards erfüllen somit eine Schnittstellen- und Vermittlungsfunktion zwischen eher technisch orientierten reinen Datenexpertinnen und Forschenden. Kompetenzen für diese Rolle bilden sich in den oben erwähnten Kompetenzbereichen Data Management sowie in den von EDISON zu allen Kompetenzfeldern erarbeiteten Fähigkeits- und Wissensbeschreibungen ab. Kompetenzbereiche beschreiben dabei übergeordnete Bereiche, in denen Fähigkeiten und Wissen praktische Anwendung finden. (Vgl. Demchenko et al., 2017) Kompetenzen bestehen in

- Strategieentwicklung für Leitlinien und Datenmanagementpläne
- Datenmodelle und Datenstandards
- Datenintegration
- Datenprovenienz
- Datensicherung, data privacy und ethische Aspekte

Wissen und spezielle Fähigkeiten werden in folgenden Bereichen definiert:

- Entwicklung und Anwendung von Datenmanagementplänen
- Datensicherungssysteme
- Anforderungen an hybride Systeme (cloud-basierte Datensicherung)
- Datenarchitekturen, -typen, -formate, -modellierung
- Datenkuration und Qualitätskontrolle
- Strategieentwicklung für Backups und Privacy
- Metadaten und Persistent Identifier Systeme
- Forschungsdatenmanagement, Open Access, Open Data, Open Science

Datenmanagement ist aus der Perspektive des EDISON-Frameworks allerdings nur ein Teil des großen Gebiets Data Science. Kompetenzen sollten daher nicht isoliert, sondern im Bezug zu anderen Bereichen der Data Science betrachtet werden. Jedoch wird durch die Aufteilung in Kompetenzbereiche und Rollen deutlich, dass für das Datenmanagement ein Schwerpunkt auf den oben beschriebenen Kompetenzen liegt.

EOSCpilot

Eine breitere Sicht auf benötigte Kompetenzen im Bezug zur Einführung der European Open Science Cloud (EOSC) wurden im Projekt EOSCpilot im Arbeitspaket 7 “Skills” erarbeitet.[8] Hintergrund hierfür sind für die EOSC formulierte Anforderungen an Fähigkeiten und Kompetenzen im Bereich Data Stewardship und (Open) Data Science. In den europäischen Mitgliedsstaaten müssten neue Professionen (Data Stewards/Datenspezialistinnen) mit Kernkompetenzen im Bereich des Datenmanagements über den gesamten Forschungskreislauf ausgebildet werden. Da bisher jedoch keine Einigkeit darüber besteht, welche Kompetenzen und Rollen es für diese Professionen gibt, wurde im EOSCpilot ein Kompetenzframework erarbeitet, das auf bestehende Frameworks (vorwiegend EDISON) aufbaut und diese insbesondere im Hinblick auf die FAIR Data Principles (Wilkinson et al., 2016) und Open-Science-Praktiken erweitert (vgl. Whyte et al., 2018). Data Stewardship wird in diesem Kontext definiert als "the formalisation of roles and responsibilities and their application to ensure that research objects are managed for long-term reuse, and in accordance with FAIR data principles" (Whyte et al., 2018, S.11). Stewardship wird auch hier als eine Aufgabe definiert, die die Bereiche Datenmanagement und -kuration, Data Science und Data Analytics, Data Service Engineering und Fachwissenschaft an ihren Schnittstellen verbindet. (vgl. ebd. S.12) Die im Framework beschriebenen Fähigkeiten werden in drei Level (verstehen - anwenden - evaluieren) unterteilt und diese Level an die jeweiligen Rollen angepasst (Forschende, Data Scientists/Analyst, Data Service Engineers und Data Manager/Curator). Das Framework unterscheidet zudem zwischen EOSC-Service-Nutzenden und -Anbietenden. Insgesamt werden 59 Fähigkeiten definiert, die folgenden Kompetenzbereichen zugeordnet werden (vgl. ebd., S. 38f.):

- Planen und designen
- Erheben und prozessieren
- Integrieren und analysieren

[8] Siehe: https://eoscpilot.eu/themes/skills

- Bewerten und bewahren
- Publizieren und veröffentlichen
- Zugänglich machen und auffinden
- Steuern und beurteilen
- Reichweite und Ressourcen
- Beraten und befähigen

Die Nutzung von Services (in diesem Fall die EOSC) erfordert von Datenmanagerinnen/Kuratorinnen insbesondere Expertise in den Kompetenzbereichen Bewertung und Speicherung/Archivierung (Dokumentation, Qualität, Formate und Migration) sowie in der Zugänglichkeit und Auffindbarkeit (Zugänglichkeit und Auffindbarkeit von Metadaten, PIDs, Qualität von Repositories) und des Governments (Offene Forschungsstrategien, Sicherheitsfragen, Kenntnisse zu Impact und Anerkennung). Schwerpunkte für Serviceanbieterinnen liegen im Bereich Planung und Design (DMPs, Datenmodelle, Metadatenspezifikationen), der Reichweite und den Ressourcen (Service und Change Management, Workflow-Design und Storage Management).

Diskussion

In einem aktuellen Diskussionspapier von Stifterverband und McKinsey & Company (vgl. Meyer-Guckel et al., 2019) wurden sieben Handlungsbereiche benannt, in denen Hochschulen ihr Bildungsangebot auf "Future Skills" ausrichten sollten. Die nachfolgenden drei Handlungsbereiche sind dabei für den Studiengang von besonderer Relevanz:

- “3. Vermittlung von Data Literacy, damit Absolventen aller Fächer mit großen Datensätzen umgehen und sie bewusst einsetzen und hinterfragen können”
- “5. Positionierung von Hochschulen als Weiterbildungsanbieter, um den großen Bedarf der Unternehmen decken zu können”
- “7. Entwicklung neue Formen der Zertifizierung und Kompetenznachweise, da nicht nur formale Hochschulabschlüsse von den Unternehmen gefordert werden, sondern auch spezielle Kompetenznachweise” (Groß, 2019)

Der Studiengang DDM greift die Handlungsbereiche auf, wobei Data Literacy als Begriff genauer zu betrachten ist, um den Studiengang und die damit verbundenen Ziele adäquat einordnen zu können. Data Literacy verstehen wir in Anlehnung an den Stifterverband und die vorgestellten Frameworks als

> *“die Kompetenz des kritischen und lösungsorientierten Umgangs mit digitalen Daten. Sie umfasst die Auseinandersetzung mit digitalen Daten, angefangen bei ihrer Entstehung über die Prozesse, Instrumente und Infrastrukturen zu ihrer Verarbeitung, Analyse und Bereitstellung inklusive Publikation bis hin zu ihrer langfristigen Sicherung und Nachnutzung. Neben dem planvollen und kritischen Einsatz von Daten für verschiedene (interdisziplinäre) Kontexte ist die kritische Auseinandersetzung, d.h. das Verstehen, Analysieren und Bewerten von rechtlichen, technischen und organisatorischen Rahmenbedingungen, Anforderungen und Lösungen bedeutend. Dieses konzeptuelle Wissen ist darüber hinaus in die verschiedenen Domänen wie Forschung und Wissenschaft, Kultur, Gesellschaft und Wirtschaft übertragbar.”* (Petras et al., 2019)

Diese Definition beinhaltet verschiedene Aspekte des Datenmanagements und bildet den Ausgangspunkt für den Studiengang DDM.

Der Studiengang ist in erster Linie auf wissenschaftspolitische, organisatorische, technische und rechtliche Aspekte im Bereich des Datenmanagements spezialisiert. Diese Ausrichtung orientiert sich an Desideraten besonders für die Qualifizierung im wissenschaftlichen Datenmanagement unabhängig von der jeweiligen Fachdisziplin. Die diskutierten Kompetenzframeworks überschneiden sich beispielsweise in den Bereichen Metadatenstandards und PID-Systeme, Konzeption und Erstellung von Datenmanagementplänen, Datenformate und -modelle und der Qualitätskontrolle von Forschungsdaten. Diese Wissensbereiche werden durch DDM abgedeckt und sind für Data Literacy und das Datenmanagement grundlegend. Auch Kompetenzen wie beispielsweise Strategieentwicklungen für das Datenmanagement, statistische Methoden und der Umgang mit technologischen Anwendungen für die Datenintegration und -analyse (Data Science) werden im Studiengang grundlegend aufgegriffen. Die Herausforderung besteht darin, weitere Wissensgebiete der Data Literacy wie die „kritische Datentheorie“ in adäquater Weise zu integrieren und zu vermitteln. In DDM wird dies in Form verschiedener Modulkurse und sogenannter Reallabore umgesetzt, die eine Vertiefung in verschiedenen, aktuellen Themenstellungen ermöglicht.

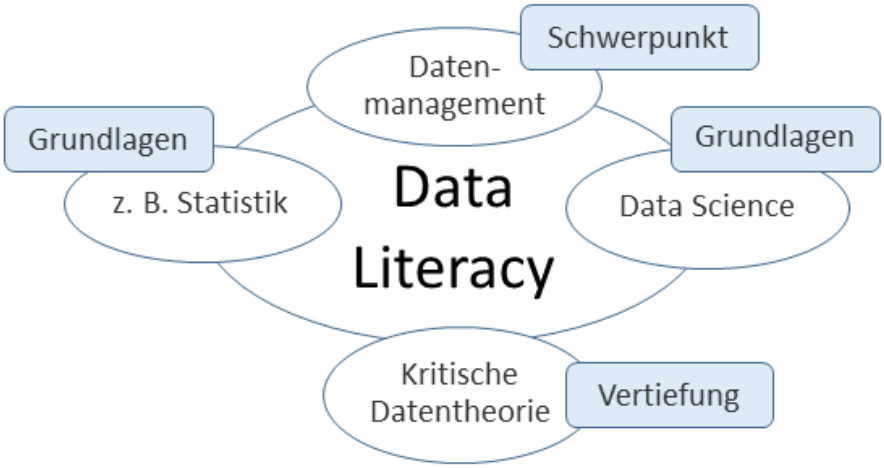


Abb. Einordnung von DDM-Studieninhalten in der Kontext der Data Literacy

Bei der Definition von Kompetenzen für das Datenmanagement ist die Betrachtung der jeweiligen Zielgruppe wichtig. DIL richtet sich speziell an Forschende und definiert keine Kompetenzen für Service-Anbieterinnen. EDISON definiert Kompetenzen für die Anwendung von datengetriebenen Forschungsprozessen (Data Science). Der Studiengang DDM deckt sich weitgehend mit dem Ansatz des EOSCpilots, der zwischen Service-Nutzerinnen und Anbieterinnen von Services wie der EOSC unterscheidet (vgl. Whyte et al., 2018). Die in DDM vermittelten Kompetenzen sollen Studierende zunächst zur grundlegenden Anwendung befähigen (Nutzerperspektive), in den Vertiefungs- und Projektmodulen auch die Konzeption von Möglichkeiten der Bereitstellung vermitteln (Anbieterinnenperspektive).

Data Stewardship wird sowohl von EDISON als auch von EOSCpilot als eine Querschnittstätigkeit gesehen, die eine vermittelnde Schnittstellenfunktion zwischen (technischem) Datenmanagement und fachwissenschaftlichen Ansprüchen einnimmt. Die Studierenden des Studiengangs DDM bringen aus ihrem vorherigen Studien- und Arbeitserfahrungen unterschiedliches Domänenwissen mit, auf das sich die im Curriculum vermittelten Kompetenzen beziehen und entsprechend aufbauen können. Der Studiengang setzt daher disziplinübergreifend an, um eine möglichst breite Ausbildung in allen Bereichen des Datenmangements zu gewährleisten.

Ziel eines Studiengangs zum Datenmanagement sollte es aus LIS-Perspektive sein, *Data Literacy* in dem oben definierten Sinne zu vermitteln und Studierende zu befähigen, kritisch und lösungsorientiert mit digitalen Daten umzugehen. Perspektivisch wird die kritische Auseinandersetzung, also das Verstehen, Analysieren und Bewerten von ethischen, rechtlichen, technischen und organisatorischen Rahmenbedingungen, Anforderungen und Lösungen an Bedeutung gewinnen. Diese Fähigkeit soll auf unterschiedliche (wissenschafts-)Domänen übertragbar sein. Absolventinnen des Studiengangs werden in der Lage sein, an der Schnittstelle zwischen Datenmanagement, Fachwissenschaft und Technologie zu arbeiten, sich kritisch mit den jeweiligen Standpunkten auseinanderzusetzen und zwischen ihnen zu vermitteln. Diese Rolle spiegelt die Beschreibungen von Data-Stewardship-Funktionen wieder, wie sie momentan im Hinblick auf die EOSC diskutiert werden.

Kooperation von FHP und IBI

Am Fachbereich Informationswissenschaften der FHP und dem IBI werden traditionell Kompetenzen im Umgang mit analogen und digitalen Informationsobjekten vermittelt. Die Spezifika von Bibliotheken und anderen Informationsinfrastruktureinrichtungen bilden den Bezugsrahmen für die Untersuchung und Gestaltung von Services für die Informationsversorgung. Zugleich werden Kenntnisse über Wissen und Information sowie Rahmenbedingungen der Informationsversorgung vermittelt, die auf andere Domänen übertragbar sind.

Das IBI und der Fachbereich Informationswissenschaften der FHP sind etablierte bibliotheks- und informationswissenschaftliche Institute im Bereich des Datenmanagements. An beiden Instituten wurde das Datenmanagement in den vergangenen Jahren in Forschung und Lehre aufgegriffen und entwickelt. Darüber hinaus haben beide Hochschulen zentrale Dienste für das Forschungsdatenmanagement aufgebaut. (Vgl. Kindling & Schirmbacher, 2013; Petras et al., 2019) Die Verbindung des IBI als Institut einer Forschungsuniversität und der FHP als anwendungsorientierte Hochschule ist für dieses Thema eine wichtige Voraussetzung. Diese Kooperation zur gemeinsamen Entwicklung und Gestaltung eines Studiengangs mit dieser inhaltlichen Ausrichtung ist in Deutschland einzigartig. (Vgl. Petras et al., 2019)

Die konkrete inhaltliche Ausarbeitung des Curriculums für DDM basiert auf Vorarbeiten wie dem Lehrkonzept für das Wahlpflichtmodul zum Thema Forschungsdaten, das am IBI durch den Lehr- und Forschungsbereich Informationsmanagement seit 2013 im Masterstudium unter dem Titel Ausgewählte Aspekte digitaler Informationsversorgung angeboten wurde. Auf Basis der Lehreinheiten dieses Mastermoduls ist der erste Entwurf des Modulhandbuchs enstanden.

Die Balance zwischen disziplinübergreifenden Kompetenzen und Querschnittsthemen und disziplinspezifischen Ansätzen zu finden, ist eine spezifische Herausforderung in der LIS, aber zugleich auch ihre Stärke. So zeigt die Erfahrung aus den bisherigen Lehrangeboten der beiden Institute beispielsweise, dass die Auseinandersetzung mit disziplinspezifischen Praktiken sehr wichtig ist, beispielsweise indem Forschende aus verschiedenen Wissenschaftsgebieten von ihrer täglichen Arbeit mit Daten berichten oder spezifische Probleme des Datenmanagements kommunizieren. Die informationswissenschaftliche "Vogelperspektive" schärft dann den Blick für Unterschiede von Forschungsgegenständen und -praktiken und ermöglicht die Verständigung über selbige. Diese Perspektive fördert die Auseinandersetzung mit generellen Themen wie organisatorischen Rahmenbedingungen und unterstützt den Ausbau von Services an multidisziplinär ausgerichteten Einrichtungen. Neben disziplinären Hintergründen können in ähnlicher Weise unterschiedliche Karrierestufen Forschender oder der Perspektivwechsel zwischen Forschenden und Infrastrukturanbieterinnen berücksichtigt werden.

Fazit

Betrachten wir Wissens- und Kompetenzgebiete der LIS als grundlegend für eine Data Literacy und insbesondere für das Datenmanagement, dann muss zugleich festgestellt werden, dass dies in den Diskursen außerhalb der LIS in Deutschland bisher wenig Beachtung findet.

Es ist deshalb wichtig, dass LIS-Einrichtungen die bisherigen nationalen und internationalen Aktivitäten und Diskurse zu "Data Literacy" oder "Data Science", die außerhalb der Disziplin geführt werden, aufgreifen, differenzierend betrachten, die Verbindung zu vorhandenen und geplanten Angeboten herstellen und sich positionieren. Hier muss es vor allem darum gehen, dass entsprechende Sichtbarkeit nicht nur für einzelne Einrichtungen wie das IBI oder die FHP hergestellt wird, sondern

dass die LIS mit ihren Forschungsthemen und Lehrangeboten in anderen Domänen sehr viel sichtbarer wird.

Die kritische Bewertung von Daten und Informationen und die der LIS inhärenten transdisziplinären Perspektiven auf Information und Wissen bieten großes Potential. Eine Verbindung von technisch-methodischen Kompetenzen im Datenmanagement mit philosophisch-ethisch geprägten Forschungsansätzen einer kritischen Datentheorie könnte ein Alleinstellungsmerkmal für LIS geprägte Studiengänge im Bereich der Data Literacy und des Datenmanagements sein. Darunter verstehen wir die kritische Auseinandersetzung mit politischen, sozialen und ethischen Implikationen von (Big) Data (vgl. Illiades und Russo, 2016) wie beispielsweise Untersuchungen zur Auswirkung von Macht oder In- bzw. Exklusion, die sich in neuen Forschungsgebieten wie Data Feminism (vgl. D'Ignazio & Klein, 2019) zeigen.

Zusammenfassend richten wir uns an unsere eigene Wissenschaftsdisziplin mit dem Wunsch, sich gerade im Bereich Data Literacy selbstbewusster aufzustellen: Die Kompetenzen dafür sind ebenso bereits vorhanden wie der Raum für die Integration angrenzender Gebiete und aktueller Diskurse.

Literatur

[ALA] American Library Association (2015) Framework for Information Literacy for Higher Education. URL: http://www.ala.org/acrl/standards/ilframework. Zuletzt abgerufen am 15.04.2019.

[Allianz] Allianz der Wissenschaftsorganisationen, Arbeitsgruppe Forschungsdaten (2018) *„Research Data Vision 2025" – ein Schritt näher: Ein Diskussionspapier der Arbeitsgruppe Forschungsdaten.* DOI: https://doi.org/10.2312/allianzoa.024

[Carlson & Johnston] Carlson, J., & Johnston, L. (Hrsg.) (2015) *Data information literacy: librarians, data, and the education of a new generation of researchers*. West Lafayette, Indiana: Purdue University Press.

[Demchenko] Demchenko, Y. (2017) *EDISON Data Science Framework, Part 4: Data Science Professional Profiles (DSPP), Release 2*. URL: http://edison-project.eu/sites/edison-project.eu/files/attached_files/node-486/edison-dspp-release2-v04.pdf. Zuletzt abgerufen am 15.04.2019.

[Demchenko et al.] Demchenko, Y., Belloum, A., & Wiktorski, T. (2017) *EDISON Data Science Framework, Part 1: Data Science Competence Framework (CF-DS), Release 2*. URL: http://edison-project.eu/sites/edison-project.eu/files/filefield_paths/edison_cf-ds-release2-v08_0.pdf. Zuletzt abgerufen am 15.04.2019.

[D'Ignazio, C. & Klein, L. (2019) Data Feminism (Draft version). URL: https://bookbook.pubpub.org/data-feminism. Zuletzt abgerufen am 15.04.2019.

[Groß] Groß, P. (2019) *Digitalisierte Arbeitswelt erfordert neues Lehren und Lernen an Hochschulen.* (Pressemitteilung). URL: https://idw-online.de/de/news711993. Zuletzt abgerufen am 15.04.2019.

[Iliades & Russo] Iliadis, A., & Russo, F. (2016) *Critical data studies: An introduction.* Big Data & Society, 3(2). DOI: https://doi.org/10.1177/2053951716674238

[Kindling & Schirmbacher] Kindling, M., & Schirmbacher, P. (2013) *„Die digitale Forschungswelt" als Gegenstand der Forschung.* Information - Wissenschaft & Praxis, 64(2–3). DOI: https://doi.org/10.1515/iwp-2013-0017

[Michel et al.] Michel, A., Baumgartner, P., Bullinger-Hoffmann, A. C., Gerdes, A., Hesse, F. W., Kuhn, S., … Spinath, B. (2018) *Framework zur Entwicklung von Curricula im Zeitalter der digitalen Transformation.* URL: https://hochschulforumdigitalisierung.de/sites/default/files/dateien/Diskussionspapier1_Framework_Curriculumentwicklung.pdf. Zuletzt abgerufen am 15.04.2019.

[Neuroth et al.] Neuroth, H., Rothfritz, L., Petras, V., & Kindling, M. (erscheint) *Digitales Datenmanagement als neue Aufgabe für wissenschaftliche Bibliotheken.* In: Bibliothek Forschung und Praxis.

[Petras et al.] Petras, V., Kindling, M., Neuroth, H., & Rothfritz, L. (2019) *Digitales Datenmanagement als Berufsfeld im Kontext der Data Literacy. ABI Technik*, *39*(1), 26–33. https://doi.org/10.1515/abitech-2019-1005

[RfII] Rat für Informationsinfrastrukturen (2016) *Leistung aus Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. URL: http://www.rfii.de/?p=1998 Zuletzt abgerufen am 15.04.2019.

[Sapp Nelson] Sapp Nelson, M. (2017) *A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs*. Journal of eScience Librarianship, 6(1). DOI: https://doi.org/10.7191/jeslib.2017.1096

[Schield] Schield, M. (2004) Information Literacy, Statistical Literacy and Data Literacy. *IASSIST Quarterly*, *28*(2/3), 6–11.

[Meyer-Guckel et al.] Volker Meyer-Guckel, Julia Klier, Julian Kirchherr, & Mathias Winde. (2019) *Future Skills*. URL: https://www.future-skills.net/. Zuletzt abgerufen am 15.04.2019.

[Whyte et al.] Whyte, A., de Vries, J., Thorat, R., Kuehn, E., Sipos, G., Cavalli, V., … Ashley, K. (2018) *Skills and Capability Framework* (Nr. Deliverable 7.3). Abgerufen von EOSCpilot website: https://eoscpilot.eu/sites/default/files/eoscpilot-d7.3.pdf

[Wilkinson et al.] Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A.,[…] Mons, B. (2016) *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data, 3. DOI: https://doi.org/10.1038/sdata.2016.18

[Wissenschaftsrat] Wissenschaftsrat (2019) *Empfehlungen zu hochschulischer Weiterbildung als Teil des lebenslangen Lernens (Drs. 7515- 19).* URL: https://www.wissenschaftsrat.de/download/2019/7515-19.pdf. Zuletzt abgerufen am 15.04.2019.

Data Librarian – ein neues Berufsbild für wissenschaftliche Bibliotheken und ein Schwerpunkt im Studiengang Data and Information Science

Simone Fühles-Ubach, Ragna Seidler-de Alwis

TH Köln

Zusammenfassung

Der nachfolgende Beitrag beschäftigt sich mit dem Entwicklungsprozess des neuen Berufsbildes “Data Librarian“, das in anglo-amerikanischen Bibliotheken bereits etabliert ist und das nun auch in und für Deutschland entwickelt wurde. In einem Prozess der Studiengangsentwicklung an der TH Köln wurde in einem mehrstufigen Verfahren mit allen Stakeholdern ein Studiengangsschwerpunkt entwickelt, der in den kommenden Jahren eine Lücke bei den hoch IT-affinen Themen in Bibliotheken schließen soll. Studieninhalte und Struktur werden umfassend und detailliert erläutert.

Abstract

The following article deals with the development process of a new direction for the profession “data librarian” in Germany, an occupation which is already widely settled in the Anglo-American library and information world. With the help of a structured process the TH Köln developed a degree programme via a multi-level procedure by involving all stakeholders. The degree programme ought to put graduates in a position to manage highly IT-related challenges in libraries. The corresponding course contents and structure of the developed degree programme will be illustrated in detail.

Studiengangentwicklung

Es ist mittlerweile zu einer Binsenweisheit geworden, dass die Bibliotheks- und Informationsbranche im Umbruch ist. Besonders sichtbar werden solche Entwicklungen allerdings, wenn sie der eigenen Institution näherkommen, wie z.B. der Entschluss, die Informationswissenschaft an der HHU in Düsseldorf nicht weiter zu verfolgen oder auch die Turbulenzen um die ZB Med, die weit über die eigene Branche hinaus Beachtung fanden. In dieser Zeit entschied sich das Kollegium des Instituts für Informationswissenschaft der TH Köln in den Prozess der Studienreform einzusteigen

und wie eine lernende Organisation den Entwicklungen des eigenen Umfeldes Rechnung zu tragen.

Darüber hinaus gab auch interne Ursachen für Veränderungsbedarf. Die Bewerberzahlen und insbesondere die Abbrecherquoten hatten sich in den vergangenen fünf Jahren negatvi entwickelt, so dass auch dieser Trend für einen inhaltlichen Reformbedarf sprach.

Auf diese Weise diente der Gesamtprozess einerseits der strategischen Positionierung des Instituts für Informationswissenschaft innerhalb der TH Köln und andererseits der Sicherung der Zukunftsfähigkeit im sich stetig ändernden Umfeld. Bisher galten Bereiche wie Informations- und Recherchekompetenz als die Schlüsselfähigkeiten der Studierenden des Instituts für Informationswissenschaft mit seinen Bachelor-Studiengängen „Bibliothekswissenschaft“, „Angewandte Informationswissenschaft“ und „Online-Redakteur“. Diese Kernkompetenz wird zukünftig in allen Studiengängen um den Bereich der Datenkompetenz (data literacy) erweitert und ergänzt. In wissenschaftlichen Bibliotheken entwickelt sich das Feld der Forschungsdaten seit mehr als 10 Jahren als eigener Zweig. (DFG, 2009)

Die Informationswissenschaft, die bisher den Aspekt „Information“ überwiegend in den Vordergrund stellte, hat mit dem Feld Daten bzw. Data Literacy einen weiteren Schwerpunkt ins Blickfeld gerückt. „Data literacy ist die Fähigkeit, planvoll mit Daten umzugehen und sie im jeweiligen Kontext bewusst einsetzen und hinterfragen zu können. Dazu gehört: Daten zu erfassen, erkunden, managen, kuratieren, analysieren, visualisieren, interpretieren, kontextualisieren, beurteilen und anzuwenden. (Stifterverband, 2018). Neben dem Content werden zukünftig also die verschiedenen Dimensionen des Forschungsfeldes Daten in den Vordergrund treten, um die Absolventinnen und Absolventen auch weiterhin mit den geforderten Kompetenzen entsprechend der sich wandelnden Anforderungen auszustatten.

Curriculumswerkstatt: Absolventinnen und Absolventen im Zentrum

Seit mehreren Jahren entwickelt die TH Köln ihre neuen Studiengänge im Prozesse einer sogenannten Curriculumwerkstatt. Dieser Prozess besteht aus vier Meilensteinen und einer Qualitätssicherungsschleife. (ZLE, 2018)

Dabei wird die Herangehensweise an die Entwicklung neuer Studiengänge von Grund auf verändert: In früheren Prozessen starteten die Überlegungen häufig bei den Forschungsschwerpunkten und Kompetenzen der Kolleginnen und Kollegen, durch die auch das Fächerspektrum, die Veranstaltungen und letztlich die Kenntnisse der Studierenden bestimmt wurden. Der neue Prozess setzt am entgegengesetzten Ende an: Hier wird im ersten Schritt das kompetenzorientierte Absolventenprofil als zentrale Basis der Studiengangentwicklung erarbeitet. Im Zentrum steht die Frage, welche Kernkompetenzen von den zukünftigen Absolventinnen und Absolventen gefordert werden. Hier: Was müssen Bibliothekarinnen und Bibliothekare in welcher Intensität können und was wird gegebenenfalls nicht mehr benötigt? Wie kann eine zukunftsorientierte, einerseits spezialisierte, andererseits jedoch auf breiten Basiskenntnissen aufsetzenden Berufsausrichtung gelingen? Erst im Anschluss an diese Überlegungen werden die didaktisch adäquaten Lehr- und Prüfungsformen und die Lehrveranstaltungen entwickelt und dem Kollegium zugeordnet.

Der Hochschulentwicklungsplan 2030 der TH Köln stellt die Kompetenzorientierung in den Fokus ihrer Bildungs- und Entwicklungsziele (TH Köln, 2019), wie dies seit mehreren Jahren auch von der Hochschulrektorenkonferenz gefordert wird (HRK, 2012). Zentraler Aspekt ist dabei die Ausrichtung der Hochschulbildung auf die sogenannte „Employability“ der Studierenden und damit auf die aktuellen und zukünftigen Entwicklungen der Arbeitswelt. Das Instrument, um eine kompetenzorientierte Studiengangsentwicklung zu erreichen, ist das sogenannte Kölner Modell der „Curriculumwerkstatt“.

Eine Besonderheit der Curriculumwerkstatt liegt darin, dass der Weg zu einem Absolventenprofil nicht eindimensional aus der internen Perspektive der Lehrenden erfolgt, sondern aus einer mehrdimensionalen Perspektive in mehreren Schritten erarbeitet wird, wie die folgende Grafik verdeutlicht.

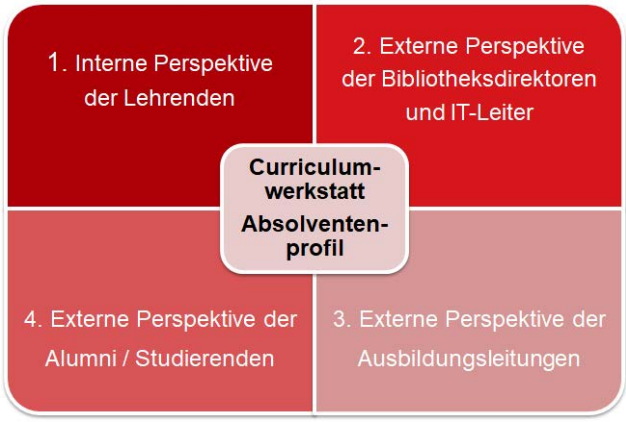


Abb. 1: Kompetenzorientierte Studiengangsentwicklung „Curriculumentwicklung“

Stufe 1: Perspektive Lehrende

In dieser Phase erfolgt eine kritische Reflektion von Inhalt und Struktur der derzeitigen Studiengänge und die Erarbeitung einer Stärken- und Schwächenanalyse (SWOT) aus der Binnensicht, um erste Ansatzpunkte für Veränderungen und Verbesserungen zu ermitteln. Darüber hinaus wurde eine Sichtung von Daten und Publikationen zur Arbeitsmarktsituation, Stellenausschreibungen und aktuellen Positionspapieren aus Verbänden und Industrie [7] vorgenommen, um die sich ändernden zukünftigen Rahmenbedingungen und Erwartungen aus dem Arbeitsmarkt gezielt zu erfassen.

Die folgenden Stufen 2-4 wurden in Form mehrerer Zukunftswerkstätten durchgeführt, die zeitlich nacheinander stattfanden und deren Ergebnisse für die nächste Zukunftswerkstatt wiederum die Grundlage bildeten. Der Abstand lag bei 2-3 Wochen und die Dauer einer Zukunftswerkstatt war mit 4-6 Stunden festgelegt. Der Ablauf wird kurz skizziert: In der Startphase besteht u.a. die Möglichkeit Kritik an bestehenden Inhalten und Strukturen zu formulieren, was in der Regel den Bedarf für einen Änderungsprozess noch einmal klar vor Augen führt. Das anschließend zu formulierende Idealbild soll die bestmögliche vorstellbare Situation beschreiben. Dahinter steckt die Absicht, sich nicht frühzeitig in seinen eigenen Vorstellungen von bestehenden Rahmenbedingungen limitieren zu lassen. Die Orientierung an der Realität findet in der letzten Phase statt, in der geprüft wird, wie viel von der Idealvorstellung sich auch in der Praxis umsetzen lassen würde. Die folgende Grafik skizziert die Vorgehensweise:

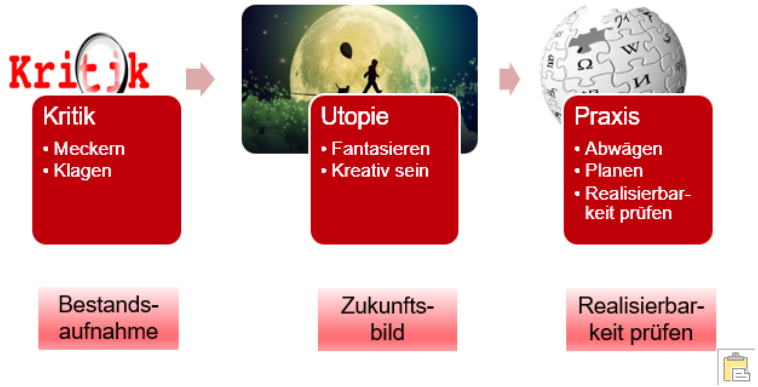


Abb. 2: Zukunftswerkstatt

Stufe 2: Externe Perspektive der Bibliotheksleitungen und IT-Verantwortlichen
Als nächster Schritt wurde eine Zukunftswerkstatt mit ca. 20 Vertretern aus Öffentlichen (ÖB) und Wissenschaftlichen Bibliotheken (WB) unterschiedlicher Größenordnung, Fachstellen und Spezialbibliotheken veranstaltet. Die bisherigen Studieninhalte wurden vorstellt und auf ihre Zukunftsfähigkeit hin analysiert. Gleichzeitig wurden neue Bedarfe artikuliert und gänzliche Streichungen vorgeschlagen und abschließend themenspezifisch gesammelt und geclustert.

Bereits nach dieser Zukunftswerkstatt zeichneten sich – ebenso wie bei der vorherigen Diskussion im Kollegium (Stufe 1) folgende Punkte ab:

- Eine **Schwerpunktsetzung nach Bibliothekssparten (WB, ÖB)** aufgrund der sich immer stärker vollziehenden Ausdifferenzierung von Wissenschaftlichen und Öffentlichen Bibliotheken und den sich daraus ergebenden verschiedenen Kompetenzprofile.
- Der **Bedarf für die Entwicklung eines neuen Profils für einen Data Librarian**, d.h. einen hoch IT-affinen Spezialisten im Bereich von Repositorien, Forschungsdaten, eScience-Prozessen und anderen neuen, datenbasierten Aufgabenfeldern in Wissenschaftlichen Bibliotheken.

Stufe 3: Externe Perspektive der Ausbildungsleiter
Im Treffen mit den Ausbildungsleitern von Bibliotheken (Schwerpunkt NRW) stand die Zuordnung von Kompetenzen und Kompetenzniveaus (Basis / Fortgeschritten) zu den bisher festgehaltenen Spezialisierungen ÖB, WB und DaLi im Vordergrund. Diskutiert wurden die Unterschiede zwischen Öffentlichen und Wissenschaftlichen Bibliotheken

z.B. in den Bereichen der Formalerschließung (u.a. RDA), Informationskompetenz oder Bibliotheksmanagement.

Stufe 4: Externe Perspektive der Alumni und Studierenden

Die Zukunftswerkstatt mit den Studierenden bezog sich in der Hauptsache auf das Erleben und die Wahrnehmung des aktuellen Studiengangs, seiner Organisation (u.a. Praxisphase) und der inneren Struktur der Module. Gleichzeitig wurden die Vorschläge der Leitungen sowie der der Ausbildungsleiter thematisiert und mit den Vorstellungen der Studierenden abgeglichen.

Data Librarian als neues Berufsbild

Die Zusammenfassung der internen und externen Ergebnisse ergab, dass insbesondere in wissenschaftlichen Bibliotheken zukünftig ein großer Bedarf an IT- bzw. Daten-Spezialisten entstehen wird, deren IT-Kenntnisse weit über Zusatzkenntnisse in Form von einzelnen Modulen hinausreichen. Der Studienschwerpunkt „Data Librarian“ wurde damit als Spezialisierungsoption des Studiengangs „Data and Information Science“ konzipiert. Die zweite Schwerpunktoption lautet „Data Analyst“.

Abb. 3: Data Librarian als eigener Schwerpunkt eines Studiengangs

Auf diese Weise entstanden die beiden neuen Bachelor-Studiengänge

- BA “Data and Information Science“ mit zwei verschiedenen Schwerpunkten „Data Librarian“ oder „Data Analyst“ sowie
- BA „Bibliothek und digitale Kommunikation“ mit Schwerpunkt „Öffentliche Bibliothek (ÖB)“ oder „Wissenschaftliche Bibliothek (WB)“

Der Studiengang Data and Information Science ist ein praxisorientierter, 7-semestriger BA Studiengang mit 27 Modulen und insgesamt 210 ECTS. Laborpraktika und eigene Forschungsprojekte sowie das Praxissemester im 4. Semester zeigen deutlich den starken Praxisbezug auf. Der Studiengang ist im Kern ein informationswissenschaftlicher Studiengang und daher nimmt die Ordnung und Priorisierung von Informationen einen besonderen Stellenwert im Curriculum ein. Die Absolventinnen und Absolventen werden befähigt relevante Zusammenhänge und Schlussfolgerungen aus gewonnenen Informationen herauszuarbeiten und zu analysieren. Sie sind in der Lage, informationswissenschaftliche Erkenntnisse unter Berücksichtigung wirtschaftlicher und gesellschaftlicher Erfordernisse in die Verwertung von Daten und Informationen zu übertragen. Daher müssen die Studierenden über ein breites Spektrum an Techniken und Methoden zur Entwicklung und Analyse von Informationsprodukten verfügen. Das Ziel des Studiengangs ist es daher, die Absolventinnen und Absolventen in die Lage zu versetzen heterogene Informationsquellen im Kontext ihrer Arbeit zu bewerten , zu nutzen und zu analysieren. Gleichzeitig sollen sie aber auch digitale Informationssysteme gestalten und integrieren können, denn Wissenschaft ist heute in weiten Teilen datenintensiv und damit gleichzeitig technikintensiv und - zumindest der Möglichkeit nach - kollaborativer als vor der Einbindung digitaler Werkzeuge und Prozesse.

In den ersten drei Semestern werden die gleichen Lehrveranstaltungen von beiden Schwerpunktbereichen (Data Librarian und Data Analyst) besucht. Die Grundlagen der Informationswissenschaft werden durch Module in den Bereichen Wissensorganisation, inhaltliche Erschließung und Information Retrieval vermittelt. Eine weitere Grundlage sind die technischen Fertigkeiten, wie z.B. Programmierung (Python etc.) und Software-Entwicklung / Software-Administration und Web-Technologien als auch die Vermittlung statistischer Verfahren und die Visualisierung von Informationen. Weitere technischen Inhalte sind das Entwickeln und der Einsatz von Dokumenten- und Content-Management-Systemen, Retrieval Systemen und von semantischen Technologien für wissensbasierte Systeme. Vorbedingung für die korrekte Extraktion von Daten und Analyse von Informationen sind gute Kenntnisse in der Bewertung von Informationsquellen und Recherchemethoden. Mit diesem Methodenmix sind Absolventinnen und Absolventen des Data and Information Science Studiengangs besonders für die Anforderungen der modernen daten- und

informationsgetriebenen Arbeitswelt gerüstet. Sie bilden mit ihrem Profil die Schnittstelle zwischen IT-Entwicklung und dem Informationsbedarf in Forschungseinrichtungen und wissenschaftlichen Bibliotheken als auch Unternehmen.

Data and Information Science – Studienverlaufsplan – Data Librarian
Modulraster: 5 Module à 4 SWS/6 ECTS/180h je Semester

1. Semester 20 SWS 30 ECTS 900h Workload	**Programmierung** • **Webentwicklung**	**Informationserschließung** • **Wissensorganisation**	**Digitale Informationsgesellschaft** • **Informationsethik** • **Berufsfelderkundung (Profil2)**	**Informationsvisualisierung**	**Information in Unternehmen**
2. Semester 20 SWS 30 ECTS 900h Workload	**Programmierung** • **Softwareentwicklung**	**Informationserschließung** • **Strukturierte Dokument-beschreibung**	**Datenmodellierung**	**Statistische Datenanalyse**	**Informationsquellen** • **Informationsrecherche**
3. Semester 20 SWS 30 ECTS 900h Workload	**Informationssysteme** • **Content- & Dokumenten-managementsysteme**	**Information Retrieval**	**Datenbanksysteme**	**Data Mining**	**Informationsanalyse**
4. Semester 30 ECTS 900h Workload	**Praxismodul** Planung & Organisation Projektmanagement	**Praxismodul** *Projektarbeit* *Praxisphasenbericht* *Präsentation*	Praxisphase	Praxisphase	Praxisphase
5. Semester 20 SWS 24 ECTS 900h Workload	**Suchmaschinentechnologie** • **Webtechnologien**	**Projektarbeit I** • Interdisziplinäres Projekt (Profil2)	**Informationsrecht & Datenschutzrecht** • Wissenschaftliches Arbeiten	**Allg. Social Credits** • **Studienportfolio**	**Information Consultancy, Wissenschaftskommunikation & Wissenschaftssoziologie und -politik**
6. Semester 20 SWS 30 ECTS 900h Workload	**Projektarbeit II**		**Forschungsdaten I** • Lizenzmanagement, Digitales Publizieren & Open Access	**Informetrie, Bibliometrie, Szientometrie** • **Empirische Forschungsmethoden**	**Automatische Erschließung**
7. Semester 10 SWS 36 ECTS 900h Workload		**Forschungsdaten II** • Digitalisierung & Langzeitarchivierung	**Seminare** • Aktuelle Themen • Seminar Bachelorarbeit	**Bachelorarbeit**	**Bachelorarbeit**

Abb. 4: Modulmatrix und Studienverlauf Data Librarian

Data Librarian - Studieninhalte und Qualifikationen

Im Studiengang Data and Information Science ist es der Studienschwerpunkt Data Librarian, der ein neues Berufsbild in Bibliotheken adressieren soll und im Studium einen Anteil von 60 ECTS umfasst. Der Data Librarian wird für die Bereiche Bildung, Wissenschaft und Forschung ausgebildet und umfasst damit auch die Wissenschaftlichen Bibliotheken und Spezialbibliotheken. Das Praxissemester wird im Bereich einer wissenschaftlichen Bibliothek (Spezialbibliothek) oder einer Informations- oder Forschungseinrichtung absolviert und führt die Studierenden in die Bibliothekswelt ein. Im Im 5. bis 7. Semester werden durch fachliche Module, die gemeinsam mit den Studierenden des Studienschwerpunktes „Wissenschaftliche Bibliothek“ des Studiengangs „Bibliothek und digitale Kommunikation“ absolviert werden, spezifische Themen des Berufsfeldes [8] vertieft:

- Information Consultancy; Wissenschaftskommunikation und Wissenschafts-soziologie und -politik (5. Semester)

- Forschungsdaten I, Lizenzmanagement, Digitales Publizieren & Open Access / eRessourcen (6. Semester)
- Bibliometrie / Szientometrie (6. Semester)
- Automatische Erschließung (6. Semester)
- Forschungsdaten II, Digitalisierung und Langzeitarchivierung / Repositorien (7. Semester)

Forschungsdaten und Forschungsdatenmanagement bilden somit einen zentralen Schwerpunkt. Zwei große, jeweils zweisemestrige Projektarbeiten sollen darüber hinaus eine weitere Intensivierung schwerpunktspezifischer Themen mit deutlichem Praxisbezug ermöglichen. [9]

Exponierte Rolle der Forschungsdaten

Forschungsdaten sind immer in Bezug zur jeweiligen Wissenschaftsdisziplin zu sehen und berücksichtigen mehrere Perspektiven:

a) Die generische Perspektive zur nachhaltigen Strukturierung, Pflege, Verfügbarhaltung und Langzeitarchivierung (DFG-Empfehlung 10 Jahre) von Metadaten ist ein originäres Aufgabengebiet der Bibliotheks- und Informationswissenschaft.
b) Die fachwissenschaftliche Perspektive thematisiert die disziplinspezifischen Anforderungen und Besonderheiten an die Aufbereitung und Interpretation von Forschungsdaten sowie der für ihr Verständnis bzw. ihre Nachnutzung unabdingbaren Metadaten.
c) Die Perspektive der wissenschaftlichen Praxis berücksichtigt neue Formen der wissenschaftlichen Zusammenarbeit über zeitliche und räumliche Grenzen hinweg (z.B. mit virtuellen Methoden arbeitende Forschergruppen, Online-Review-Verfahren, Virtuelle Labore/Classrooms), gemeinsames Erfassen und Verwerten von Forschungsdaten sowie den Aufbau von Fachcommunitys im Netz.
d) Die ethische Perspektive befasst sich unter anderem mit Fragen der Privatsphäre, Auswirkungen der Rekontextualisierung und Verknüpfung sowie Analyse großer Datenmengen (Big Data) sowie der Einhaltung von Grundlagen und Regeln guter wissenschaftlicher Praxis.

Daraus ergibt sich eine interdisziplinäre Aufgabenstellung, die nicht allein im Institut für Informationswissenschaft geleistet werden kann. Die Zusammenführung von generischer und disziplinspezifischer Perspektive ist an der TH Köln in der hochschulübergreifenden interdisziplinären Projektwoche (HIP) geplant und

infrastrukturell vorgesehen. In interdisziplinärer Zusammenarbeit können alle Akteurinnen und Akteure aus Ingenieur- und Naturwissenschaften sowie Geistes- und Sozialwissenschaften eine gemeinsame Projektwoche gestalten, deren Ziel es ist, die Bedeutung und Funktion interdisziplinärer Prozesse wie das Forschungsdatenmanagement in den Vordergrund zu rücken und ein Bewusstsein dafür zu entwickeln. Das Thema Forschungsdaten bietet sich hier in idealer Weise an.

Strategische Personalentwicklung im Kollegium

Der inhaltliche Wandel im Bereich der Bibliotheks- und Informationswissenschaft erfordert auch eine langfristige Umorientierung im Bereich der Professuren, um den neuen Anwendungsgebieten und Forschungsfeldern des Instituts für Informationswissenschaft gerecht zu werden.

Mit Unterstützung des Präsidiums der TH Köln kann für 5 Jahre eine Innovationsprofessur besetzt werden, die in zwei 50%-Professuren geteilt wurde.

- Professur eScience und Forschungsdatenmanagement (Besetzung zum 01.09.2019)
- Professur Open Access und Management digitaler Ressourcen (Verfahren läuft noch)

Diese beiden Professuren ergänzen und vervollständigen den Teil des Kollegiums, der sich mit überwiegend bibliotheksnahen Themen auseinandersetzt. Noch im laufenden Jahr 2019 werden darüber hinaus aufgrund von (Vor-)Ruhestand zwei weitere Stellen in den Bereichen „IT in Bibliotheken“ und „Informationsdienstleistungen“ ausgeschrieben werden, deren zukünftige fachliche Ausrichtung in der Studiengangentwicklung bereits angedacht wurde. Gleichzeitig braucht diese Neuorientierung, z.B. die stärkere Datenorientierung auch klare Signale nach außen, um zu verdeutlichen, dass neue Gebiete mit spezialisierten Kolleginnen und Kollegen besetzt werden. Das Institut hat seine Personalentwicklung daher bereits in den vergangenen Jahren mit Blick auf diesen Wandel ausgerichtet. Freiwerdende Professuren wurden in den vergangenen Jahren u.a. wie folgt besetzt:

- Professur für Information Retrieval – besetzt mit einem Informatiker mit Schwerpunkt Data Harvesting und Analyse

- Prof. für Web Recherche – besetzt mit einem Physiker mit Schwerpunkt Text und Data Mining
- Prof. für Informationskompetenz (aber auch Data Science) – besetzt mit einem Bioinformatiker mit Schwerpunkt Data Carpentry
- Prof. für Medien- und Webwissenschaft - besetzt mit einer Spezialistin für Social Media

Diese Kollegen ergänzen insbesondere im Datenschwerpunkt die drei Kolleginnen und Kollegen, die als Mathematiker im bisherigen Kollegium überwiegend mit IT-Themen beschäftigt waren.

Interessensbekundungen in großem Umfang

Ein Blick in verschiedene Jobbörsen zeigt, dass einerseits die Berufsbezeichnungen für den Data Librarian recht vielfältig sind - sie reichen vom Data Librarian bis zum IT-Systembibliothekar - und dass andererseits das Berufsbild stark nachgefragt wird. In der ersten Woche der Berufsfelderkundung, einer Projektwoche, die 8 Wochen nach Studienbeginn stattfindet, konnten mehrere Bibliotheken bzw. Forschungseinrichtungen (u.a. ZBMed, GESIS, Bibliothek des FZ Jülich) mit geographischer Nähe zur TH Köln von den interessierten Studierenden besucht werden. Viele Einrichtungen, wie z.B. die ZBW in Hamburg, die USB Köln, die ULB Bonn, die DNB Frankfurt und die Universitätsbibliothek in Marburg zeigen bereits proaktives Interesse an Studierenden dieses Studiengangs und bieten Studierenden an, in ihrer Einrichtung das Praxissemester absolvieren zu können. Auch an der TH Köln gab es im Vorfeld in der Studienberatung ein ausgeprägtes Interesse von Seiten der Studieninteressierten.

Ausblick

Derzeit befinden sich die ersten Studierenden des BA Data and Information Science im zweiten Semester und diese werden ab März 2020 Zugang zu bibliothekarischen Einrichtungen über das verpflichtende Praxissemester suchen. Um diese neue Kooperation zum Erfolg zu führen ist eine besonders enge Zusammenarbeit zwischen Hochschule und Praxissemester notwendig, die über spezifische Vorstellungsrunden und gemeinsame Projektbetreuung sichergestellt werden wird. Noch vor Beginn der ersten Praxisphase werden entsprechende Kriterien und Qualitätssicherungsmaßnahmen und mit der Berufspraxis abgeglichen, um dieses

neue Berufsbild bereits im ersten Einstieg in der Bibliothek bzw. der Forschungseinrichtung inhaltlich richtig zu positionieren und so erfolgreich einzusetzen.

Literatur

Alle Online-Quellen zuletzt recherchiert am 26.4.2019

[1] Deutsche Forschungsgemeinschaft, Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme Unterausschuss für Informationsmanagement : Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten. Januar 2009. <http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf>

[2] Stifterverband der Deutschen Wissenschaft: Data Literacy Education: Datenkompetenzen für die Studierenden aller Fächer. <https://www.stifterverband.org/data-literacy-education>

[3]Der Studiengang Online-Redaktion wird weiterhin im Institut für Informationswissenschaft angeboten, wird jedoch wegen der deutlich geringeren inhaltlichen Überschneidungen hier nur am Rand erwähnt.

[4] ZLE / Referat 4: Organisationsstruktur und Prozess einer Curriculumwerkstatt, 2018. <https://www.th-koeln.de/mam/downloads/deutsch/hochschule/profil/lehre/curriculumswerkstatt.pdf>

[5] Hochschulentwicklungsplan 2030 der TH Köln, 2019, S 11.<https://www.th-koeln.de/mam/downloads/deutsch/hochschule/profil/hochschulentwicklungsplan2030.pdf>

[6] Vgl. HRK-Fachgutachten zur Kompetenzorientierung in der Lehre, Hochschulrektorenkonferenz - Projekt Nexus (Schaper, 2012). Die Hochschulrektorenkonferenz (HRK) befasst sich seit 2010 gemeinsam mit dem Bundesministerium für Bildung und Forschung (BMBF) im Projekt nexus mit der Weiterentwicklung der Studienprogramme und der Sicherung der Studienqualität.

[7] Hier bspw. DBV-Berichte zur Lage der Bibliotheken der letzten Jahre (http://www.bibliotheksverband.de/dbv/publikationen/bericht-zur-lage-der-bibliotheken.html); Sichtung der Stellenbörsen Bibliojobs (www.bib-info.de/verband/berufsfeld-information-bibliothek/bibliojobs.html) und OpenBibliojobs (https://jobs.openbiblio.eu/) sowie von aktuellen Positionspapieren, bspw. IFLA Trendreport (https://trends.ifla.org) und Positionspapier des ekz-Fachbeirats „Berufsbild und Entwicklung“.

[8] Die detaillierten Inhalte der Module können im Modulhandbuch unter https://www.th-koeln.de/mam/downloads/deutsch/studium/studiengaenge/f03/dis-modulbuch_standapril2019.pdf eingesehen werden.

[9] Anmerkung: Das Zentrum für bibliotheks- und informationswissenschaftlichen Weiterbildung (ZB IW) der TH Köln bietet für bereits Berufstätige und andere Interessierte Personenkreise einen mehrwöchigen Zertifikatskurs “Data Librarian” ab Herbst 2019 an.

Datenmanagementpläne

Research Data Management Organiser

Ulrike Wuttke[1], Heike Neuroth[1], Jochen Klar[2], Harry Enke[3], Janine Straka[1], Kerstin Wedlich-Zachodin[5], Claudia Kramer[5], Olaf Michaelis[3], Jens Ludwig[4]

[1] Fachhochschule Potsdam (FHP)
[2] Berater und Softwareentwickler
[3] Leibniz-Institut für Astrophysik Potsdam (AIP)
[4] Stiftung Preußischer Kulturbesitz
[5] Karlsruher Institut für Technologie (KIT)

Zusammenfassung

Das Erstellen eines Datenmanagementplans (DMP) und die Planung des Datenmanagements erfordern ein frühzeitiges Nachdenken darüber, wie mit Daten während der Forschungsaktivitäten und nach Projektende umgegangen werden soll. Es stellen sich Fragen zu Nutzungsrechten, zur Datenpublikation und zu Archivierungsmöglichkeiten. Der Research Data Management Organiser (RDMO) ist eine webbasierte Open Source-Software, die Forschungsprojekte bei der Planung, Umsetzung und Verwaltung aller Aufgaben des Forschungsdatenmanagements über den gesamten Datenlebenszyklus unterstützt. Zusätzlich ermöglicht RDMO die textuelle Ausgabe eines DMP nach den Vorgaben unterschiedlicher Förderer und kann als Teil des Infrastrukturangebots einer Einrichtung an deren Bedarfe (z. B. eigene Frage- und Antwortvorlagen) und deren eigene Corporate Identity angepasst werden.

In diesem Beitrag wird das DFG-Projekt Research Data Management Organiser (RDMO) vorgestellt, das sich momentan in der zweiten Projektphase befindet. In dieser wird durch die Projektpartner des Leibniz-Instituts für Astrophysik Potsdam, der Fachhochschule Potsdam und der Bibliothek des Karlsruher Instituts für Technologie die Software weiterentwickelt und verbessert. Außerdem werden in enger Zusammenarbeit mit den Nutzenden neue Funktionalitäten hinzugefügt, die Community bei der Integration von RDMO in lokale Infrastrukturen unterstützt, Einführungs- und Schulungsmaterialien erstellt und ein Nachhaltigkeitskonzept für RDMO erarbeitet.

Abstract

Data management planning and creating a Data Management Plan (DMP) requires early consideration of how to handle the data during the project and after the end of the project. Questions arise concerning usage rights, data publication, and long-term

storage. The Research Data Management Organiser (RDMO) is a web-based open source software that supports research projects in planning, implementing, and managing all research data management tasks over the entire data lifecycle. In addition, RDMO enables the textual output of a DMP according to the specifications of different funders. It can be adapted as an infrastructural service of an institution according to its requirements (e.g. individual question and answer templates) and its own corporate identity.

The article presents the DFG funded project Research Data Management Organiser (RDMO), which is in its second project phase. Currently, the software is being further developed and improved by the project partners Leibniz Institute for Astrophysics Potsdam, University of Applied Sciences Potsdam and the library of the Karlsruhe Institute of Technology. Furthermore, new functionalities will be added in close cooperation with the users, the community will be supported in integrating RDMO into local infrastructures, introductory and training materials will be created and a sustainability concept for RDMO is being developed.

1. Einleitung

1.1 Aktuelle Entwicklungen

Im Jahr 2018 hat das Bundesministerium für Forschung und Entwicklung (BMBF) nach einer Re-Strukturierung des Ministeriums ein eigenes Referat 421 für Forschungsdaten ins Leben gerufen[1]. Parallel dazu wurde in Deutschland das Förderprogramm Nationale Forschungsdateninfrastruktur (NFDI) gestartet. Bereits 2016 legte dafür der Rat für Informationsinfrastrukturen (RfII) mit der Veröffentlichung des Berichts „Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland“ (RfII 2016) den Grundstein für die Strategie eines nationalen Ansatzes des Forschungsdatenmanagements. Darauffolgende Veröffentlichungen des RfII gaben weitere Impulse für die Ausgestaltung der NFDI.[2] Neben verschiedenen Fachverbänden nahmen aus Bibliotheks- und Infrastrukturperspektive verschiedene Verbände und Institutionen das Thema mit Nachdruck auf und legten weitere, die

[1] https://www.bmbf.de/pub/orgplan.pdf. Alle Links wurden am 15.04.2019 geprüft.
[2] Vgl. die weiteren Diskussionsimpulse des RfII, u. a. RfII 2017a, RfII 2017b, RfII 2018a und RfII 2018b.

Wichtigkeit des Managements von Forschungsdaten im Rahmen der nationalen Strategieentwicklung unterstreichende, Impuls- und Strategiepapiere vor.[3]
Im internationalen Kontext sei in diesem Zusammenhang auf die Tätigkeiten der Initiative „Research Data Alliance - Research Data Sharing without barriers“[4] verwiesen, die ein breites Spektrum im Rahmen von Arbeitsgruppen abdeckt. Auf europäischer Ebene wurden im Rahmen der „European Open Science Cloud“ (EOSC)[5] im November 2018 zwei wichtige Berichte veröffentlicht: 1.) „Prompting an EOSC in practice“ (Muscella et al. 2018), der sich hauptsächlich mit Regularien für verschiedene Beteiligungsszenarien an der EOSC beschäftigt, und 2.) „Turning FAIR into reality” (European Commission Expert Group on FAIR Data 2018). Der zuletzt genannte Bericht beschreibt „the broad range of changes required to turn „FAIR data into reality““ (ebd., S. 13). Alle Initiativen, Entwicklungen und Diskussionen sowie konkreten Förderprogramme basieren auf den FAIR-Prinzipien („Findable – Accessible – Interoperable – Reusable“)[6], die in Zukunft einen menschen- und vor allem maschinenlesbaren Zugang zu Forschungsdaten und deren Nachnutzung realisieren sollen.

Die im November 2018 veröffentlichte „Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)“ (GWK 2018) weist in ihrer Präambel auf die Notwendigkeit von potentiellen NFDI-Konsortien als „wissenschaftlich breit nutzbare Datenschätze mit gesellschaftlichem Mehrwert“ hin und macht in ihren Förderkriterien darauf aufmerksam, dass „ein stimmiges Konzept zu Datennutzung und -zugang sowie Auffindbarkeit und Nachnutzbarkeit der Daten, welches entlang der FAIR-Prinzipien ausgerichtet ist“ (ebd., S. 4), vorliegen muss. Dafür stehen im Zeitraum von 2019 bis 2028 bis zu 90 Mio. Euro pro Jahr im Endausbau für die Projektförderung der NFDI zur Verfügung. Flankierend eröffnete der Stifterverband ein eigenes Förderprogramm für neue Lehr- und Lernkonzepte im Bereich „Data Literacy“.[7]

Zusammenfassend kann festgestellt werden, dass Forschungsdaten eines der wichtigen Themen der Zukunft darstellen und der Umgang mit und das Management von Forschungsdaten weiter in den Mittelpunkt der Aufmerksamkeit rücken. Dazu

[3] Vgl. u. a. DFG 2018a, Kapitel B-2d, DBV 2018, DFG 2018b.
[4] https://www.rd-alliance.org/.
[5] https://eoscpilot.eu.
[6] Vgl. z. B. https://www.go-fair.org/fair-principles/.
[7] https://www.stifterverband.org/data-literacy-education.

gehört auch die Entwicklung und Bereitstellung geeigneter Werkzeuge und Dienste, die alle beteiligten Akteure aus der Wissenschaft, Infrastruktur-Dienstleister etc. darin unterstützen, den Umgang mit Forschungsdaten gemäß den FAIR-Prinzipien zu ermöglichen.

1.2 Hintergrund des DFG-Projekts

Seit 2015 fördert die DFG das Projekt Research Data Management Organiser: Zunächst in einer ersten Förderphase für 18 Monate unter dem Namen „Entwicklung und Implementierung eines Werkzeugs für die Planung, Umsetzung und Kontrolle des Forschungsdatenmanagements (FDMP-Werkzeug)“ und seit November 2018 in einer zweiten Förderphase für 30 Monate unter dem jetzigen Titel und Akronym „Research Data Management Organiser (RDMO)“. Seit der ersten Phase waren die beiden Einrichtungen Leibniz-Institut für Astrophysik Potsdam (AIP) und der Fachbereich Informationswissenschaften an der Fachhochschule Potsdam (FHP) für die Durchführung verantwortlich, ab der zweiten Phase hat sich die Bibliothek des Karlsruher Instituts für Technologie (KIT) angeschlossen. In beratender Funktion war stets Jens Ludwig, (ehemals Staatsbibliothek zu Berlin, jetzt: Strategien & Technologien Koordination Digitale Transformation, Stiftung Preußischer Kulturbesitz) als Mitinitiator des Projekts beteiligt.

Nach dem Ende der ersten Förderphase stand ein Werkzeug für die strukturierte Planung, Umsetzung und Verwaltung des Forschungsdatenmanagements zur Verfügung. Da sich schon nach relativ kurzer Zeit zeigte, dass Forschungsdatenmanagement ein aktiver, dynamischer Prozess ist, der mit der Planung von Forschungsvorhaben beginnen sollte und kontinuierlich angepasst und dokumentiert werden muss, wurde das Werkzeug von FDMP zu RDMO umbenannt, was die Schwerpunktverschiebung von der Planung zur Organisation des Forschungsdatenmanagements verdeutlichte. Die Umbenennung sollte auch unterstreichen, dass viele Akteure am Forschungsdatenmanagement beteiligt sind und das Werkzeug auch für die mittel- bis langfristige strategische Planung (z. B. im Technologiebereich, für die Finanzkalkulation einer gesamten Einrichtung etc.) verwendet werden kann.

Ein weiteres wesentliches Merkmal von RDMO ist, dass die Software von jeder Einrichtung eigenständig implementiert und so dem eigenen „Look & Feel“ angepasst

werden kann. RDMO versucht so weit wie möglich, gängige Schnittstellen zur Verfügung zu stellen, so dass das Werkzeug mit bereits vorhandenen Diensten interoperabel verknüpft werden kann. Weiterhin verbleiben alle Daten in der Hoheit der Einrichtung. Die RDMO-Community kann sich in kollaborativ orchestrierten Prozessen an der Weiterentwicklung und Pflege der Software und den Fragenkatalogen etc. beteiligen.

Schwerpunkte der zweiten Förderphase sind einerseits die technologische Stabilisierung des Werkzeugs inklusive Bereitstellung weiterer Schnittstellen für die einfache Integration in lokale Technologieumgebungen, andererseits die stabile Verankerung in der Community. Je mehr Einrichtungen und Forschergruppen das Werkzeug einsetzen und sich aktiv an der Weiterentwicklung beteiligen, desto nachhaltiger kann der Betrieb gesichert werden. Im Folgenden werden die Aktivitäten rund um RDMO näher beschrieben.

2. Research Data Management Organiser (RDMO)

2.1 Ziele

In Deutschland werden Datenmanagementpläne (DMP) in der Regel momentan nur einmal, als Teil des Antragstexts, erstellt und im Verlauf des Forschungsvorhabens nicht weiter aktualisiert. Im Mittelpunkt des RDMO-Projekts steht die Weiterentwicklung des Organisers zu einem multilingualen Research Data Management Organiser (RDMO), der es ermöglicht, das Forschungsdatenmanagement während des gesamten Projekts zu organisieren und entsprechend den jeweiligen Bedarfen anzupassen. Unter Einbeziehung des Nutzerfeedbacks und entsprechender Bedarfe der Zielgruppe und aufbauend auf den Ergebnissen des Vorgängerprojekts werden im aktuellen Projekt vier Teilziele verfolgt.

Erweiterung: RDMO ist so aufgebaut, dass alle relevanten Daten zum Projektstart erfasst und daraus ein Datenmanagementplan erstellt werden kann, mit dessen Hilfe das Datenmanagement während des gesamten Projektverlaufs organisiert werden kann (aktives Datenmanagement). Diese mit RDMO erstellten aktiven DMPs reflektieren immer den aktuellen Stand aller für das Datenmanagement wichtigen Aspekte und enthalten Angaben zu konkreten Aufgaben und Rollen (Stichwort *actionable data management plan*) (vgl. Miksa et al. 2019). Dazu kommt die besondere Unterstützung zentraler Aufgaben, wie der Kostenabschätzung, des Ingest-Prozesses

und der Interoperabilität mit z. B. Nachweissystemen wie re3data[8] durch die Entwicklung entsprechender Module und eines Metadatenmodells. Die Implementierung von Schnittstellen und Identifier-Systemen zur Verlinkung zwischen RDMO und den tatsächlichen Daten und zur Übernahme von Metadaten aus Forschungsinformationssystemen (FIS) wird außerdem beabsichtigt.

Integration in die Infrastruktur: Ein weiteres Ziel des Projekts ist die Vereinfachung der Integration von RDMO in die jeweiligen Infrastrukturen von Infrastrukturanbietern durch eine standardisierte Installation (z. B. über Docker-Container), die Entwicklung eines integrierten Update-Mechanismus zur besseren Unterstützung der Wartung, sowie die Erweiterung der Unterstützung von Authentifizierungs- und Autorisierungsverfahren. Die vereinfachte Integration ist eine wichtige Voraussetzung für die Etablierung von RDMO als verlässliches Service-Angebot von Infrastrukturanbietern.

Etablierung in der Community: Das dritte Ziel von RDMO ist die Etablierung in verschiedenen Fachdisziplinen (wie Biodiversität, Psychologie etc.), universitären und außeruniversitären Wissenschaftseinrichtungen (wie Universitäten und Leibniz-Instituten) sowie Wissenschaftsorganisationen (wie der Fraunhofer-Gesellschaft), als zentrales Tool für aktives Datenmanagement. Außerdem steht in diesem Bereich die internationale Anschlussfähigkeit im Mittelpunkt (mit z. B. DMPonline[9] des Digital Curation Centers in Großbritannien oder dem DMPTool[10] der University of California in den USA) sowie die Förderung der konkreten Nutzung des Tools vor Ort in den verschiedenen Communities. Hierfür sind sowohl die zentrale Bereitstellung von Schulungsmaterialien und Online-Tutorials als auch dezentrale Ansätze sinnvoll.

Nachhaltigkeit / Verstetigung: Zentrale Aufgabe in diesem Bereich ist die Erhaltung von RDMO als eigenständiges Tool, damit sich die in den Communities eingesetzten RDMO-Instanzen in der Zukunft austauschen und spezifische Entwicklungen integriert werden können und RDMO insgesamt beständig weiterentwickelt wird. Hierfür arbeitet das Projekt mit verschiedenen Initiativen wie E-Science Baden-Württemberg[11] sowie (außer-)universitären Forschungseinrichtungen zusammen.

[8] https://www.re3data.org/.
[9] https://dmponline.dcc.ac.uk/.
[10] https://dmptool.org/.
[11] https://mwk.baden-wuerttemberg.de/de/forschung/forschungslandschaft/e-science/.

2.2 Community

Für eine Software wie RDMO sind Community Building und Bestrebungen zur Sicherung der Nachhaltigkeit eng miteinander verzahnt. Die Nutzung von RDMO als Werkzeug zum Datenmanagement in den Communities, d. h. den verschiedenen Fachdisziplinen und Institutionen, kann die Nachhaltigkeit des Tools sicherstellen. Ziel ist es letztendlich, dass sich die verschiedenen RDMO-Instanzen selbst tragen und auf diese Weise RDMO insgesamt weiterentwickelt wird. Daher ist eine der Hauptaufgaben des RDMO-Projekts die Etablierung einer aktiven Community und der Aufbau einer dezentralen Struktur.

Für das Community-Building bedient sich das Projekt verschiedener Instrumente. Neben klassischer Dissemination und Training spielen insbesondere der Aufbau von Kooperationsstrukturen und die Förderung von Aktivitäten in der Community eine wichtige Rolle. Zentrale Pfeiler sind hierbei technische Dokumentationslösungen und die Erleichterung der Integration von RDMO in die eigene Infrastruktur, sowie die Etablierung verschiedener Foren zum Austausch. Neben den Projekt-eigenen digitalen Kommunikationskanälen (wie z. B. GitHub, Slack, allgemeine Mailingliste), haben sich vor allem die bisher an verschiedenen Orten durchgeführten RDMO-Anwendertreffen bewährt.[12] In diesen konnten sich (potentielle) Anwender*innen zu Fragen der Installation und Anwendung mit dem RDMO-Team und untereinander austauschen. Außerdem konnten weitere Bedarfe gesammelt werden, die nach Prüfung durch das Projektteam ggf. in die Verfeinerung bzw. Anpassung der Projektaufgaben einfließen.

Zur Halbzeit des Projekts sind durch diese unterschiedlichen Aktivitäten verschiedene Kooperationsszenarien etabliert bzw. befinden sich in der Entwicklung. Der Einsatz von RDMO in verschiedenen Zusammenhängen, wie z. B. in Projekten, in Universitäten bzw. Universitätskooperationen, in DFG-Sonderforschungsbereichen, an Leibniz-Instituten, oder regionalen Verbünden zeigt, dass RDMO bereits eine gute Aufnahme gefunden und sich eine Anwender-Community gebildet hat.[13]

[12] Die lokalen Workshops fanden sowohl bei den unmittelbaren RDMO-Projektpartnern als auch bei RDMO-Kooperationspartnern statt. Siehe die Übersicht auf der RDMO-Webseite: https://rdmorganiser.github.io/workshops/.

[13] Siehe die interaktive Visualisierung der Projektkooperationen in Form einer Karte auf der RDMO-Projektwebseite: https://rdmorganiser.github.io/kooperationen/.

Darüber hinaus ist der Einsatz von RDMO im Projektantrag „Domain-Data-Protokolle für die empirische Bildungsforschung – DDP-Bildung“ zum BMBF-Call „Kuration und Qualitätsmerkmale von Forschungsdaten“ als Bestandteil des Vorhabens vorgesehen. RDMO wurde als Pilot-Service in die Informationsplattform forschungsdaten.info[14] eingebettet, das Hosting und technische Betreuung erfolgen beim RDMO-Projektpartner KIT (Karlsruher Institute of Technologie). Weitere Kooperationen sind geplant, wie z. B. die Kooperation zwischen RDMO und dem RADAR-Projekt[15] für die Nutzung von RDMO als ein Interface für das RADAR-Angebot[16] und die Integration von RDMO in die erweiterte Cloud-basierte Forschungsdatenmanagement-Umgebung Sciebo[17].

Alle RDMO-Instanzen sind auf der RDMO-Webseite als interaktive Karte visualisiert, wobei zwischen Produktiv- und Testinstanzen unterschieden wird.[18] Während Testinstanzen nur für eine kleine Zahl Testnutzer*innen zugänglich sind, um die Funktionalitäten auszuprobieren, sind Produktivinstanzen allen Anwender*innen der jeweiligen anbietenden Institution zugänglich und können aktiv für das Datenmanagement eingesetzt werden. Es folgt eine Kategorisierung des Einsatzes von RDMO anhand konkreter Beispiele (siehe Tab. 1-3). Die Anordnung erfolgt dabei auf Basis des Umfangs der jeweiligen direkten bzw. potentiellen Anwender-Community, von Projektzusammenhängen zu regionalen Verbünden.

Einsatz von RDMO im Rahmen von Forschungsprojekten	
Name	**Standort(e)**
EmiMin - Verbundvorhaben Emissionsminderung Nutztierhaltung (BMEL)[19]	RDMO-Instanz ZB Med/Publisso (Köln)
Domain Data Protokolle in der empirischen Bildungsforschung[20]	Kooperationsprojekt, Leitung GESIS (Köln), Beteiligung AIP (Potsdam)

[14] https://rdmo.forschungsdaten.info/.
[15] https://www.radar-service.eu/de.
[16] Diese geplante Kooperation war Thema eines Lightening Talks während der Heidelberger E-Science Tage 2019: https://www.e-science-tage.de/de/programm.
[17] https://www.sciebo.de/.
[18] https://rdmorganiser.github.io/kooperationen/.
[19] https://www.ktbl.de/themen/emimin/.
[20] Noch keine Projekt-Webseite.

FoDaKo - Forschungsdaten in Kooperation (BMBF)[21]	Universitäten Düsseldorf-Siegen-Wuppertal
MaMoMaR - Management Molekularer Daten im Research Data Life Cycle (DFG)[22]	Fachhochschule Potsdam und Robert Koch-Institut, Berlin
RADAR[23]	FIZ Karlsruhe

Tab. 1: RDMO-Kooperationsszenarien: Projekte

RDMO als Service einer Infrastruktureinrichtung (z. B. Bibliothek oder Rechenzentrum) an einer Hochschule bzw. außeruniversitären Einrichtung	
Name	**Standort(e)**
Alfred-Wegner-Institut (AWI), Helmholtz-Zentrum für Polar- und Meeresforschung[24]	Bremen
Deutsches Zentrum für Luft- und Raumfahrt (DLR)[25]	Weßling
Fachhochschule Potsdam (FHP)[26]	Potsdam
Forschungszentrum Jülich[27]	Jülich
Friedrich-Alexander Universität Erlangen-Nürnberg[28]	Erlangen, Nürnberg
Göttingen eResearch Alliance[29] der Georg-August-Universität Göttingen	Göttingen
Helmholtz-Zentrum Berlin für Materialien und Energie (HZB)[30]	Berlin

21 https://www.fodako.de/.
22 Noch keine Projekt-Webseite.
23 https://www.radar-service.eu/de.
24 https://www.awi.de/.
25 https://www.dlr.de/.
26 https://www.fh-potsdam.de/.
27 https://www.dlr.de/.
28 https://fau.de.
29 http://www.eresearch.uni-goettingen.de.
30 https://www.helmholtz-berlin.de/.

Karlsruher Institut für Technologie (KIT), Helmholtz-Gemeinschaft	Karlsruhe
Leibniz-Institut für Astrophysik Potsdam (AIP)[31]	Potsdam
Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK)[32]	Gatersleben
Max Planck Digital Library[33]	München
Physikalisch-Technische Bundesanstalt (PTB)[34]	Braunschweig
Ruhr Universität Bochum[35]	Bochum
RWTH Aachen[36]	Aachen
Stiftung Preußischer Kulturbesitz[37]	Berlin
Technische Universität Braunschweig[38]	Braunschweig
Technische Universität Darmstadt[39]	Darmstadt
Technische Universität Hamburg[40]	Hamburg
Universität Bayreuth[41]	Bayreuth
Universität Darmstadt[42]	Darmstadt
Universität Duisburg-Essen[43]	Duisburg, Essen
Universität Hamburg[44]	Hamburg

[31] https://www.aip.de/de.
[32] https://www.ipk-gatersleben.de/.
[33] https://www.mpdl.mpg.de/.
[34] https://www.ptb.de/.
[35] https://www.ruhr-uni-bochum.de/de.
[36] https://www.rwth-aachen.de/.
[37] https://www.preussischer-kulturbesitz.de/.
[38] https://www.tu-braunschweig.de/.
[39] https://www.tu-darmstadt.de/.
[40] https://www.tuhh.de/tuhh/startseite.html.
[41] https://www.uni-bayreuth.de/de/index.html.
[42] https://www.tu-darmstadt.de/.
[43] https://www.uni-due.de.
[44] https://www.uni-hamburg.de/.

Universität Hildesheim[45]	Hildesheim
Universität Stuttgart[46]	Stuttgart
Universität Trier[47]	Trier
Westfälische Wilhelms-Universität Münster[48]	Münster

Tab. 2: RDMO-Kooperationsszenarien: Infrastruktureinrichtungen

RDMO im Rahmen regionaler Verbünde	
Name	**Standort(e)**
bwFDM-Info (Koordiniertes Forschungsdatenmanagement in Baden-Württemberg)[49]	Baden-Württemberg, Karlsruher Institut für Technologie, Universität Konstanz
HeFDI (Hessische Forschungsdateninfrastrukturen)[50]	Hessen, Projektleitung Philipps-Universität Marburg
Landesinitiative NFDI der Digitalen Hochschule NRW[51]	Nordrhein-Westfalen, Projektleitung Universität Duisburg-Essen

Tab. 3: RDMO-Kooperationsszenarien: Regionale Verbünde

2.3 Software

RDMO ist eine interaktive Webanwendung mit der die Nutzer*innen primär über einen Webbrowser interagieren. Als Basis für den serverseitigen Teil der Software dient das in der Programmiersprache Python geschriebene Django-Framework[52]. Über die Python-eigene Paketverwaltung pip wird auf eine Reihe weiterer, etablierter Programmbibliotheken aus der Python- und Django-Community zurückgegriffen.

[45] https://www.uni-hildesheim.de/.
[46] https://www.uni-stuttgart.de/.
[47] https://www.uni-trier.de/.
[48] https://www.uni-muenster.de/.
[49] http://bwfdm.scc.kit.edu/. Einbindung von RDMO in das zentrale Angebot Forschungsdaten.info unter https://rdmo.forschungsdaten.info.
[50] https://www.uni-marburg.de/de/forschung/kontakt/forschungsdatenmanagement/projekte/hefdi-hessische-forschungsdateninfrastrukturen.
[51] https://fdm-nrw.de/.
[52] https://www.djangoproject.com/.

Als interaktive Webanwendung verfügt RDMO auch über ausgeprägte clientseitige Funktionalitäten, die in der Programmiersprache JavaScript und mit Hilfe des AngularJS 1[53]-Frameworks realisiert sind. Client und Server stehen dabei über REST-Schnittstellen in Verbindung. Für die Unterstützung verschiedener Ausgabeformate wird pandoc[54] verwendet. Bei der Auswahl dieser Komponenten wurde insbesondere auf einen ausreichenden Reifegrad der jeweiligen Projekte und der Verbreitung nicht nur in der akademischen Webentwicklungs-Community geachtet.

RDMO wurde vom ersten Tag an als Open Source-Software entwickelt und unter der Apache2-Lizenz[55] bereitgestellt. Als zentrale Plattform für die Entwicklung, aber auch die Interaktion mit der Community in technischen Fragen, wird die Internetplattform GitHub[56] verwendet. GitHub dient hierbei nicht nur als Repositorium, sondern ist auch durch Integrationen mit Werkzeugen der Continuous Integration (Travis CI, Coveralls) und mit Zenodo zur Vergabe von DOIs für Releases verbunden.

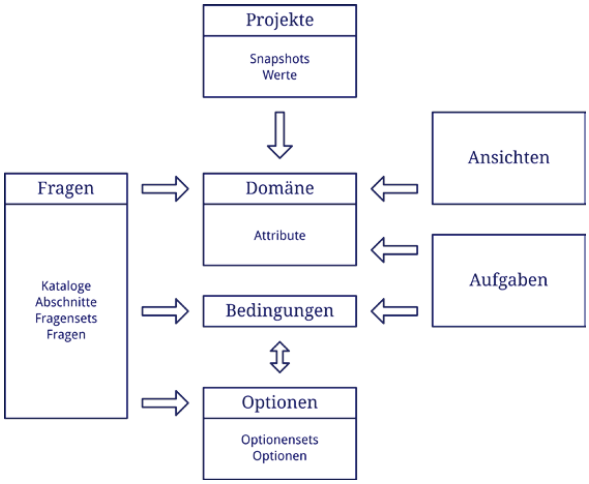


Abb. 1: Überblick über das RDMO-Datenmodell und die verschiedenen Module und ihre Abhängigkeiten.

Die Applikation organisiert sich anhand einer Reihe von die verschiedenen Teilfunktionalitäten von RDMO abbildenden Modulen (siehe Abb. 1). Primärer

[53] https://angularjs.org/.
[54] https://pandoc.org/.
[55] https://www.apache.org/licenses/LICENSE-2.0.
[56] RDMO auf GitHub: https://github.com/rdmorganiser.

Interaktionspunkt mit den Nutzer*innen ist das Fragen-Modul in dem verschiedene Fragenkataloge konfiguriert werden können, die sich wiederum in Abschnitte, Fragensets und schließlich in einzelne Fragen gliedern. Die Fragen werden durch die Nutzer*innen in einem interaktiven, strukturierten Interview beantwortet. Das Optionen-Modul stellt kontrollierte Vokabulare als Antwortoptionen zur Verfügung. Die Bedingungen legen fest, welche Fragen auf Grund schon gegebener Antworten übersprungen werden können. Nach der Beantwortung der Fragen können die getätigten Eingaben einerseits in identischer Form oder aber, durch das Ansichten-Modul, anhand von hinterlegten Vorlagen, in der von Förderern verlangten textuellen Form neu zusammengesetzt ausgegeben werden. Der Download unterstützt hierbei gängige Formate, wie Microsoft Word, Open Office oder LaTeX. Zusätzlich können auf den Nutzereingaben basierende Aufgaben konfiguriert werden und steht eine API für den programmatischen Zugriff, beispielsweise durch andere Tools, zur Verfügung.

Eingabe- und Ausgabeteil sind durch das RDMO-Domänenmodell verbunden. Dieses besteht aus einem internen, baumförmigen Vokabular sogenannter Attribute, die Fragen und Antworten miteinander verbinden. Die gleichen Attribute werden auch für die Ansichten und Aufgaben verwendet, so dass sich eine Entkopplung zwischen Eingabe und Ausgabe ergibt. Dadurch, dass die Nutzereingaben mit Bezug auf die Attribute gespeichert werden, können Projekte auch zwischen verschiedenen RDMO-Instanzen über ein spezielles XML-Format exportiert bzw. importiert werden (so lange das gleiche Domänenmodell verwendet wird). Eine ähnliche Funktionalität erlaubt auch eine Weitergabe von Fragenkatalogen, Ansichten und anderen Inhalten zwischen den Betreibenden von RDMO-Instanzen (siehe auch unten).

Besonderes Augenmerk gilt der Anpassbarkeit von RDMO. Die beschriebenen Inhalte (Fragenkatalog, Ansichten, Domäne etc.) können alle über ein Webinterface bearbeitet werden. Hierfür sind kein Dateizugriff auf den Server oder Programmierkenntnisse nötig. Auch die Software ist so aufgebaut, dass sich leicht Anpassungen am Aussehen der Seite vornehmen lassen (Corporate Identity), die eigene RDMO-Instanz jedoch weiter kompatibel zur zentral gepflegten RDMO-Bibliothek bleibt. Die Installation selber setzt nur einen Linux-Server und normale Administrationskenntnisse voraus. RDMO besitzt eine eigene Nutzerverwaltung, integriert aber auch OAUTH2 Provider (z. B. ORCID) und LDAP-System oder lässt sich innerhalb einer Shibboleth-Föderation betreiben.

2.4 RDMO Metadatenmodell

Wie bereits oben beschrieben, erfolgt die Eingabe in RDMO in Form eines strukturierten Interviews (Fragenkatalog) über ein interaktives Webinterface. Die eingegebenen Informationen können auf verschiedene Art und Weise aggregiert, kombiniert und ausgegeben werden: Ausgabe der Fragen und Antworten als Text; Ausgabe über spezielle DMP-Vorlagen, wie z. B. Horizon 2020; Ausgabe bestimmter Aufgaben um zukünftige Handlungsbedarfe anzuzeigen, wie z. B. Termine zur Datenveröffentlichung; Informationen zur Kostenabschätzung für das Datenmanagement; sowie Nachnutzung durch andere Software-Werkzeuge über programmierbare Schnittstellen (APIs) (vgl. auch Neuroth et al. 2018, S. 62). Diese verschiedenen Optionen unterstreichen die Funktionalität von RDMO, nicht – wie viele andere Tools für das Forschungsdatenmanagement – die Generierung eines textuellen DMP in den Vordergrund zu rücken. RDMO unterstützt das Forschungsdatenmanagement über den ganzen Projektzeitraum unter Einbeziehung aller daran beteiligten Akteure (wie z. B. Wissenschaftler*innen, Datenmanager*innen, IT-Expert*innen, Institutsleiter*innen etc.).

Die Grundlage für die komplexen Abfrage- und Ausgabefunktionalitäten bildet das interne Datenmodell von RDMO (siehe Abb.1, vgl. Neuroth et al. 2018, S. 62).[57] Die Fragen (und die dazu eingegebenen Antworten) des Fragenkatalogs sowie die verschiedenen Ansichten und Aufgaben referenzieren auf ein baumförmiges Domänenmodell. Das jeweils aktuelle Domänenmodell steht auf GitHub im XML-Format zum Download zur Verfügung.[58] Innerhalb des RDMO-Domänenmodells werden die Fragen und Antworten auf sogenannte Attribute referenziert, wobei die Attribute einzelne Informationseinheiten (Frage- und Antwortoptionen) bezeichnen. Die einzelnen Elemente des Domänenmodells sind mit eindeutigen Bezeichnern versehen, um die semantische Interoperabilität der einzelnen RDMO-Instanzen untereinander, aber auch mit Datenmodellen wie DataCite[59], oder perspektivisch CERIF[60] (Common European Research Information Format) oder dem Kerndatensatz

[57] Vgl. auch Heger 2016. Diese Arbeit ist jedoch aufgrund der Überarbeitung des RDMO-Domänenmodells im Jahr 2018 in einigen Teilen überholt.
[58] https://github.com/rdmorganiser/rdmo-catalog/blob/master/rdmorganiser/domain/rdmo.xml.
[59] https://datacite.org/.
[60] https://www.eurocris.org/cerif/main-features-cerif.

Forschung[61] zu ermöglichen. Die semantische Interoperabilität der verschiedenen RDMO-Instanzen untereinander erlaubt den Datenexport für die Übertragung der in einer Instanz vorhandenen Daten auf eine andere Instanz, z. B. im Falle eines Umzugs an eine neue Institution. Darüber hinaus bildet die Interoperabilität mit anderen Datenmodellen, wie zum Beispiel DataCite, die Voraussetzung für die breitere internationale Anschlussfähigkeit von RDMO und die Standardisierung des Datenmanagements (Stichwort *Machine Actionable Datamanagement Plans*).
Im Vorläuferprojekt wurden bereits im Rahmen der Anforderungsanalyse etwa 70 User Stories bezüglich inhaltlicher Aspekte und der Usability von RDMO aus der Sicht verschiedener Stakeholder/Akteure definiert (ein Ansatz der agilen Software-Entwicklung) (vgl. Neuroth et al., S. 58-59). Einige dieser User Stories betreffen Szenarien in denen z. B. Datenmanager*innen oder Institutsleiter*innen Statistiken über verschiedene Einrichtungen oder eine spezifische Einrichtung erstellen oder Infrastrukturanbieter bestimmte Antworten von DMPs abfragen wollen, um Planungen, z. B. zur rechtzeitigen Bereitstellung von Speicherplatz, vornehmen zu können (vgl. ebd., S. 59). Die Anforderungen an RDMO zur Erfüllung dieser beispielhaften Szenarien betreffen die Fragenstruktur und -logik bzw. die Antwortmöglichkeiten – so erhöht der weitgehende Verzicht auf Freitextfelder die Vergleichbarkeit der Antworten über verschiedene Projekte hinweg – und ist eng mit dem RDMO-Domänenmodell zusammenhängenden Bereichen wie Exportmöglichkeiten und Schnittstellen verbunden (vgl. ebd., S. 59).

Die Bereitstellung des RDMO-Quellcodes als Open Source-Software erlaubt es interessierten Institutionen RDMO ausgiebig zu testen und dem Projekt agil auf Feedback aus der Community zu reagieren. So war Feedback aus der Community der Anlass zur Vereinfachung des RDMO-Domänenmodells in der ersten Phase der aktuellen Projektlaufzeit (2018). Im Zuge dieser Überarbeitung wurden die Import- und Exportfunktionen erweitert und einige grundsätzliche Änderungen vorgenommen, mit dem Ziel einer höheren Austauschfähigkeit. Zusätzlich reflektieren diese Änderungen aus der Teilnahme an der RDA-Interest Group Active Data Management Plans[62] und

[61] https://kerndatensatz-forschung.de/. Der Kerndatensatz Forschung enthält momentan nur rudimentäre Ansätze zu mit Forschungsdaten verbundenen Fragen.
[62] https://rd-alliance.org/groups/active-data-management-plans.html.

Feedback im Rahmen der Vorstellung des RDMO-Datenmodells beim RDA-Treffen 2018 in Berlin[63] gewonnene Erkenntnisse.

In der zweiten Phase des Projekts steht die weitere Vereinfachung des RDMO-Domänenmodells mit Hinsicht auf ein zukünftiges definierte RDMO-Metadatenmodell im Mittelpunkt. Hierbei wird besondere Aufmerksamkeit der Unterstützung der Aufgaben Kostenabschätzung und Rolleneinteilung, sowie des Ingest-Prozesses und der Interoperabilität mit Nachweissystemen, insbesondere DataCite, d. h. dem DataCite-Metadata Schema[64], sowie der Einbeziehung der Ergebnisse der RDA-Interest Group Active Metadata Plans[65], gewidmet. Weiterhin ist es vorgesehen, das RDMO-Datenmodell (sowie ggf. das zukünftige RDMO-Metadatenmodell), ganz im Sinne der FAIR-Prinzipien, offen für die Community bereit zu stellen.

Referenzen

DBV - Deutscher Bibliotheksverband (2018): Wissenschaftliche Bibliotheken 2025. Verfügbar unter: https://www.bibliotheksverband.de/fileadmin/user_upload/Sektionen/sektion4/Publikationen/2018_02_27_WB2025_Endfassung_endg.pdf.

DFG - Deutsche Forschungsgemeinschaft (2018a): Stärkung des Systems wissenschaftlicher Bibliotheken in Deutschland. Ein Impulspapier des Ausschusses für Wissenschaftliche Bibliotheken und Informationssysteme der Deutschen Forschungsgemeinschaft. Verfügbar unter: http://www.dfg.de/download/pdf/foerderung/programme/lis/180522_awbi_impulspapier.pdf.

DFG - Deutsche Forschungsgemeinschaft (2018b): Förderung von Informationsinfrastrukturen für die Wissenschaft. Ein Positionspapier der Deutschen Forschungsgemeinschaft. Verfügbar unter: http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_informationsinfrastrukturen.pdf.

European Commission Expert Group on FAIR Data (2018): Turning FAIR into Reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data. Verfügbar unter: https://doi.org/10.2777/1524.

GWK - Gemeinsame Wissenschaftskonferenz (2018): Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI). Verfügbar unter: https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf.

Heger, Martin (2016): Datenmodellie-rung für Forschungsdatenmanagementpläne. Masterarbeit an der Fachhochschule Potsdam im Fachbereich 5 Informationswis-senschaften. Verfügbar unter: http://rdmorganiser.github.io/docs/Heger_MA.pdf.

Miksa, Tomasz, Simms, Stephanie, Mietchen, Daniel, Jones, Sarah (2019): Ten principles for machine-actionable data management plans. PLoS Comput Biol 15(3): e1006750. Verfügbar unter: https://doi.org/10.1371/journal.pcbi.1006750.

[63] https://www.rd-alliance.org/dmp-common-standards-wg-rda-11th-plenary-meeting.
[64] Momentan Version 4.2. vom 20.03.2019, vgl. https://schema.datacite.org/, Zugriff 25.03.2019.
[65] https://rd-alliance.org/groups/active-data-management-plans.html. Die Ergebnisse der Interest Group sollen während des RDA-Treffens in Philadelphia (02.-04.04.2019) vorgestellt werden.

Muscella, Silvana, Campos Plasencia, Isabel, Komatsoulis, George A., Mortensen, Andreas, Räim, Toivo, Robida, François, Strick, Linda, Tochtermann, Klaus, Turk, Žiga, Wilkinson, Ross (2018): Prompting an EOSC in practice: final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC). Luxembourg: European Commission. Verfügbar unter: https://doi.org/10.2777/112658.

Neuroth, Heike, Engelhardt, Claudia, Klar, Jochen, Ludwig, Hens, Enke, Harry (2018). Aktives Forschungsdatenmanagement. In: ABI Technik 38:1, S. 55-64. Verfügbar unter: http://doi.org/10.1515/abitech-2018-0008.

RfII - Rat für Informationsinfrastrukturen (2016): Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Verfügbar unter: http://www.rfii.de/?p=1998.

RfII - Rat für Informationsinfrastrukturen (2017a): Schritt für Schritt – oder: Was bringt wer mit? Ein Diskussionsimpuls zu Zielstellung und Voraussetzungen für den Einstieg in die Nationale Forschungsdateninfrastruktur (NFDI). Diskussionspapier. Verfügbar unter: http://www.rfii.de/?wpdmdl=2269.

RfII - Rat für Informationsinfrastrukturen (2017b): Entwicklung von Forschungsdateninfrastrukturen im internationalen Vergleich. Bericht und Anregungen. Verfügbar unter: http://www.rfii.de/?wpdmdl=2346.

RfII - Rat für Informationsinfrastrukturen (2018a): Zusammenarbeit als Chance. Zweiter Diskussionsimpuls zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) für die Wissenschaft in Deutschland. Verfügbar unter: http://www.rfii.de/?wpdmdl=2529.

RfII - Rat für Informationsinfrastrukturen (2018b): Zusammenarbeit als Chance. Zweiter Diskussionsimpuls zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) für die Wissenschaft in Deutschland. Verfügbar unter: http://www.rfii.de/?wpdmdl=2529.

Datenmanagementpläne an der RWTH Aachen University – Wie groß ist das Beratungsspektrum?

Ute Trautwein-Bruns, Daniela Hausen, Stephan von der Ropp

Universitätsbibliothek, RWTH Aachen University

Zusammenfassung

Ein Datenmanagementplan (DMP) ist ein Instrument zur Dokumentation des Forschungsdatenmanagements (FDM) und der Forschungsdaten eines Forschungsprojekts. Er wird von einigen Forschungsförderern und Institutionen zur Sicherung der Qualität der Forschung und der Nachhaltigkeit von Forschungsdaten gefordert und hat darüber hinaus das Potenzial, ein gewinnbringendes Konzept für alle Beteiligten zu werden. Um Einstiegshürden zu senken und das Konzept DMP mit Mehrwert für die Forschung handhabbar zu gestalten, bietet die RWTH Aachen University neben Informationen, Beratung und Training auch Workshops zur instituts- oder fachspezifischen Anpassung und Umsetzung von DMPs an.

Das Erstellen sinnvoller DMPs ist sowohl für die Forschenden als auch für die Unterstützung auf Seiten der Infrastruktur-Dienstleister aufwendig, nicht zuletzt, weil Community-Lösungen für viele Fachbereiche gleichermaßen wie Umsetzungsbeispiele zur Integration von DMPs in die Forschungsprozesse fehlen. Deshalb sollten zum jetzigen Zeitpunkt DMPs so kompakt und niederschwellig wie möglich gehalten werden, damit der Aufwand bei der Erstellung gering ist und das Konzept von den Forschenden akzeptiert wird. Zukünftig ist zu erwarten, dass mit der Implementierung der Nationalen Forschungsdateninfrastruktur (NFDI), Fachcommunities den Umgang mit Forschungsdaten reflektieren und angemessene Konzepte zur disziplinspezifischen Nutzung von Forschungsdaten und zur Umsetzung der FAIR Data Principles erarbeiten.

Abstract

A data management plan (DMP) is an instrument for documenting the research data management (FDM) and the research data within a research project. It is required by some research funders and institutions to ensure the quality of research and the sustainability of research data. Furthermore, it has the potential to become a useful concept for all stakeholders. In order to introduce the DMP concept and make it practicable for researchers, the RWTH Aachen University offers information,

consultancy and training as well as workshops for the institute- or subject-specific adaptation and implementation of DMPs.

The creation of meaningful DMPs is very time-consuming both for the researchers and for the support by infrastructure service providers, because community solutions are missing for many disciplines as well as examples for the integration of DMPs into research processes. Therefore, at present time DMPs should be kept as short and straightforward as possible, in order to minimise the effort and to increase researchers' acceptance of the concept. In future, it is expected that with the development of the National Research Data Infrastructure (NFDI), research communities will reflect on the handling of research data and define appropriate concepts for the discipline-specific use of research data and for the implementation of the FAIR Data Principles.

1. Bedeutung von DMPs

Ein DMP ist ein Instrument zur strukturierten Erfassung aller relevanten Informationen über die in einem Forschungsprojekt verwendeten Daten (Arbeitsgruppe Forschungsdaten 2018[i]). Er beinhaltet eine Beschreibung der Forschungsdaten, des Daten- und Metadatenmanagements sowie die Pläne für Aufbewahrung und gemeinsame Nutzung, basierend auf Ressourcen, Verantwortlichkeiten und Einschränkungen.

In anderen Ländern ist die Erstellung von DMPs schon länger explizit in den Ausschreibungen von verschiedenen Forschungsförderern wie beispielsweise der National Science Foundation (NSF)[1] oder des Wellcome Trust[2] verankert. Seit 2014 gewinnen DMPs auch in Deutschland an Bedeutung, da die Europäische Kommission diese im Förderprogramm Horizon 2020 im Rahmen des Open Research Data Pilot (ORD Pilot) explizit als Deliverables fordert.[3] Ziel ist es, eine qualitativ hochwertige Forschung entsprechend der im Fachbereich üblichen Standards sicherzustellen und Forschungsdaten langfristig so aufzubewahren, dass sie entsprechend den FAIR-Prinzipien[4] auffindbar, zugänglich, interoperabel und nachnutzbar - idealerweise auch

[1] https://www.nsf.gov/eng/general/dmp.jsp, Zugriff 15.04.2019
[2] https://wellcome.ac.uk/funding/guidance/developing-outputs-management-plan, Zugriff 15.04.2019
[3] http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm, Zugriff 15.04.2019
[4] https://www.force11.org/group/fairgroup/fairprinciples, Zugriff 15.04.2019

durch Maschinen - sind. Die offene Zugänglichkeit (open access), die für Textpublikationen verpflichtend ist, wird auf Forschungsdaten ausgeweitet, soweit dies rechtlich und technisch möglich sowie fachlich sinnvoll ist[5].

Die großen nationalen Forschungsförderer in Deutschland verlangen zwar nicht explizit einen DMP, wohl aber bei Antragstellung konkrete Angaben zum Umgang mit Forschungsdaten. „Die Deutsche Forschungsgemeinschaft verfügt seit 2015 über eine eigene Forschungsdaten-Richtlinie und sieht in ihren Antragsleitfäden jeweils ein Kapitel zum Umgang mit Forschungsdaten vor. Das Bundesministerium für Bildung und Forschung fordert in vielen Förderausschreibungen ebenfalls solche Angaben, wobei allerdings die Vorgaben je nach betroffener Fachdisziplin unterschiedlich streng ausfallen“ (Soßna & Wespel 2019[ii]).

Die RWTH Aachen University hat in ihrer Leitlinie zum FDM[6] DMPs für alle Forschungsprojekte mit Forschungsdaten verankert und unterstützt die Forschenden bei der Erstellung von DMPs durch Informationen, Beratung und Trainings sowie durch das Angebot einer eigenen, auf den naturwissenschaftlich-technischen Bereich fokussierten DMP-Vorlage[7] sowie den Research Data Management Organizer (RDMO)[8] als DMP-Tool. Dabei wird das Ziel verfolgt, dass der DMP mehr als ein formales Dokument sein soll, sondern ein gewinnbringendes Werkzeug für die Forschenden, indem er bei der Strukturierung des FDMs hilft, zu besserem FDM anleitet und die Qualität der Forschungsdaten im Projekt sicherstellt. Darüber hinaus sollen die Forschenden für das Thema Nachnutzung von Daten sensibilisiert werden, um das Potenzial der Veröffentlichung von Forschungsdaten in ihrer Fachdisziplin bewerten zu können. Um diese Sicht zu verdeutlichen und den Forschenden nahezubringen, verfolgt die RWTH das Konzept, DMP-Vorlagen so nah wie möglich an der jeweiligen Wissenschaft sowie an den jeweils genutzten Infrastrukturangeboten zu orientieren.

DMPs werden oft als zu bürokratische, statische Dokumente wahrgenommen, die nicht in den Forschungsprozess integrierbar sind. Eine Änderung dieses Zustands wird auch auf internationaler Ebene, beispielweise von der Research Data Alliance (RDA) in der

[5] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf, Zugriff 15.04.2019
[6] www.rwth-aachen.de/global/show_document.asp?id=aaaaaaaaaaqwpfe&download=1, Zugriff 15.04.2019
[7] http://www.rwth-aachen.de/global/show_document.asp?id=aaaaaaaaaasvnen, Zugriff 15.04.2019
[8] https://rdmo.itc.rwth-aachen.de/, Zugriff 15.04.2019

Interest Group Active Data Management Plans[9] sowie den Working Groups DMP Common Standards[10] und Exposing Data Management Plans[11], angestrebt. „Active DMPs“ sollen den gesamten Forschungsprozesses mit Mehrwerten für alle Beteiligten unterstützen (Miksa et al. 2018[iii]) . Das Konzept verfolgt die Umstellung von rein händisch ausgefüllten DMPs hin zu (größtenteils) maschinell erstellten. Auf dem RDA 13th Plenary Meeting wurden das DMP Common Model[12] vorgestellt und weitere Implementierungen diskutiert[13].

Auch das RDMO-Projekt[14] verfolgt mit dem Motto „vom Tool zum Organiser“ das Ziel, ein DMP-Werkzeug zur Integration in den Forschungsprozess zu schaffen. Dazu bietet bzw. entwickelt RDMO neben einem strukturierten Fragenkatalog mit Hilfetexten, weiterführenden Links und teils vorgegebenen Antwortmöglichkeiten, auch Schnittstellen zu anderen Systemen und eine Aufgabenfunktion, um andere im FDM involvierter Akteure einzubinden (Neuroth et al. 2018[iv]). Darüber hinaus besteht die Möglichkeit, fach- und institutsspezifische DMP-Vorlagen zur Verfügung zu stellen sowie das Tool an das Corporate Design der Hochschule anzupassen.

In diesem Beitrag präsentieren wir unsere Ansätze, die Erstellung von DMPs zu unterstützen, und diskutieren das Beratungsspektrum, das aus unserer Sicht zentrale Einrichtungen zur Unterstützung der Forschenden leisten können.

2. Unterstützung von DMPs an der RWTH Aachen University

2.1 Informations-, Beratungs- und Schulungsangebot

FDM an der RWTH Aachen University ist ein zentrales, strategisches Thema und wird durch den Aufbau von Infrastruktur- und Serviceangeboten gestützt. DMPs sind im wissenschaftlichen Alltag noch nicht integriert. Um für das Thema zu sensibilisieren und den Einstieg zu erleichtern, bietet das FDM-Team, das aus Mitarbeitenden des IT Centers und der Universitätsbibliothek besteht, Informationen, Schulungen sowie fachspezifische und konzeptionelle Beratungen an.

In der Weiterbildungsstruktur der RWTH für Promovierende und akademischen Mittelbau werden regelmäßige Veranstaltungen zum FDM angeboten, die u.a. auch

[9] https://rd-alliance.org/groups/active-data-management-plans.html, Zugriff 15.04.2019
[10] https://rd-alliance.org/groups/dmp-common-standards-wg, Zugriff 15.04.2019
[11] https://rd-alliance.org/groups/exposing-data-management-plans-wg, Zugriff 15.04.2019
[12] https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard, Zugriff 15.04.2019
[13] https://www.rd-alliance.org/wg-dmp-common-standards-rda-13th-plenary-meeting, Zugriff 15.04.2019
[14] https://rdmorganiser.github.io/projekt/, Zugriff 15.04.2019

das Thema DMP adressieren. Das 2-stündige Überblickseminar „Management von Forschungsdaten (FDM I)“ erklärt die Grundbegriffe und führt das Konzept DMP ein. Im darauf aufbauenden ganztägigen Workshop zum „Management von Forschungsdaten (FDM II)“ werden in praktischen Übungen hands-on Erfahrungen zu den Konzepten und Werkzeugen des FDM vermittelt. Zudem gibt es ein Überblickseminar für Mitarbeitende in Technik und Verwaltung, da insbesondere Administratorinnen und Administratoren sowie Laborpersonal die Forschenden in der Umsetzung des FDMs unterstützen können. Alle Veranstaltungen werden auf Anfrage auch für individuelle Gruppen (Institute, Forscherkollegs, Projekte) angeboten, wobei Themenschwerpunkte, wie das Erstellen von DMPs, zielgruppengerecht abgestimmt und kombiniert werden können. Aktuell wird ein Blended-Learning-Angebot finalisiert, welches neben dem Präsenzworkshop (FDM II) eine online-Vorbereitung zur selbständigen und individuellen Erarbeitung der grundlegenden Themen als Alternative zum Überblickseminar (FDM I) enthält. In diesem Zusammenhang wurden auch zwei Lehrvideos zum Thema DMP produziert. Im Video „Datenmanagement nach Plan“ (Schmitz et al. 2018[v]) wird erklärt, was ein DMP ist und welche Vorteile er für die Forschung bietet. Im darauf aufbauenden zweiten Video „Inhalte eines Datenmanagementplans“ (Schmitz et al. 2018[vi]) wird auf die konkreten Inhalte eines DMPs eingegangen.

Das Thema FDM sollte im Idealfall schon bei der Beantragung von Projekten berücksichtigt werden, um Aufwendungen für das FDM in die Ressourcenplanung einfließen zu lassen. Auch wird damit der Erwartungshaltung großer Forschungsförderer Rechnung getragen, die ein professionelles FDM entsprechend der Community-üblichen Standards erwarten, wie sie beispielsweise von der Gesellschaft für Biologische Daten e.V[15] aufgestellt werden. In der Beratungspraxis ist allerdings zu beobachten, dass bei den Forschenden das Thema FDM in dieser frühen Phase noch einen sehr unterschiedlichen Stellenwert hat. Für einige steht FDM in Konkurrenz zu den fachlichen Inhalten. Dies führt teilweise dazu, dass das Thema erst sehr spät Berücksichtigung findet und somit nur wenig Zeit für konkrete Ausführungen bleibt. Die Herausforderung in der Beratung ist es zudem, einen für das jeweilige Vorhaben und die jeweilige Fachdisziplin angemessenen Rahmen zu definieren, was dadurch noch erschwert wird, dass bisher nur wenige Erfahrungen hinsichtlich der konkreten Erwartungshaltung der Forschungsförderer sowie des Stellenwerts von FDM in der

[15] https://www.gfbio.org/, Zugriff 15.04.2019

Begutachtung vorliegen. Grundsätzlich empfehlen wir, sich bereits im Forschungsantrag freiwillig zur Erstellung eines DMPs zu verpflichten, um dem Fördermittelgeber zu signalisieren, dass die Notwendigkeit eines strukturierten FDMs erkannt ist.

2.2 Beratung zu DMPs in Horizon2020

DMPs sind in Projekten der Europäischen Kommission im Rahmen des ORD-Pilots obligatorisch. Das Spektrum der Beratungsanfragen reicht vom allgemeinen Informationsbedarf zu DMPs bis hin zu sehr spezifischen Fragestellungen. Teilweise wird auch nur um Feedback zu bereits fertig erstellten DMP-Entwürfen gebeten.
Unsere Erfahrungen zeigen, dass mit der Vorgabe der Europäischen Kommission DMPs ausführlich und gewissenhaft erstellt werden. Jedoch ist zu beobachten, dass der Open Science Gedanke bei den Forschenden unterschiedlich ausgeprägt und der Grad der konkreten Umsetzung der FAIR Data Prinzipien ebenso wie die fachlichen Rahmenbedingungen sehr heterogen sind. Dies resultiert teilweise darin, dass Aussagen bezüglich Zugänglichkeit der Daten fehlen bzw. ohne Begründung sehr restriktiv behandelt werden. In einigen Fällen wird auch das Data Sharing nur über die Projektlaufzeit gedacht und die Möglichkeit einer Nachnutzung der Daten nach Projektende nicht in Erwägung gezogen. Dagegen werden die Themen des alltäglichen Datenumgangs (z.B. Dateibenennung, Ordnerstruktur) häufig sehr ausführlich behandelt. Hierin kann sich auch das Bestreben der Forschenden äußern, die verpflichtende Erstellung des DMPs für sich gewinnbringend zu nutzen, um das FDM projektweit für alle Partner zu regeln und festzuhalten.

Weiterhin beobachten wir, dass Forschende, die unsere Weiterbildungsangebote genutzt haben, einen sehr guten Überblick zu den Anforderungen eines DMPs zeigen und diesen entsprechend des Entwicklungsstandes des FDMs in ihrer Community anfertigen können. Für viele Fachbereiche fehlen aber noch die Grundlagen, um die Fair Data Principles in vollem Maße umsetzen zu können. Es fehlen Community-Empfehlungen und etablierte Standards sowie fachspezifische Metadatenschemata, um die Interoperabilität der Forschungsdaten sicherstellen zu können. Durch diese Rahmenbedingungen ist es uns noch nicht möglich, unser Angebotsportfolio für die Beratung und Weiterbildung so fachspezifisch auszurichten, wie wir es für wünschenswert halten.

2.3 Institutsspezifische Anpassung von DMP-Vorlage

Der Ansatz der institutsspezifischen Anpassung von DMP-Vorlagen basiert auf dem Wunsch der Forschenden nach konkreteren DMP-Vorlagen, die nur die für ihre Forschung relevanten Themen abdecken und sowohl die Angebote der RWTH als auch die institutsspezifischen Umsetzungen zum FDM bereits enthalten. Dementsprechend wurde bereits im Rahmen des Projektes zur „Einführung von FDM an der RWTH“ (Hausen et al. 2018[vii]) in Zusammenarbeit mit Doktorandinnen und Doktoranden zweier Institute die allgemeine RWTH-Vorlage auf den jeweiligen Bedarf angepasst. Daraus resultierten deutlich an Fragen reduzierte, aber sehr spezifische DMP-Vorlagen, die bereits die am Institut vorgegebenen Prozesse und Standards enthielten. Durch Umkehrformulierung einzelner Fragen sind nur Spezifizierungen notwendig, wenn von der Vorgabe abgewichen wird, was die Zeit für die Erstellung des DMPs deutlich reduziert. Die Vorlagen wurden als Word-Dokumente umgesetzt, was jedoch für die Forschenden weniger attraktiv war und heute mit der Integration in RDMO optimiert werden könnte.

Eine weitere bedarfsspezifische Vorlage wurde in enger Zusammenarbeit mit der Koordinatorin und gleichzeitig FDM-Verantwortlichen eines Forschungskollegs erarbeitet. Wichtig dabei war der Einbezug aller Kollegiaten, um eine möglichst hohe Akzeptanz zu schaffen. Da dieses Vorgehen auf die zukünftige Arbeit mit Excellenzclustern, Instituten, Graduierten- oder Forschungskollegs übertragen werden soll, wird es im Weiteren genauer vorgestellt:

In einem ersten Beratungsgespräch wurden zunächst die Schwerpunkte der Kollegskoordinatorin zusammengestellt, die aufgrund der heterogenen Zusammensetzung des Kollegs aus mehr als acht verschiedenen Instituten auf der Datenaufbewahrung, der Archivierung sowie auf den erfassten Metadaten lagen. In einem halbtägigen Workshop wurden den Kollegiaten die Grundlagen des FDMs sowie das DMP-Konzept vermittelt. Sensibilisiert für die Anforderungen benannten nun auch die Kollegiaten die für sie wichtigsten Themen und Fragestellungen, die zusammen mit den Schwerpunkten der Koordinatorin in einer kollegspezifischen DMP-Vorlage integriert wurden. Diese besitzt einen Umfang von 20 Fragen und basiert auf der RWTH-Vorlage. Um die Arbeiten bei der darauffolgenden Erstellung des DMPs besser zu strukturieren und den Zeitaufwand überschaubar zu gestalten, wurden den Kollegiaten in einem zweiwöchigen Rhythmus jeweils wenige Fragen eines Themenfeldes geschickt, die innerhalb weniger Minuten zu beantworten waren.

Alle Antworten wurden in einem gemeinsamen DMP zusammengefasst und im RDMO der RWTH für das gesamte Kolleg bereitgestellt. In einem abschließenden Workshop wurden die Ergebnisse vorgestellt und zu einigen Vorgehensweisen Verbesserungsvorschläge gegeben.

Als Ergebnis lässt sich zusammenfassen, dass eine Reduktion der Fragen von 58 in der RWTH-Vorlage auf 20 für das Forschungskolleg möglich war. Die angepasste Vorlage enthält Kollegsspezifika, wie die gemeinsame Kollaborationsplattform oder eigene Vorlagen zur Dokumentation, ebenso wie RWTH-Spezifika, z.B. das institutionelle Repositorium RWTH Publications. Beim Erstellen des DMPs bestanden die größten Unsicherheiten in den Themen Lizenzen und Nutzungsrechte von Forschungsdaten.

2.4 Fachspezifische Anpassung von DMP-Vorlagen

Ein weiterer Ansatz, der ebenfalls auf der Hypothese basiert, dass DMPs mehr als ein formales Dokument sein sollen und als Werkzeug zur Qualitätssicherung der Forschungsdaten im Projekt verstanden werden können, wechselt vom sehr engen Instituts- bzw. Projektbezug auf das Fach. Ziel war es, eine generische DMP-Vorlage bedarfsgerecht anzupassen sowie Antwortoptionen und Hilfetexte fachspezifisch zu erweitern, so dass sie zu gutem FDM im Projekt anleiten.

Die Arbeiten wurden gemeinsam von der Universitätsbibliothek der RWTH Aachen University und der Universitäts- und Landesbibliothek der TU Darmstadt im Rahmen von NFDI4Ing[16] und der gleichzeitigen Einführung von RDMO durchgeführt. Es wurden als Basis der RDMO-Fragenkatalog und als Zielgruppe der Fachbereich Maschinenbau gewählt. Finanziert wurde das Projekt „Forschungsdatenmanagementplanung im Maschinenbau“[17] im Rahmen des „Ideenwettbewerb zum Digitalen Wandel in der Wissenschaft“[18], der vom Forschungszentrum Jülich im Rahmen eines Forschungsvorhabens des BMBF ausgerichtet wurde.

[16] https://nfdi4ing.de/, Zugriff 15.04.2019
[17] https://nfdi4ing.de/projekte/dmp-vorlagen/, Zugriff 15.04.2019
[18] https://www.bildung-forschung.digital/de/ideenwettbewerb-zum-digitalen-wandel-in-der-wissenschaft-2007.html, Zugriff 15.04.2019

Um fachlichen Input zu erhalten und Antworten auf die Fragen zu finden, was gutes FDM im Maschinenbau bedeutet und was Kriterien für die Qualität von Forschungsdaten in diesem Bereich sind, wurde an beiden Hochschulen je ein Workshop mit Wissenschaftlerinnen und Wissenschaftlern der Fachrichtung Maschinenbau veranstaltet (Trautwein-Bruns et al. 2019[viii]).

Die eintägigen Workshops waren in drei Abschnitte gegliedert: Der erste Block diente der Informationsvermittlung zum Thema FDM und DMP. Es wurden die Begriffe FDM, DMP und DMP-Vorlage erklärt und das DMP-Werkzeug RDMO vorgestellt. Zudem wurden die lokalen Angebote der jeweiligen Universität präsentiert. Der zweite Block diente der Reflexion und dem Erfahrungsaustausch der Teilnehmenden untereinander zu den individuellen Praktiken des FDMs. Dabei wurde auf Arten, Formate und Umfang von Forschungsdaten, Arbeitsprozesse, Best Practice Lösungen und die spezifischen Rahmenbedingungen und Anforderungen eingegangen. Insbesondere wurde die Frage diskutiert, was für die Teilnehmenden Qualität von Forschungsdaten ausmacht. Im dritten Block wurde in Form eines „World-Café“ die generische RDMO-Vorlage bezüglich Umfang, Verständlichkeit und Relevanz für den Maschinenbau diskutiert. In einem Etherpad wurden gemeinschaftlich Fragen, Antwortoptionen und Hilfetexte kommentiert oder ergänzt.

Das Ergebnis der Workshops war ein Textdokument mit allen Kommentaren der Teilnehmenden aus Aachen und Darmstadt, die im Folgenden in „Anpassung Frage“, „Anpassung Antwortoption“ oder „Anpassung Hilfetext“ kategorisiert und in die allgemeine RDMO-Vorlage eingearbeitet wurden. Neben der fachspezifischen Anpassung wurden gleichzeitig die lokalen Infrastrukturangebote aufgenommen. Diese modifizierte Vorlage wurde dann den Teilnehmerinnen und Teilnehmern der Workshops im RDMO-Tool bereitgestellt. Diese sollten damit zur Qualitätskontrolle jeweils einen DMP erstellen und Feedback geben, welches wiederum in die Vorlage eingearbeitet wurde. Zukünftig wird die fachspezifisch angepasste „NFDI4Ing-MB-RDMO“-Vorlage im GitHub des RDMO-Projektes[19] zur Nutzung und Weiterentwicklung durch die Community bereitgestellt werden.

Insgesamt musste aufgrund der Heterogenität innerhalb der Disziplin immer noch ein relativ hohes Abstraktionsniveau gehalten werden. Auch wenn sich beispielsweise die

[19] https://github.com/rdmorganiser/rdmo-catalog/tree/master/rdmorganiser/questions,Zugriff 15.04.2019

meisten Teilnehmenden einig waren, dass für sie der ethische Fragenkomplex nicht relevant ist, gab es Teilnehmer, die mit personenbezogenen Daten arbeiteten.

An der allgemeinen RDMO-Vorlage wurden insgesamt 14 Änderungen an Fragetexten, 30 Änderungen an Antwortoptionen und 83 Ergänzungen zu Hilfetexten vorgenommen. Die vorhandenen Fragen wurden grundsätzlich akzeptiert, insbesondere vor dem Hintergrund, dass Fragen optional zu betrachten sind und nicht zwingend beantwortet werden müssen. In einigen Fällen wurden die Fragen jedoch zwecks besserer Verständlichkeit umformuliert bzw. der Hintergrund der Frage in den Hilfetexten erläutert, wenn Frage von den Teilnehmenden nicht oder anders verstanden wurde. Antwortoptionen wurden nur teilweise fachspezifisch angepasst.

Beispielsweise wurde die DFG-Fächersystematik auf den ingenieurwissenschaftlichen Bereich eingeschränkt. Die Vorgabe konkreter Antwortoptionen mit zum Teil sehr individuellen Punkten erschien aber als nicht sinnvoll, da dies bei einer gemeinsamen Vorlage für den gesamten Maschinenbau zu einem Verlust an Handhabbarkeit und Übersichtlichkeit führen würde. In diesen Fällen wurden die Anmerkungen als Hinweis oder Beispiele in den Hilfetext aufgenommen.

Die Motivation der Teilnehmerinnen und Teilnehmer, in der Nacharbeitsphase tatsächlich einen DMP zu erstellen, war sehr gering gegenüber dem Bedarf an Erfahrungsaustausch zum Umgang mit den Forschungsdaten während der Workshops. Es besteht der Wunsch nach Best Practice-Lösungen insbesondere im aktiven FDM, aber entsprechende Standards oder Vorgaben sind zum jetzigen Zeitpunkt oft nicht bekannt. Dies lässt sich in unterschiedliche Richtungen interpretieren. Es könnte sein, dass der Aufwand in der Erstellung des DMPs auch mit der fachspezifisch angepassten Vorlage noch zu hoch gegenüber dem Gewinn, der sich vom DMP versprochen wird, eingeschätzt wird. Da aus unserer Sicht das Thema FDM im Bereich Maschinenbau allgemein noch keinen so hohen Reifegrad erreicht hat, könnte es auch sein, dass noch nicht der richtige Zeitpunkt für die Einführung des Konzepts DMP erreicht ist, sondern dass es zielführender ist, jetzt das Hauptaugenmerk auf die Etablierung disziplinspezifischer Standards und Vorgaben zu legen. Sobald diese existieren, besteht die Chance, die DMP Vorlage konkreter zu gestalten, den Aufwand zu reduzieren und damit ein besseres Aufwand-Nutzen Verhältnis zu realisieren.

3. Bewertung und Diskussion des Beratungsspektrums zu DMPs

Die Anforderungen zum FDM haben sich in den letzten Jahren vor dem Hintergrund von Open Science und wachsenden technischen Möglichkeiten verändert, indem ein größerer Wert auf Nachnutzbarkeit und Interoperabilität von Forschungsdaten gelegt wird. Den Wissenschaftlerinnen und Wissenschaftlern erscheint ein DMP als qualitätssichernde Maßnahme zur Erfüllung dieser Anforderungen zwar oft sinnvoll, jedoch fehlen für viele Bereiche noch Lösungen bzw. konkrete Umsetzungsbeispiele, so dass ein Fragenkatalog allein das Erstellen von DMPs nicht forcieren kann. DMP-Vorlagen müssen von passenden Leitfäden, Empfehlungen und Hilfestellungen begleitet werden. Die Betreuung durch Infrastruktureinrichtungen erscheint sinnvoll, wobei diese das Problem der noch fehlenden Community-Lösungen und Umsetzungsbeispiele nicht alleine lösen können. Die DFG fordert daher die „Fächer, Fachgesellschaften und Communities dazu auf, ihren Umgang mit Forschungsdaten zu reflektieren und angemessene Regularien zur disziplinspezifischen Nutzung und ggf. offenen Bereitstellung von Forschungsdaten zu entwickeln“(DFG 2015[ix]). Weitere Impulse werden insbesondere aus dem Aufbau der Nationalen Forschungsdateninfrastruktur (NFDI) erwartet.

Die meisten DMP-Vorlagen orientieren sich am Forschungsdatenlebenszyklus, können aber je nach Zielsetzung oder fachspezifischem Hintergrund unterschiedliche Schwerpunkte setzen oder inhaltlich variieren. So rückt die Horizon 2020-Vorlage die FAIR-Prinzipien in den Vordergrund, und die DMP-Vorlage der RWTH fokussiert auf den naturwissenschaftlich-technischen Bereich. Mit der fachspezifischen Anpassung der RDMO-Vorlage für den Maschinenbau wurde der Fokus noch ein wenig enger gesetzt. Aber selbst auf der Ebene eines Fachbereichs ist die Zielgruppe noch zu heterogen und entsprechende Vorlagen zu komplex. Der Nachteil dieser komplexen Vorlagen ist, dass sie den Gewinn für die Forschenden nicht direkt erkennen lassen, der Aufwand für die Erstellung von DMPs zu hoch ist und diese in der Folge als zu bürokratisch empfunden und nicht freiwillig erstellt werden.

Die institutsspezifische Anpassung von DMP-Vorlagen hat dagegen gezeigt, dass mit einer homogeneren Gruppe die Konkretisierung mit Reduktion der Fragen und konkreten Antwortoptionen oder Hinweisen funktionieren kann, jedoch sind die Arbeiten sehr aufwändig und bedürfen einer intensiven Mitarbeit seitens der Forschenden. Denn nur wenn deren konkreter Bedarf Berücksichtigung findet,

entstehen Vorlagen, die einen tatsächlichen Mehrwert bieten. Grundsätzlich beobachten wir aber auch eine große Diskrepanz zwischen DMPs, wie sie von Fördermittelgebern mit dem Ziel von Open Data gefordert werden, und DMPs als Teil des eigenen FDMs, wie sie der Forschende als Unterstützung seiner eigenen Forschungstätigkeit empfinden würde. Der Forschende ist vor allem daran interessiert, die Forschungsprozesse während der Projektlaufzeit zu dokumentieren und effektiv zu gestalten. Die Nachnutzung der Daten ist dabei zweitrangig.

Mit den gesammelten Erfahrungen geht die Weiterentwicklung unseres Angebotsspektrums verstärkt in zwei Richtungen: Zum einen werden wir an einer kurzen DMP-Vorlage arbeiten, die nur die wesentlichen Aspekte des FDMs abdeckt und von den Forschenden mit geringem Zeitaufwand befüllt werden kann. Ziel dieser kurzen Vorlage ist es, der Leitlinie der RWTH Aachen University zum FDM zu entsprechen, die wesentlichen Aspekte des FDM frühzeitig zu bedenken und die Infrastrukturangebote der RWTH zum FDM zu vermitteln und ggf. zur Nutzung anzuregen. Zum anderen wollen wir Institute, Verbundprojekte oder Forschergruppen bedarfsorientiert bei der Erstellung der DMPs unterstützen. Hier werden in enger Zusammenarbeit mit den Datenverantwortlichen, z.B. den Datenmanagern der Exzellenzcluster, angepasste DMP-Vorlagen entwickelt, die in Umfang und Inhalt auf die konkreten Anforderungen und die Arbeitsumgebung zugeschnitten sind.

Das Erstellen von DMPs bedeutet viel Aufwand, sowohl auf Seiten der Forschenden als auch auf Seiten der Infrastruktur-Dienstleister im Bestreben, die passende Unterstützung zu liefern. Der Ansatz, den DMP von einem formalen, statischen Textdokument zu einem dynamischen Werkzeug für alle Stakeholder zu wandeln, klingt vielversprechend, ist derzeit aber noch eine Zukunftsvision. Zum jetzigen Zeitpunkt sollten DMPs so kompakt und niederschwellig wie möglich gehalten werden, damit der Aufwand bei der Erstellung gering ist und das Konzept von den Forschenden akzeptiert wird.

Referenzen

[i] Arbeitsgruppe Forschungsdaten (2018). Forschungsdatenmanagement. Eine Handreichung. Potsdam: Deutsches GeoForschungsZentrum GFZ. http://hdl.handle.net/21.11116/0000-0000-BC80-B

[ii] Volker Soßna & Johannes Wespel: Forschungsdatenmanagement in der Antragsberatung: Grundlagen und Empfehlungen für den universitären Forschungsservice. - Hannover: Institutionelles Repositorium der Leibniz Universität Hannover, 2019. DOI: https://doi.org/10.15488/4281

[iii] Tomasz Miksa, João Cardoso, & José Borbinha. (2018). Framing the scope of the common data model for machine-actionable Data Management Plans. Zenodo. http://doi.org/10.5281/zenodo.2161855

[iv] Heike Neuroth, Claudia Engelhardt, Jochen Klar, Jens Ludwig & Harry Enke (2018). Aktives Forschungsdatenmanagement. ABI Technik, 38(1), pp. 55-64, DOI: https://doi.org/10.1515/abitech-2018-0008

[v] Dominik Schmitz, Daniela Hausen, Ute Trautwein-Bruns. Datenmanagement nach Plan. RWTH Aachen University. 2018. Verfügbar unter DOI: 10.18154/RWTH-2018-231100

[vi] Dominik Schmitz, Daniela Hausen, Ute Trautwein-Bruns. Inhalte eines Datenmanagementplans. RWTH Aachen University. 2018. Verfügbar unter DOI: 10.18154/RWTH-2018-224185

[vii] Daniela Hausen, Ulrike Eich, Bela Brenger, Florian Claus, Benedikt Magrean, Elke Müller, Matthias S. Müller, Marius Politze, Stephan von der Ropp, Dominik Schmitz, Ute Trautwein-Bruns & Ann-Kathrin Wluka (2018). Introducing coordinated research data management at RWTH Aachen University : a brief project report. http://doi.org/10.18154/RWTH-2018-224588

[viii] Ute Trautwein-Bruns, Gerald Jagusch, & Stephan von der Ropp (2019). Workshop zur Anpassung von Datenmanagementplanvorlagen im Maschinenbau. Zenodo. http://doi.org/10.5281/zenodo.2598797

[ix] DFG (2015). Leitlinien zum Umgang mit Forschungsdaten. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf

Poster

Implementierung des Forschungsdatenmanagement an der SUH in Hildesheim – Die Rolle der Universitätsbibliothek Hildesheim auf dem Campus

Annette Strauch

UB Uni Hildesheim

Abstract

Die Universitätsbibliothek (UB) Hildesheim ist seit Jahren ein Serviceanbieter für die Forschungsunterstützung im Rahmen von E-Science. Sie stellt digitale Informationsressourcen und Fachportale bereit und unterstützt Forschende beim digitalen Publizieren. Dazu hat sie ein Repositorium für elektronische Dokumente eingerichtet (HilDok). Sie unterstützt die Forschenden technisch und organisatorisch beim Open-Access-Publizieren und verwaltet einen Open-Access-Publikationsfonds. In Kooperation mit dem Rechenzentrum bietet die UB zudem ein Repositorium zur Speicherung, Archivierung, Verwaltung, Verknüpfung und Bereitstellung auch von forschungsrelevanten Daten an (HilData). HilData und HilDok ermöglichen zudem die Verknüpfung von Forschungsdaten mit Forschungspublikationen.

Die SUH hat die strategische Herausforderung für sich erkannt und in der Universitätsbibliothek seit März 2018 eine Stelle Forschungsdatenmanagement besetzt. Im ersten Schritt geht es darum, forschungsaktive Wissenschaftlerinnen und Wissenschaftler im Forschungsdatenmanagement (FDM) konkret zu unterstützen bzw. ein “Awareness“ zu schaffen, um die Forschenden für das Thema FDM zu sensibilisieren sowie ihre Bedarfe und Anforderungen zu ermitteln. Berücksichtigt wird dabei auch, konkrete Lösungen für generische als auch spezielle Anforderungen der Forschenden unter Berücksichtigung der lokalen Bedingungen zu ermitteln.

Die SUH setzt die Beteiligung am Ausbau von Forschungdateninfrastrukturen auf lokaler, regionaler, nationaler und internationaler Ebene fort. In Deutschland hat sich inzwischen der Research Data Management Organiser RDMO als ein wichtiges Werkzeug etabliert und soll auch an der SUH eingesetzt werden. Das seit März 2018 installierte Tool ist an der SUH bisher eine Testinstanz (Stand: Oktober 2018) und wird in den nächsten Monaten von den Softwareentwicklern innerhalb der RDMO-Community kontinuierlich weiterentwickelt bevor erste Workshops an der SUH und ein Nutzen der Forschenden von RDMO möglich sein wird. Fortsetzung der aktiven Beteiligung an der nationalen Plattform forschungsdaten.org (Veranstaltungen, eigene Beiträge, Kooperationen im Rahmen von DINI (Deutsche Initiative für Netzwerkinformation) durch Beiträge auf Veranstaltungen und Jahrestagungen (Poster, Vorträge), Zusammenarbeit mit DARIAH-DE / Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften sowie mit DARIAH-EU, Mitarbeit in der neuen „Kommission für forschungsnahe Dienste“ des Vereins Deutscher Bibliothekarinnen und Bibliothekare (VDB), Kooperation mit der Gesellschaft für wissenschaftliche Datenverarbeitung (GWDG), u.a. im Kontext der „Sync- & Share“-Plattform Academic Cloud und weiteren Fragen zu Speicherungslösungen sowie der digitalen Langzeitarchivierung von Forschungsdaten, Kooperation mit der OER World Map:
https://oerworldmap.org/resource/ im Rahmen von Forschungsdaten und Open Educational Ressources (OER) und Open Science an deutschen Hochschulen.


Forschungsdatenmanagement
Universitätsbibliothek Hildesheim
"Bad Libraries build collections. Good libraries build services
(of which a collection is only one). Great libraries build Communities."
Integration des Forschungsdatenmanagement als Service in einer Bibliothek am Campus.
Datenverbreitung zum Nachnutzen ("Dissemination"), Beratung z.B. zu
- Recht und Lizenzen
- Auffinden von Daten
Sichtbarkeit, Teilen der Forschung ("Sustainability").
Planung des FDM im Sinne der "guten wissenschaftlichen Praxis" (DMP)
- Beratung zu Datenmanagementplänen
- RDMO
Gute Bewilligungschance für Drittmittelanträge.
Generierung
Bearbeitung
Nachnutzung
Zugang
Anreicherung
Archivierung
Stiftung Universität Hildesheim 2003
open access
Lösungen zur Speicherung und Langzeitarchivierung
- Speicherinfrastruktur und Langzeitsicherung
- Zugriffsregelungen ("Access")
Nachhaltiges, qualitätsgerechtes Forschungsdatenmanagement.
Praxisorientiertes Arbeiten mit Forschungswerkzeugen, z.B.
- "Tools"
- Formaten
"Living-lab"-Ansatz. Ermittlung welche Werkzeuge sich bewährt haben und welche nicht.
re3data.org
ACADEMIC CLOUD
Gut sichtbare und frei zugängliche (Open Access) Publikationen, HilDok, HilData, DOI / re3data.org
- Gitlab der SUH, Verknüpfen mit Software
- Datenzitation
Datenrepositorium, Beachtung von rechtlichen und ethischen Aspekten.
Strukturierte Datenorganisation, z.B.
- "File-Transfer" via Academic Cloud
- Projektmanagementplattformen (Redmine, SharePoint, etc.)
- Arbeit mit Metadaten, Metadatenstandards, Klassifikationen, Ontologien
Sparen von Zeit und Aufwand. Metadatenmanagement vereinfacht die Datennachnutzung durch klare Strukturen.
doi
RZ
Das Forschungsdatenmanagement der Stiftung Universität Hildesheim unterstützt praxisbezogen (inkl. Einzelberatungen und zielgerichteten Workshops) mit bereits vorhandenen Angeboten. Forscherinnen und Forscher werden schon vor der Antragstellung und während ihrer gesamten Forschungstätigkeit beraten.
Kontakt: fdm@uni-hildesheim.de
Annette Strauch
Forschungsdaten-Policy und Kooperationen (lokal, regional, national und international)

Forschungsunterstützung im Forschungsdatenmanagement

Birte Lindstädt

ZB MED – Informationszentrum Lebenswissenschaften

Abstract

Forschungsdatenmanagement (FDM) betrifft den gesamten Datenlebenszyklus und erfordert daher ganzheitliche Lösungen, die im Konsens zwischen den Beteiligten gefunden werden müssen. Bibliotheken können in diesem Prozess eine treibende und vermittelnde Kraft sein. Aus dieser Einsicht heraus entwickelt ZB MED unterstützende Services für Forschende und Multiplikatoren (z.B. Bibliotheken) in den Lebenswissenschaften entlang des Lebenszyklus von Forschungsdaten.

Dies beginnt bei der Erstellung eines Datenmanagementplans zur dynamischen Fortschreibung im Projektverlauf. Dazu erprobt ZB MED aktuell die Anwendung des Research Data Management Organizer (RDMO) in einem agrarwissenschaftlichen Forschungsprojekt und stellt eine RDMO-Instanz zur Nutzung durch andere lebenswissenschaftliche Forschungseinrichtungen zur Verfügung.

Im Hinblick auf die Dokumentation von Daten spielen in den Lebenswissenschaften Elektronische Laborbücher (Electronic Lab Notebook ELN) eine wichtige Rolle. Um den Beratungsbedarf zu diesem Thema zu begegnen hat ZB MED einen Leitfaden dazu erstellt und bietet den sog. ELN-Finder zur Filterung geeigneter Tools an.

Für die Dokumentation sind darüber hinaus die Einhaltung von Standards, insbesondere Metadatenstandards, zentral. Im Rahmen von Projektanträgen in einzelnen lebenswissenschaftlichen Fachcommunities werden deshalb Kurationskriterien und Qualitätsstandards von Forschungsdaten gemeinsam mit Forschungseinrichtungen erprobt.
Im Schritt der Publikation bietet ZB MED zum einen eine Publikationsberatung an, zum anderen wird das Fachrepositorium Lebenswissenschaften (FRL) aufgebaut, um ggf. vorhandene Lücken in der Versorgung lebenswissenschaftlicher Forschungseinrichtungen mit Forschungsdatenrepositorien zu schließen. Zum Publikationsservice gehören auch die Vergabe von persistenten Identifikatoren (DOI) – ZB MED ist DOI-Vergabestelle und Data Cite-Mitglied – sowie der Nachweis der Forschungsdaten im Suchportal LIVIVO.

Die Digitale Langzeitarchivierung (dLZA) als einer der letzten Phasen im Datenlebenszyklus wird im Rahmen eines Pilotprojekts mit einem außeruniversitären Forschungsinstitut erprobt, so dass das dauerhafte Lesbarhalten von Forschungsdaten und darauf basierend die dLZA an ZB MED ausgebaut werden kann.

Begleitet werden diese Services durch das Angebot von Workshops zu unterschiedlichen Themen, Vorträge auf lebenswissenschaftlichen und bibliothekarischen Fachkonferenzen sowie von Webinaren und Tutorials.

Forschungsunterstützung im Forschungsdatenmanagement

Services und Beratung im Lebenszyklus lebenswissenschaftlicher Forschungsdaten

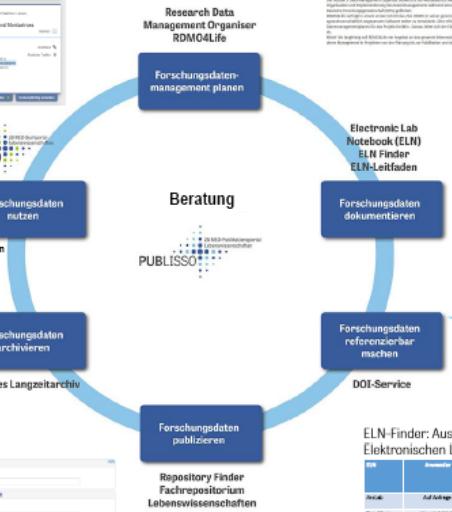
Birte Lindstädt
lindstaedt@zbmed.de
ZB MED

Strategie

- Identifizierung und Schließen von Lücken in der Forschungs-dateninfrastruktur
- Neue Tools erproben und als Service bereitstellen
- Services mit der Forschung bei ZB MED entwickeln
- Passive und aktive Beratungsformate

Services entlang des Lebenszyklus von Forschungsdaten

Darstellung von Forschungsdatensätzen in LIVIVO



Erfassungsmaske für Forschungsdaten im Fachrepositorium Lebenswissenschaften. Erstellt in Kooperation mit: hbz

ELN-Finder: Auswahl und Bewertung von Elektronischen Laborbüchern

Ausblick

- Beteiligung am Aufbau der **Nationalen Forschungs-dateninfrastruktur NFDI** im Rahmen mehrerer Konsortien
- Beteiligung an Forschungsprojekten als Datenmanager („**Data Librarian**“)
- Umsetzung der **FAIR-Prinzipien** in eigenen Infrastrukturen

www.publisso.de

Multiplikator*innen einbeziehen – ein Train-the-Trainer-Programm zum Thema Forschungsdatenmanagement

Petra Buchholz[1] , Kerstin Helbig[2] , Dominika Dolzycka[1], Katarzyna Biernacka[2] , Katrin Cortez[2]

[1]Freie Universität Berlin
[2]Humboldt-Universität zu Berlin

Abstract

Angesichts der knappen Personalressourcen der Infrastruktureinrichtungen versucht das im BMBF-Projekt FDMentor entwickelte Train-the-Trainer Programm zum Thema Forschungsdatenmanagement neben dem hauptamtlich mit dem Thema Forschungsdaten betrauten Personal weitere Personenkreise in die Schulungsaktivitäten einzubeziehen. Dies können einerseits weitere Beschäftigte von Infrastruktureinrichtungen der Universitäten z. B. Fachreferentinnen und Fachreferenten aus den Bibliotheken sein. Anderseits werden auch Wissenschaftlerinnen und Wissenschaftler angesprochen, die in ihren Arbeitsbereichen oder im Rahmen der Lehre die erworbenen Kenntnisse als Multiplikator*innen weitergeben möchten.

Das interaktiv ausgelegte Multiplikatorenprogramm vermittelt Grundlagenkenntnisse in den wichtigsten Themenbereichen des Forschungsdatenmanagements mit Methoden, die von den Teilnehmenden direkt auf eigene Schulungsveranstaltungen übertragen werden können. Ergänzt wird dies durch didaktische Elemente und organisatorisches Know-how zur Durchführung von Schulungen.

Das Konzept und die umfangreichen Begleitmaterialien wurden zur freien Nachnutzung veröffentlicht.

FDMentor

Multiplikator*innen einbeziehen – ein Train-the-Trainer-Programm zum Thema Forschungsdatenmanagement

Ausgangslage

- Personalkapazitäten für FDM bei den Infrastruktureinrichtungen sind häufig erst im Aufbau
- Kenntnisstand zum Forschungsdatenmanagement (FDM) oftmals sehr gering
- Erhöhen der Kompetenzen im FDM ist eine wichtige Aufgabe

Train-the-Trainer-Programm

Zweitägiger Workshop für ca. 12 Teilnehmende

Zielgruppen

- direkt – Trainer*innen mit oder ohne Vorkenntnisse (Bibliothekar*innen, Wissenschaftler*innen, Doktorand*innen, Mitarbeiter*innen von Sonderforschungsbereichen und Dozent*innen ...)
- indirekt – Workshop-Teilnehmende ohne Vorkenntnisse

Workshop-Einheiten sind unverändert auf die späteren Schulungen der Multiplikator*innen übertragbar.

Grundsätze des Programms

Verwendung aktivierender Methoden
Beispiel: Drehen und Wenden
Ziel: Erarbeitung des Forschungsdaten-Lebenszyklus
- Karten mit Schlüsselbegriffen eines Prozesses, Modells oder einer Theorie
- In Gruppen die Karten in eine logische Reihenfolge bringen und das Ergebnis den anderen Teilnehmenden vorstellen

Vorteil: Ermöglicht eine aktive Auseinandersetzung und Diskussion zum Thema

20-Minuten-Regel – maximal 20 Minuten neue Inhalte

Wechselnde Gruppeneinteilungen

Arbeitsmaterialien

PowerPoint-Folien

Druckvorlagen für Methoden

Lehrdrehbücher inklusive Vorlage

Vorlagen für Flipchart-Blätter

Vorlagen für Teilnahmebescheinigungen

Druckvorlagen für Übungen

Checklisten

Feedback-Bogen

Inhalte

Grundlagen des FDM
Digitale Forschungsdaten, Forschungsdaten-Policies, Datenmanagementplan, Ordnung und Struktur, Dokumentation und Metadaten, Speicherung und Backup, Langzeitarchivierung, Zugriffssicherheit, Publikation, Nachnutzung, rechtliche Aspekte, institutionelle Infrastruktur, praktische Übung

Workshop-Gestaltung und Didaktik
Didaktisches Vorgehen, formaler Rahmen, Konzeptentwicklung, didaktische Methoden

Sozialer Rahmen
Begrüßen und Kennenlernen, Orientierung, Abschluss des ersten Tages, Feedback und Verabschiedung

Pilotierungen

9 Pilot-Workshops à 2 Tage, 82 Teilnehmende

Feedback
- direkt im Anschluss an den Workshop
- ca. 4 Monate später per E-Mail

Fazit

Konzeptionell:
- 20-Minuten-Regel fördert Lernerfolg
- Methodenvielfalt fördert Motivation
- Gute Arbeitsatmosphäre ist Aufgabe der Workshop-Leitung

Generell:
- Das Interesse an Train-the-Trainer Veranstaltungen zu FDM ist groß
- Spezifische Angebote werden gebraucht (Disziplinen/Vorkenntnisse)

Veröffentlichung

Das Konzept und alle Arbeitsmaterialien wurden **frei nachnutzbar** auf Zenodo publiziert:
Dominika Dolzycka, Katarzyna Biernacka, Kerstin Helbig und Petra Buchholz, (2019). Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement (Version 2.0). Zenodo. http://doi.org/10.5281/zenodo.2581292

Das Konzept ist zentrales Begleitdokument für Trainer*innen und bietet weitergehende Informationen zu den Themengebieten und zur Durchführung der Einheiten.

Projektlaufzeit: 1. Mai 2017 bis 30. April 2019
AutorInnen: Petra Buchholz, Kerstin Helbig, Dr. Dominika Dolzycka, Katarzyna Biernacka, Katrin Cortez

Kontakt: fdmentor@hu-berlin.de
Twitter: @fd_mentor
https://hu.berlin/fdmentor

Diese Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

gefördert von: Bundesministerium für Bildung und Forschung

Sind wir bereit für die Forschungsdatenhaltung?

Kathleen Neumann[1], Wiebke Oeltjen[2], Ulrike Stahl[3], Robert Stephan[4]

[1]VZG
[2]Universität Hamburg
[3]JKI
[4]Universität Rostock

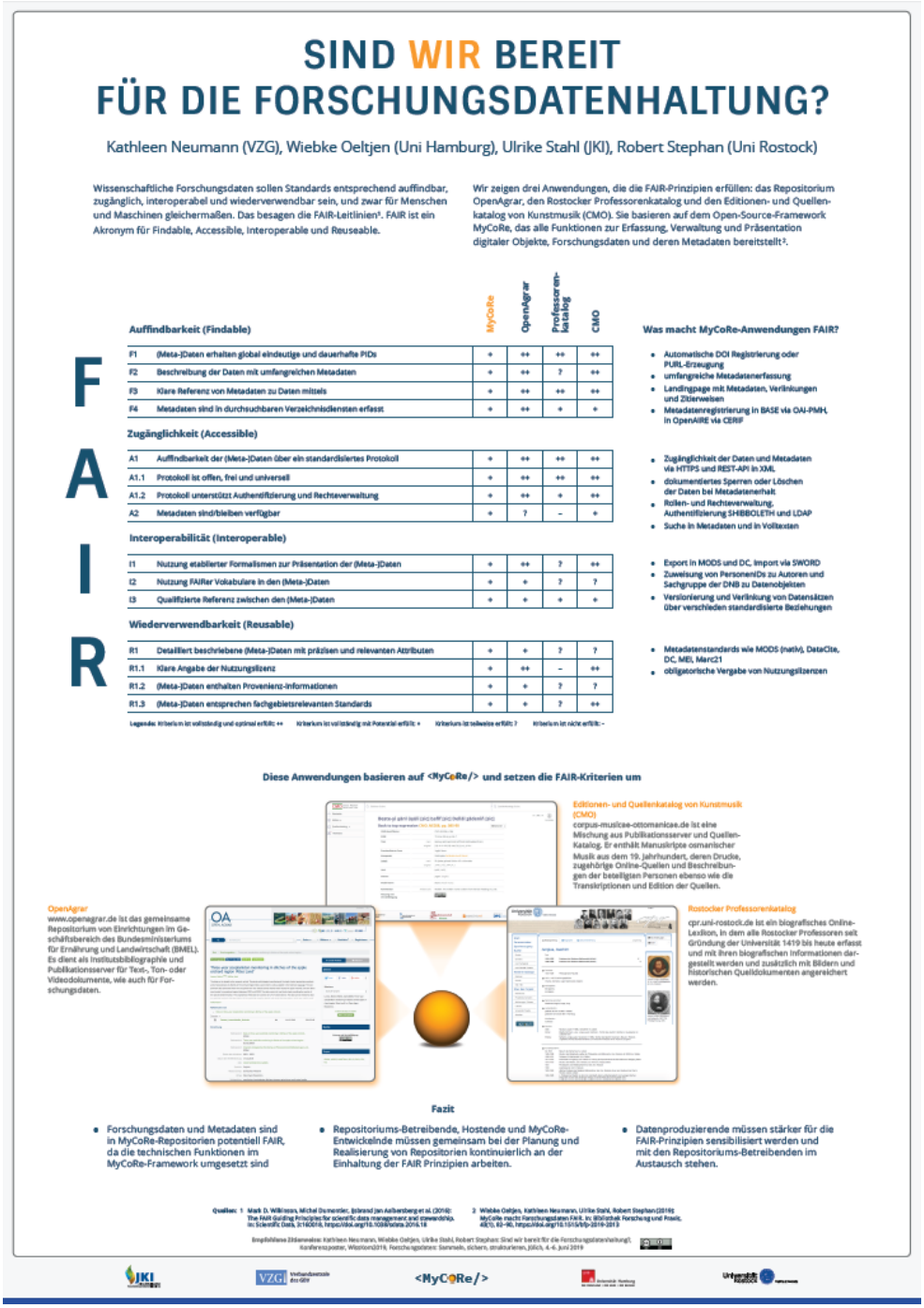
Abstract

Forschungsdaten bilden einen Grundpfeiler wissenschaftlicher Erkenntnis und sind die Basis für weitere Forschung. Eine transparente Dokumentation der Forschungsdaten, ein verantwortungsvolles Forschungsdatenmanagement (FDM) einschließlich qualitätsgesicherter Archivierung und/oder Veröffentlichung sorgen für die Nachvollziehbarkeit und Reproduzierbarkeit von Forschungsprozessen und deren Ergebnissen und ermöglichen eine vielfältige Nachnutzung.

MyCoRe ['maikɔːr] ist ein Open-Source-Framework zur Erfassung, Verwaltung und Präsentation digitaler Objekte und deren Metadaten. Die bis heute mehr als 80 realisierten Anwendungen (z.B. institutionelle Repositorien, Archive und Online-Lexika) enthalten auch verschiedene Forschungsdaten. Von der zitierfähigen Ablage einzelner Forschungsdaten auf Publikationsservern bis zu fachspezifischen Datenbanken und Portalen, zeigt sich dabei ein breites Spektrum.

Seit 2001 wird MyCoRe von einer bundesweiten Gemeinschaft an Universitätsbibliotheken, universitären Rechenzentren und der VZG kontinuierlich weiterentwickelt. Dabei standen schon immer Prinzipien im Mittelpunkt, die wir heute unter anderem als FAIR-Leitprinzipien kennen: Daten und Metadaten sollten in MyCoRe-Webanwendungen im Rahmen einer entsprechenden Infrastruktur auffindbar (Findable), zugänglich (Accessible), interoperabel (Interoperable) und wiederverwendbar (Reusable) sein. Dafür stellt das Framework Schnittstellen und Funktionen bereit, die zum Verwalten, Speichern, Präsentieren und Austauschen von Metadaten und den digitalen Ressourcen benötigt werden.

Im Rahmen der einzelnen Anwendungsentwicklungen und dem laufenden Betrieb hat sich gezeigt, dass es nicht reicht eine Software bereitzustellen, die den FAIR-Prinzipien entspricht. Es kommt auch darauf an, dass in den Projekten, die den Aufbau eines Repositoriums planen und realisieren, kontinuierlich an der Einhaltung der Prinzipien mitgewirkt wird. Dies setzt u.a. eine stärkere Sensibilisierung der Fachwissenschaftler bzw. Repository-Betreiber für die FAIR-Prinzipien voraus. Wir zeigen hier realisierte Umsetzungen der FAIR-Prinzipien ebenso wie nötige Entwicklungsperspektiven sowohl auf rein technischer als auch auf organisatorischer und konzeptioneller Ebene auf.



SIND WIR BEREIT FÜR DIE FORSCHUNGSDATENHALTUNG?

Kathleen Neumann (VZG), Wiebke Oeltjen (Uni Hamburg), Ulrike Stahl (JKI), Robert Stephan (Uni Rostock)

Wissenschaftliche Forschungsdaten sollen Standards entsprechend auffindbar, zugänglich, interoperabel und wiederverwendbar sein, und zwar für Menschen und Maschinen gleichermaßen. Das besagen die FAIR-Leitlinien[1]. FAIR ist ein Akronym für Findable, Accessible, Interoperable und Reuseable.

Wir zeigen drei Anwendungen, die die FAIR-Prinzipien erfüllen: das Repositorium OpenAgrar, den Rostocker Professorenkatalog und den Editionen- und Quellenkatalog von Kunstmusik (CMO). Sie basieren auf dem Open-Source-Framework MyCoRe, das alle Funktionen zur Erfassung, Verwaltung und Präsentation digitaler Objekte, Forschungsdaten und deren Metadaten bereitstellt[2].

		MyCoRe	OpenAgrar	Professoren-katalog	CMO
Auffindbarkeit (Findable)					
F1	(Meta-)Daten erhalten global eindeutige und dauerhafte PIDs	+	++	++	++
F2	Beschreibung der Daten mit umfangreichen Metadaten	+	++	?	++
F3	Klare Referenz von Metadaten zu Daten mittels	+	++	++	++
F4	Metadaten sind in durchsuchbaren Verzeichnisdiensten erfasst	+	++	+	+
Zugänglichkeit (Accessible)					
A1	Auffindbarkeit der (Meta-)Daten über ein standardisiertes Protokoll	+	++	++	++
A1.1	Protokoll ist offen, frei und universell	+	++	++	++
A1.2	Protokoll unterstützt Authentifizierung und Rechteverwaltung	+	++	+	++
A2	Metadaten sind/bleiben verfügbar	+	?	–	+
Interoperabilität (Interoperable)					
I1	Nutzung etablierter Formalismen zur Präsentation der (Meta-)Daten	+	++	?	++
I2	Nutzung FAIRer Vokabulare in den (Meta-)Daten	+	+	?	?
I3	Qualifizierte Referenz zwischen den (Meta-)Daten	+	+	+	+
Wiederverwendbarkeit (Reusable)					
R1	Detailliert beschriebene (Meta-)Daten mit präzisen und relevanten Attributen	+	+	?	?
R1.1	Klare Angabe der Nutzungslizenz	+	++	–	++
R1.2	(Meta-)Daten enthalten Provenienz-Informationen	+	+	?	?
R1.3	(Meta-)Daten entsprechen fachgebietsrelevanten Standards	+	+	?	++

Legende: Kriterium ist vollständig und optimal erfüllt: ++ Kriterium ist vollständig mit Potential erfüllt: + Kriterium ist teilweise erfüllt: ? Kriterium ist nicht erfüllt: –

Was macht MyCoRe-Anwendungen FAIR?

- Automatische DOI Registrierung oder PURL-Erzeugung
- umfangreiche Metadatenerfassung
- Landingpage mit Metadaten, Verlinkungen und Zitierweisen
- Metadatenregistrierung in BASE via OAI-PMH, in OpenAIRE via CERIF
- Zugänglichkeit der Daten und Metadaten via HTTPS und REST-API in XML
- dokumentiertes Sperren oder Löschen der Daten bei Metadatenerhalt
- Rollen- und Rechteverwaltung, Authentifizierung SHIBBOLETH und LDAP
- Suche in Metadaten und in Volltexten
- Export in MODS und DC, Import via SWORD
- Zuweisung von PersonenIDs zu Autoren und Sachgruppe der DNB zu Datenobjekten
- Versionierung und Verlinkung von Datensätzen über verschieden standardisierte Beziehungen
- Metadatenstandards wie MODS (nativ), DataCite, DC, MEI, Marc21
- obligatorische Vergabe von Nutzungslizenzen

Diese Anwendungen basieren auf <MyCoRe/> und setzen die FAIR-Kriterien um

Editionen- und Quellenkatalog von Kunstmusik (CMO)
corpus-musicae-ottomanicae.de ist eine Mischung aus Publikationsserver und Quellen-Katalog. Er enthält Manuskripte osmanischer Musik aus dem 19. Jahrhundert, deren Drucke, zugehörige Online-Quellen und Beschreibungen der beteiligten Personen ebenso wie die Transkriptionen und Edition der Quellen.

OpenAgrar
www.openagrar.de ist das gemeinsame Repositorium von Einrichtungen im Geschäftsbereich des Bundesministeriums für Ernährung und Landwirtschaft (BMEL). Es dient als Institutsbibliographie und Publikationsserver für Text-, Ton- oder Videodokumente, wie auch für Forschungsdaten.

Rostocker Professorenkatalog
cpr.uni-rostock.de ist ein biografisches Online-Lexikon, in dem alle Rostocker Professoren seit Gründung der Universität 1419 bis heute erfasst und mit ihren biografischen Informationen dargestellt werden und zusätzlich mit Bildern und historischen Quelldokumenten angereichert werden.

Fazit

- Forschungsdaten und Metadaten sind in MyCoRe-Repositorien potentiell FAIR, da die technischen Funktionen im MyCoRe-Framework umgesetzt sind
- Repositoriums-Betreibende, Hostende und MyCoRe-Entwickelnde müssen gemeinsam bei der Planung und Realisierung von Repositorien kontinuierlich an der Einhaltung der FAIR Prinzipien arbeiten.
- Datenproduzierende müssen stärker für die FAIR-Prinzipien sensibilisiert werden und mit den Repositoriums-Betreibenden im Austausch stehen.

Quellen: 1 Mark D. Wilkinson, Michel Dumontier, Ijsbrand Jan Aalbersberg et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In: Scientific Data, 3:160018, https://doi.org/10.1038/sdata.2016.18
2 Wiebke Oeltjen, Kathleen Neumann, Ulrike Stahl, Robert Stephan (2019): MyCoRe macht Forschungsdaten FAIR. In: Bibliothek Forschung und Praxis, 43(1), 82–90, https://doi.org/10.1515/bfp-2019-2013

Empfohlene Zitierweise: Kathleen Neumann, Wiebke Oeltjen, Ulrike Stahl, Robert Stephan: Sind wir bereit für die Forschungsdatenverwaltung?, Konferenzposter, Wissenform2019, Forschungsdaten: Sammeln, sichern, strukturieren, jülich, 4.-6. Juni 2019

CC BY

SARA-Service: Forschungsbegleitend Software-Artefakte archivieren und veröffentlichen

Volodymyr Kushnarenko[1], Franziska Rapp[2], Daniel Scharon[3], Stefan Kombrink[2], Matthias Fratz[3], Pia Schmücker[2], Marcel Waldvogel[3], Stefan Wesner[1,2]

[1]Institut für Organisation und Management von Informationssystemen (OMI), Universität Ulm
[2]Kommunikations- und Informationszentrum (kiz), Universität Ulm
[3]AG Verteilte Systeme, FB Informatik und Informationswissenschaft, Universität Konstanz

Abstract

Heutzutage geht der Forschungsprozess in vielen Disziplinen sehr eng mit der Entwicklung und dem Einsatz von Software einher. Software-Artefakte (Quellcode, Skripte, ...) werden für die Erfassung, Verarbeitung und Analyse der Forschungsdaten verwendet. Manchmal ist Software selbst Gegenstand der Forschung. Die verwendeten Software-Artefakte sollten ebenso wie die dazugehörigen Forschungsdaten langfristig verfügbar gemacht werden, um die Nachvollziehbarkeit und Reproduzierbarkeit von Forschung zu verbessern. Hierfür entwickelt das Projekt SARA (Software Archiving of Research Artefacts) einen wissenschaftlichen Dienst, der die Langzeitspeicherung und Publikation von Software-Artefakten auf einfache und benutzerfreundliche Weise anbietet [1].

Das Poster präsentiert die Hauptmerkmale des entwickelten Services und zeigt, wie Forschende die im Forschungsprozess verwendeten Software-Artefakte in wenigen Schritten archivieren und, falls gewünscht, zusätzlich mit einem persistenten Link zitierfähig veröffentlichen können. Forschende können ihre Projekte aus Git-Repositorien – GitLab-Instanzen und GitHub – zur langfristigen Speicherung in ein angebundenes Git-Archiv überführen und in einem angebundenen institutionellen Repositorium (IR) zitierbar veröffentlichen. Über die Webplattform kann das gesamte Projekt oder können auch nur einzelne Branches für die Speicherung bzw. Veröffentlichung selektiert werden. Forschende können entscheiden, ob nur eine bestimmte Version des Software-Artefaktes veröffentlicht wird, die ganze Entwicklungshistorie oder eine verkürzte Form der Historie.

Als Archiv wird ein GitLab-Server verwendet, der auch als institutionsübergreifender Service angeboten werden könnte. Für die Anbindung an institutionellen Repositorien liegt der Fokus derzeit auf Repositorien, die auf der Software DSpace basieren. Eine beliebige Anzahl von IR-Instanzen kann an den SARA-Service angebunden werden. Die Unterstützung von weiterer Repositorien-Software ist geplant. Bei der Veröffentlichung werden Metadaten definiert und ans gewünschte IR übermittelt. So startet der SARA-Service die Veröffentlichungsprozedur und leitet den Nutzer für die finale Korrektur und Publikationsbestätigung ins IR weiter.

Der SARA-Server kann als Service innerhalb einer einzelnen Einrichtung angeboten werden oder von mehreren Einrichtungen gemeinsam genutzt werden. Er kann den Forschungsprozess begleitend zur Speicherung der verschiedenen Versionen der Software-Artefakte genutzt werden und dazu beitragen, dass softwaregestützte oder softwarebezogene Forschung besser nachvollziehbar und reproduzierbar wird.

[1] https://www.sara-service.org

SARA

SARA
Software Archiving of Research Artefacts

https://sara-service.org

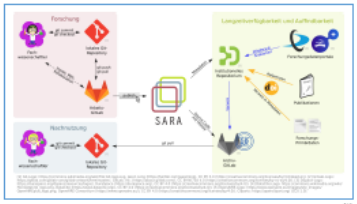
Kushnarenko, V.; Rapp, F.; Scharon, D.; Kombrink, S.; Fratz, M.; Schmücker, P.; Waldvogel, M.; Wesner, S.
Kontakt: info@sara-service.org

SARA-Service: Forschungsbegleitend Software-Artefakte archivieren und veröffentlichen

Motivation

Software spielt in vielen Disziplinen eine wichtige Rolle im Forschungsprozess. Damit Forschungsergebnisse später nachvollziehbar sind, sollte Software **langfristig verfügbar** und **referenzierbar** gemacht werden.

Mit SARA gibt es einen Webservice, der die **Archivierung** und **Veröffentlichung von Software** aus GitHub und GitLab-Instanzen auf einfache Weise ermöglicht. Dabei stehen die langfristige Verfügbarkeit der Software über ein zentrales Git-Archiv und ihre **Zitierbarkeit** über die Veröffentlichung in einem institutionellen Repositorium (IR) im Vordergrund. SARA ist von Wissenschaftler für Wissenschaft kreiert und orientiert sich immer auf Wünsche von Community.



Merkmale

- Extraktion von Daten aus GitLab-Instanzen / GitHub
- Umfang der SW-Entstehungshistorie ist anpassbar
- Langfristige Speicherung durch Git-Archiv
- DOI und Landing Page mit Metadaten
- Unterstützung von DSpace-basierten Repositorien
- Community getriebene Dienstentwicklung

1. Git-Repositorium auswählen

SARA hilft Wissenschaftlern die Artefakten aus GitLab-Instanzen oder GitHub zu archivieren und publizieren. Authentifizierung über **OAuth** und **Shibboleth** ist unterstützt.

2. Historie-Umfang anpassen

Archivierung und Publikation - von ganzer Commit-Historie bis zu bestimmten Ihren Teilen.

3. Metadaten beschreiben

SARA sichert, dass ein minimaler Metadaten-Satz erfasst wird. Metadaten-Extraktion aus GitLab / GitHub ist unterstützt.

4. Lizenz & Zugriff

Lizenzauswahl ist wichtig für weitere Nachnutzung. SARA bietet eine Liste gängiger Lizenzen an und versucht auch, die aktuell verwendete Lizenz zu erkennen. Der Zugriff zu den Artefakten ist konfigurierbar – öffentlich für alle oder privat.

5. Archivierung & Publikation

Software-Artefakt, Lizenz und Metadaten werden nach Git-Archiv übertragen. Falls öffentlicher Zugriff als Option gewählt ist, wird auch die Publikation in einem IR angestoßen.

Wissenschaft

1. Softwarewerkzeuge als Forschungsteil sind langfristig archiviert
2. Entstehungshistorie der Software kann mitarchiviert werden
3. Veröffentlichung, Verbreitung und weitere Zitierung der Software über DOI und Landing Page in IR
4. Bestimmung des Umfangs der Veröffentlichung
5. Automatisierungsmechanismen bei der Metadatenbeschreibung

Ergebnis: Nachnutzung & Nachvollziehbarkeit

Infrastruktur

1. **Individueller Betrieb**
   - SARA-Dienst innerhalb nur einer Einrichtung
   - SARA-Server, GitLab-Instanz
   - Institutionelles Repositorium, Git-Archiv
2. **Kooperativer Betrieb**
   - eine SARA-Instanz für mehrere Einrichtungen
   - leichte Integration, Anbindung und Anpassung

Community

1. Open Source Projekt auf GitHub: https://github.com/sara-service
2. Interesse an Kooperationen und Erweiterungen des Dienstes
3. Zusammenarbeit mit Wissenschaftlern und Infrastruktureinrichtungen

Universität Konstanz
Universität Konstanz:
AG Verteilte Systeme,
FB Informatik

ulm university
universität ulm
Universität Ulm:
Kommunikations- und Informationszentrum (kiz),
Institut für Organisation und Management von Informationssystemen (IOMI)

Gefördert vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg
Baden-Württemberg

DORO: ein Repositorium zur Forschungsdatenarchivierung

Robert Stephan[1], Karsten Labahn[2]

Universitätsbibliothek Rostock

Abstract
Im Rahmen von universitären und drittmittelgeförderten Projekten werden Forschungsdaten (u.a. in Form von Digitalisaten) erzeugt und in spezialisierten Anwendungen und Portalen präsentiert. Die Universitätsbibliothek Rostock (UB) und das IT- und Medienzentrum (ITMZ) unterstützen Wissenschaftler_innen dabei, die von den Forschungsförderern (z.B. DFG) geforderte Aufgabe zu erfüllen, die Daten langfristig zu archivieren und im Internet zur Verfügung zu stellen. Mit dem Repository DORO[3] (Digital Objects at Rostock University) stellt die UB dafür eine Plattform bereit. Das DORO-Repository dient als Backend-Lösung für verschiedene Spezialanwendungen, die auf die gespeicherten Daten und Metadaten über Schnittstellen zugreifen und für die Präsentation nutzen. Es basiert auf dem Repository-Framework MyCoRe[4], dessen Komponenten die für den Betrieb von klassischen Dokumentenserver notwendigen Datenformate und Schnittstellen unterstützen. Es soll gezeigt werden, dass diese auch gut für die Bereitstellung von Forschungsdaten geeignet sind.

Metadaten lassen sich mit MyCoRe frei modellieren oder in einem beliebigen XML-Format (z.B. METS / MODS) speichern. Verschiedene Persistent Identifier (wie DOI, URN und PURL) können verwaltet und registriert werden. Durch die SOLR-Integration kann auf eine ausgereifte, konfigurierbare Software für die Indexierung und Recherche in den Datensätzen und deren Metadaten zurückgegriffen werden. Ein Imageviewer ermöglicht es, Digitalisate sowie Volltexte im ALTO oder TEI-Format zu präsentieren. Über die IIIF-API kann eine Vielzahl weiterer Viewer bedient werden. Ausgehend von den Metadaten im MODS-Format, das sich auch für die Speicherung von Metadaten für Forschungsdatensätze eignet, und der Beschreibung der Dateistruktur in METS können durch XSL-Transformationen weitere Datenformate erzeugt und für die Meldung an weitere Portale und Aggregatoren (z.B. Datacite / OpenAIRE) über eine flexibel konfigurierbare OAI-Schnittstelle ausgeliefert werden.

Eine dauerhafte Pflege der im Rahmen von Forschungsprojekten entstehenden, hochkomplexen und spezialisierten Anwendungen kann über die Laufzeit der jeweiligen Projekte hinaus nicht unbegrenzt abgesichert werden. Gründe dafür sind u.a. die Realisierung mit unterschiedlichster Software und Funktionalität, technische Alterung der Systeme oder Forscher, die die Universität verlassen haben. Deshalb dient DORO auch als Fallback-Lösung. Die Daten bleiben langfristig verfügbar und über Persistent Identifier zitierfähig und recherchierbar. Das Frontend von DORO kann die Funktionalität der Spezialanwendung nicht ersetzen, stellt aber einen einfachen Zugang zu den Daten sicher.

Schließlich ergeben sich aus der Verwendung von MyCoRe als Basis für das Forschungsdatenrepositorium (DORO) und den Publikationsserver (RosDok) Synergieeffekte beim Aufbau von Knowhow, der Entwicklung und dem Betrieb.

[1] robert.stephan@uni-rostock.de, https://orcid.org/0000-0001-7605-7415
[2] karsten.labahn@uni-rostock.de, https://orcid.org/0000-0002-8482-807X
[3] http://doro.uni-rostock.de
[4] http://www.mycore.org

Universität Rostock
Traditio et Innovatio

DORO ein Repositorium zur Forschungsdatenarchivierung

Ausgangssituation

Mit DORO stellt die Universitätsbibliothek Rostock Wissenschaftler_innen eine Plattform zur Verfügung, um Forschungsdaten langfristig zu archivieren und dauerhaft im Internet zu publizieren.

DORO dient als Backend-Lösung für Spezialanwendungen, die auf die Daten und Metadaten über Schnittstellen zugreifen und diese für die Präsentation nutzen können.

Gleichzeitig dient es als Fallback-Lösung: Kann eine Spezialanwendung aus verschiedenen Gründen (begrenzte Projektlaufzeit / technischer Alterung / Forscher_innen, die die Universität verlassen / ...) nicht mehr betrieben werden, bleiben die Daten über ihren Persistent Identifier zitier- und zugreifbar und eine vereinfachte Präsentation über DORO wird gewährleistet.

Projekte (Datenlieferanten)

IPAC - Elektronische Zettelkataloge der Universitätsbibliothek Rostock
In einem Digitalisierungsprojekt wurden verschiedene Zettelkataloge der UB Rostock mit insgesamt ca. 550.000 Zetteln digitalisiert und im Internet zugänglich gemacht.

DARL - Digitales Archiv zum Rostocker Liederbuch
Das Rostocker Liederbuch ist eine mittelalterliche Handschrift mit heute noch 44 Blättern. Das Archiv enthält digitalisierte Materialien (Abbildungen, Parallelüberlieferungen, Publikationen), die einen Bezug zu diesem Werk haben.

WossiDIA-Wossidlo Digital Archive
Die handschriftliche Sammlung des mecklenburgischen Feldforschers Richard Wossidlo (1859-1939) umfasst ca. 2,3 Mio Belege mit Feldforschungsdaten, Korrespondenzen, Exzerpten aus Publikationen,... die im Rahmen eines DFG-Projektes digitalisiert und in einem Webportal veröffentlicht wurden.

DEMel - Diccionario del Español Medieval electrónico
In diesem laufenden DFG-Projekt wird ein umfangreiches Datenarchiv zum mittelalterlichen Spanisch digitalisiert und publiziert. Grundlage bilden ca. 900.000 Karteikarten, die Wortbelege aus ca. 600 Werken bzw. Textsammlungen des 10. bis 15. Jh. verzeichnen.

Schnittstellen

REST-API	für das Lesen und Schreiben von Metadaten und Dateien sowie für die Suche (Proxy auf den integrierten SOLR-Server)
OAI-PMH	Konfiguration mehrerer OAI-PMH-Endpoints für den Austausch mit Nachweissystemen
IIIF-Image-API	für die Bereitstellung von Bilddaten Auslieferung von Bildkacheln für verschiedene Zoomstufen

powered by <hyCoRe/>

Formate und Standards

MODS - Metadata Object Description Standard
Die Forschungsdatensätze werden mit einem minimalen MODS-Metadatensatz beschrieben. Dieser enthält mindestens einen internen Objekt-Identifier, DOIs (wenn vergeben), Titel, eine Objektklassifikation und die bildernde Einrichtung.

METS - Metadata Encoding and Transmission Scheme
Die Beschreibung der Binnenstruktur eines Forschungsdatensatzes erfolgt mit METS. Damit lassen sich beschreibende, administrative und strukturelle Metadaten der Objekte sowie wichtige Datei-Metadaten (wie Größe, Änderungsdatum, Checksumme) speichern.

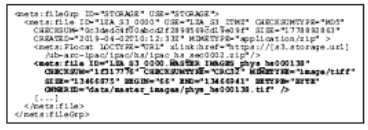
Persistent Identifier
Als Persistent Identifier werden derzeit DOIs und PURLs unterstützt. Diese werden i.d.R. von der internen ID des Forschungsdatensatzes abgeleitet.

S3-Storage (Archivierung)
Die dauerhafte Speicherung der Forschungsdatensätze erfolgt redundant in je einem S3-Storage beim IT- und Medienzentrum (ITMZ) der Universität Rostock und der Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG) in Göttingen.

BagIt (Binnenstruktur des S3-Objektes)
Die Daten werden gemäß dem BagIt-Standard abgelegt und mit den dafür notwendigen Metadaten- und Checksummen-Dateien angereichert.

BagIt-Datenstruktur *bagit.txt*

ZIP64 uncompressed
Der ZIP-Standard wurde 1989 entwickelt und ist heute Public Domain. In der Erweiterung ZIP64 lassen sich ZIP Dateien mit einer Gesamtgröße von 16 ExaByte speichern. Schaltet man die Komprimierung aus, lassen sich einzelne Dateien (sofern ihre Byte-Position bekannt ist) ohne Kenntnis des ZIP-Standards entnehmen, da ihr Byte-Strom unverändert übernommen wurde.



METS-Filegroup eines ZIP-Containers (mit enthaltenen Dateien)

Fazit

- Erste Erfahrungen mit dokumenten-/bildbasierten Forschungsdaten sind positiv.
- Für naturwissenschaftliche und technische Daten fehlen derzeit noch praktische Erfahrungen.
- Es ist offen, ob und wie weitere Systeme (relationale Datenbanken, XML-Datenbanken, Triple Stores) für eine performante Recherche integriert werden sollen.
- Herausfordernd bleibt die Zusammenstellung (Bildung von Intellectual Entities) der Forschungsobjekte - Wie granular sollen die Daten gespeichert werden?

UNIVERSITÄTSBIBLIOTHEK ROSTOCK
Robert Stephan, Karsten Labahn
digilib.ub@uni-rostock.de

Universitäts Bibliothek Rostock

Management digitaler Forschungsdaten im akademischen Umfeld - Lessons learned aus der Einführung von RADAR

Kerstin Soltau[1], Matthias Razum[1], Dorothea Strecker[2]

[1]FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur
[2]Humboldt-Universität zu Berlin

Abstract

RADAR (www.radar-service.eu) ist ein disziplinübergreifender Cloud-Dienst für die langfristige Archivierung und Publikation digitaler Forschungsdaten und wird von FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur betrieben. Derzeit wendet sich der Dienst primär an Hochschulen und außeruniversitäre Forschungs-einrichtungen, die keine eigene Forschungsdateninfrastruktur betreiben oder die RADAR ergänzend zu existierenden disziplinspezifischen Angeboten nutzen möchten.

Das Poster stellt das aktuelle Dienstleistungsspektrum, die modular aufgebaute Systemarchitektur und das Rollen- und Rechtemanagement von RADAR dar, welches die delegierte Administration durch nutzende Einrichtungen erlaubt.

Seit RADAR 2017 den Betrieb aufnahm, hat sich das Umfeld für Forschungs-datenrepositorien dynamisch weiterentwickelt. Mit der Nationalen Forschungsdaten-infrastruktur (NFDI) entsteht ein übergreifendes Forschungsdatenmanagement für das deutsche Wissenschaftssystem.

RADAR sieht sich dabei als ein generischer Baustein für Konsortien innerhalb der NFDI. Das Poster zeigt, wie die sich daraus ergebenden Herausforderungen durch eine strategische Öffnung für alternative Betriebsmodelle und zielgerichtete Funktionserweiterungen angegangen werden. Disziplinspezifische Anforderungen können z. B. über die Implementierung fachspezifischer Metadatenschemata und kontrollierter Vokabulare abgedeckt werden. Auch die zukünftige Kooperation mit anderen Forschungsdatendiensten ist möglich, beispielsweise die Verknüpfung mit Datenmanagementplan-Werkzeugen oder die Anbindung disziplinspezifischer Repositorien an RADAR. Weiterhin steht RADAR alternativen Einsatzszenarien offen, z.B. der Einbindung von Rechenzentren von Institutionen oder Konsortien als Storage Layer oder dem Betrieb der RADAR-Software inklusive Datenspeicherung auf eigenen Infrastrukturen.

RADAR
FIZ Karlsruhe

Archivierung und Publikation von Forschungsdaten mit RADAR

RADAR Dienstleistungen

Forschungsdaten archivieren

- Sichere Verwahrung ohne Veröffentlichung
- Flexible Haltefrist (5, 10, 15 J.)
- Disziplin- und formatunabhängig
- Flexible Zugriffsverwaltung: privat (Standard)/geteilt/öffentlich

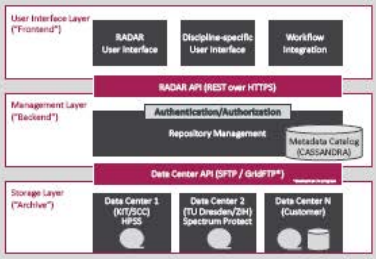
Forschungsdaten publizieren

- Sichere Verwahrung mit Veröffentlichung
- garantierte Haltefrist (mind. 25 J.)
- Disziplin- und formatunabhängig
- DOI-Registrierung und -Reservierung
- Optionale Embargofrist
- Auswahl von Lizenzen (z.B. Creative Commons)
- Automatische Metadaten-Indexierung (DataCite, OAI-PMH...)

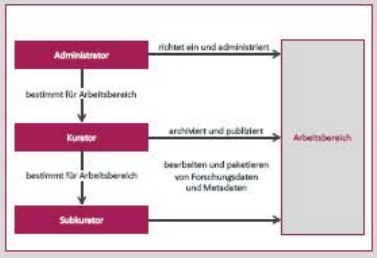
Peer-Review von Forschungsdaten

- Begutachtung vor Veröffentlichung
- privater Link für externe Gutachter
- Verlängerung oder Abbruch des Review-Zeitraums

Systemarchitektur



Rollen- und Rechtemanagement



Vorteile von RADAR für Institutionen

- Generisches Repositorium für Forschungsdaten aller Fachdisziplinen
- Reputationsfördernde Sichtbarkeit des institutionellen Forschungsoutputs
- Keine Notwendigkeit für Unternalt einer eigenen Infrastruktur: Schont Ressourcen und Kosten, ist schnell implementierbar.
- Standardbasierte Datenablage, offene Systemarchitektur und Programmierschnittstellen verhindern die Gefahr eines „Vendor Lock-ins"
- RADAR-Service und -Infrastruktur unterliegen deutschem Recht
- Delegierte Administration durch die nutzende Einrichtung dank klar definierbarem Rollen- und Rechtemanagement
- Flexibel an bestehende Workflows zum Forschungsdatenmanagement anpassbar (Integration über API, Authentifizierung über DFN-AAI, Corporate Design, Kostenkontrolle, Embargos etc.)
- Disziplinübergreifende Metadatenverwaltung. Interoperables Metadatenschema (u.a. DataCite) fördert die Umsetzung der FAIR-Prinzipien

RADAR als generischer Baustein für die NFDI

Mit disziplinspezifischen Erweiterungen:
→ Fachspezifische Metadaten / kontrollierte Vokabulare (Annotation, Suche und Harvesting)

Mit alternativen Betriebsmodellen:
→ Rechenzentren bzw. Speicherdienste des Konsortiums werden als Storage Layer eingebunden
→ RADAR wird durch die nutzende Einrichtung bzw. das Konsortium betrieben

In Kooperation mit anderen Forschungsdaten-Diensten:
→ Andere generische Dienste, z.B. ein Datenmanagementplan-Werkzeug, werden mit RADAR verknüpft
→ Disziplinspezifische Repositorien werden an RADAR (als institutionelles Repositorium) angebunden

NFDI

Kontakt

www.radar-service.eu
info@radar-service.eu
Twitter: @RADAR_Service

Möchten Sie RADAR nutzen oder mit uns kooperieren? Haben Sie weitere Ideen? Sprechen Sie uns an!

Testoption

test.radar-service.eu

Institutionen können RADAR vorab kostenlos und unverbindlich testen! Mehr unter: www.radar-service.eu/de/radar-testsystem

CC BY
Lizenziert unter CC-BY 4.0 | https://creativecommons.org/licenses/by/4.0/
Mai 2019

Kerstin.Soltau@fiz-karlsruhe.de *
Matthias.Razum@fiz-karlsruhe.de *
Dorothea.Strecker@fiz-karlsruhe.de **
* FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur
** Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

ConfIDent - Eine verlässsliche Plattform für wissenschaftliche Veranstaltungen

Stephanie Hagemann-Wilholt

TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften, Hannover

Abstract

Konferenzen stellen einen zentralen und etablierten Rahmen zur informellen Veröffentlichung von Forschungsergebnissen an unterschiedlichen Stationen innerhalb des Lebenszyklus eines Forschungsprojektes dar: von der Idee bis zum veröffentlichten Ergebnis. Sie boten auch schon vor der sich intensivierenden Debatte um die Publikation von Forschungsdaten die Möglichkeit, Daten als Ressourcen von Forschung zu präsentieren und zu diskutieren. Oftmals erhalten Konferenzbeiträge jedoch erst mit Veröffentlichung von Proceedings Sichtbarkeit in der Forschungscommunity. Dabei leisten Konferenzbeiträge selbst einen wichtigen Beitrag für die Forschung: Sie präsentieren zeitnah den aktuellen Stand eines Vorhabens und dienen als Diskussionsforum zur Schärfung von Konzepten.
Die TIB plant gemeinsam mit dem Institut für Informatik III an der Universität Bonn den Aufbau einer Konferenzplattform (ConfIDent), die diese benannten Mehrwerte von Konferenzen nutzbar macht. ConfIDent zielt auf die qualitätsorientierte, kollaborative Kuratierung von semantisch strukturierten Metadaten wissenschaftlicher Veranstaltungen. Die Plattform wird zuverlässige und transparente Daten und Arbeitsabläufe für Forschende sowie andere Stakeholder wie Universitäten, Bibliotheken, Sponsoren, Verlage oder Fachgesellschaften bereitstellen.
Ausgehend von zwei Pilotcommunities – der Informatik sowie der Verkehrs- und Mobilitätsforschung – besteht das Ziel darin, den Bedarf der Nutzenden zu ermitteln, um die Benutzerfreundlichkeit und Akzeptanz der Plattform zu fördern. Langfristiges Ziel ist es, die Ergebnisse des Projekts als Modell für Communities aus Naturwissenschaften und Technik bereitzustellen.
Ein benutzerzentrierter Ansatz, die Anbindung an bestehende Dienste zum Datenaustausch und die Einhaltung der FAIR-Prinzipien sollen die Nachhaltigkeit des Dienstes gewährleisten. Das gegenwärtige Desiderat langfristig auffindbarer, offener, referenzierbarer und nachnutzbarer Metadaten zu wissenschaftlichen Veranstaltungen soll behoben werden. Das Projekt basiert auf der bereits bestehenden Plattform OpenResearch.org, die semantische Beschreibungen wissenschaftlicher Ereignisse verwendet. Die Definition von Metadatenanforderungen dient der Strukturierung von Inhaltsinformationen, der Zuordnung von persistenten Identifikatoren und der Entwicklung von Qualitätskriterien zu Veranstaltungsdaten. Die Entwicklung und Anwendung eines vielseitigen Toolkits ermöglicht es Datenlieferanten und Anwendern zwischen verschiedenen Stufen der Datengranularität zu wählen, um die Auswahl von Qualitätskriterien an ihre Bedürfnisse anzupassen. ConfIDent wird eine weitergehende Analyse der Metadaten und die Entwicklung spezifischer Indikatoren und Metriken unterstützen, um eine bessere Sichtbarkeit wissenschaftlicher Veranstaltungen als unabhängige Leistung von Forschenden und als Teil des Lebenszyklus von Forschung zu gewährleisten: Die Daten können Kontextinformationen zu CfPs, Proceedings, Referenten, Forschungsdaten, -themen oder Konferenzreihen umfassen.

CONFIDENT – FÜR FAIRE KONFERENZMETADATEN

ENTWICKLUNG EINER NACHHALTIGEN PLATTFORM FÜR DIE DAUERHAFTE UND VERLÄSSLICHE SPEICHERUNG UND BEREITSTELLUNG VON KONFERENZMETADATEN.

// FAKE SCIENCE

- Flüchtigkeit von Konferenzinformationen
- Fehlende Qualitätskriterien und -standards für Konferenzen und Konferenzmetadaten
- Ambiguität von Informationen
- Fachspezifische Charakteristika von Konferenzen

// PERSISTENTE METADATEN & QUALITÄTSKRITERIEN

- FAIRe Daten gemäß der FAIR DATA Principles
- transparente, fachspezifische Qualitätskriterien
- Förderung von Open Science
- mehr Sichtbarkeit von Forschungsleistungen
- bessere Vernetzung

// STAKEHOLDER

- Welche Rolle spielen Konferenzen in den jeweiligen Communities?
- Welche Informationsbedürfnisse haben die jeweiligen Stakeholder? Welche Anforderungen leiten sich für Metadaten und Qualitäts-indikatoren daraus ab?
- Servicestrukturen aufbauen
- Peer Review

// PERSISTENT IDENTIFIER (PID)

- Entwicklung eines PIDs durch die internationale Initiative PIDs for conferences
- Einbindung von PIDs in die Plattform
- Beitrag zum PID Graph: Vernetzung von Metadataten

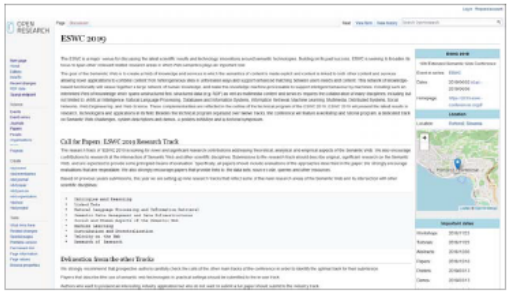
// METRIKEN & VISUALISIERUNG

- Entwicklung alternativer Metriken für Konferenzen & Visualisierungen von Konferenzinformationen
- Einbindung in Forschungs-informationssysteme
- Kulturwandel für die Anerkennung von Forschungsleistungen jenseits des Journal Publishing befördern

Source: TIB, Co-author Network (https://vivo.tib.eu/lib/vis/author-network/0000-0001-2695-7741)

// TECHNISCHE ARCHITEKTUR

- Wiki-basierte Plattform zur kollaborativen Bearbeitung von Daten (Semantic Media Wiki)
- offene Schnittstellen, Interoperabilität mit existierenden Metadatenschemata
- Visualisierungen als Analyse- und Servicetools
- Dissemination von Forschungsideen: Einbindung in einen Open Research Knowledge Graph über Linked Open Data
- Einbindung fachspezifischer Ontologien
- Transparenz durch Informationen zur Datenprovenienz



Source: TIB, OpenResearch (https://www.openresearch.org/wiki/ESWC_2019)

Mehr Informationen:
Dr. Stephanie Hagemann-Wilholt
Technische Informationsbibliothek (TIB)
Welfengarten 1B
30167 Hannover
stephanie.hagemann@tib.eu

TIB LEIBNIZ-INFORMATIONSZENTRUM TECHNIK UND NATURWISSENSCHAFTEN UNIVERSITÄTSBIBLIOTHEK

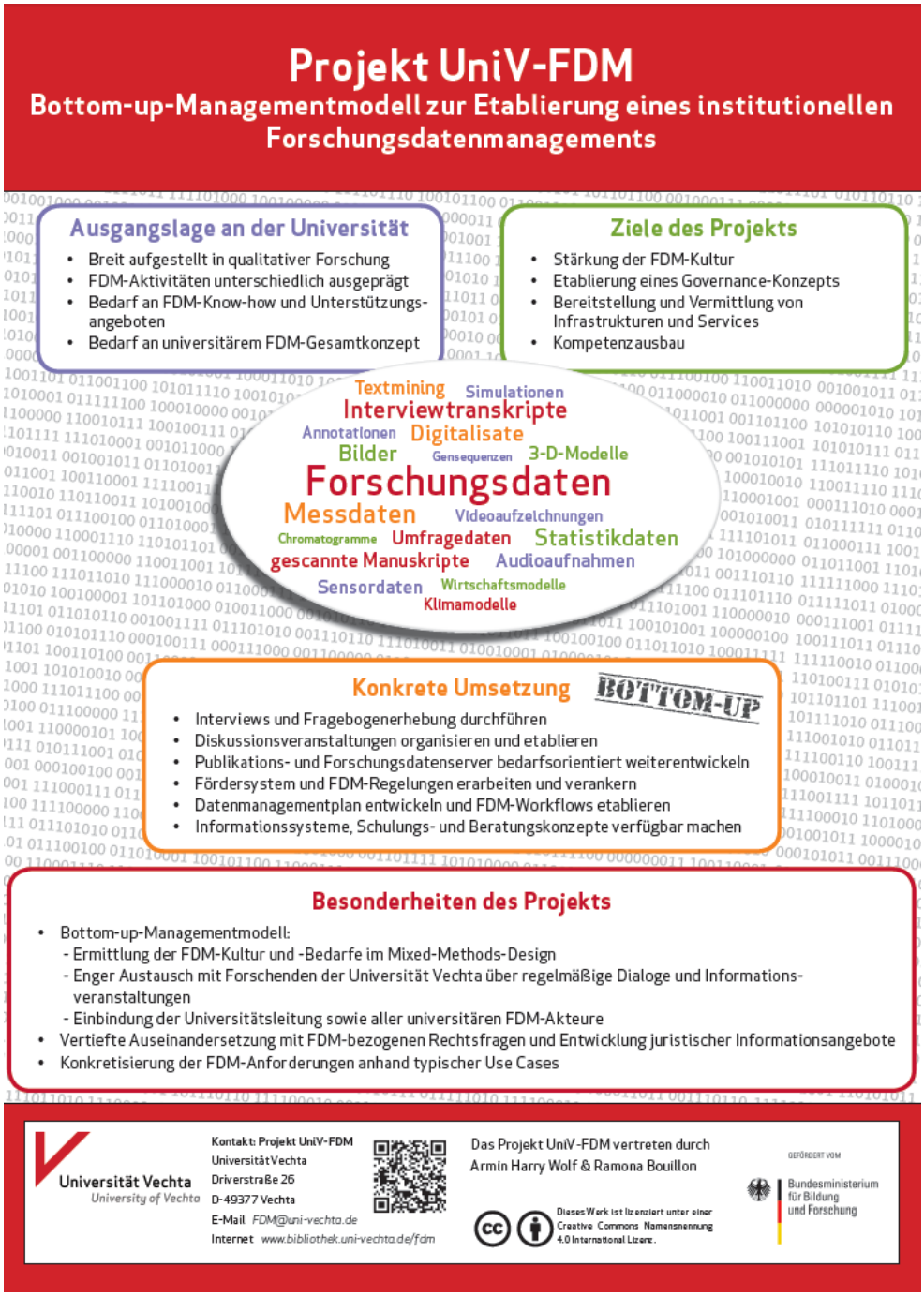
Projekt UniV-FDM
Bottom-up-Managementmodell zur Etablierung eines institutionellen Forschungsdatenmanagements

Armin Harry Wolf

Universität Vechta

Abstract

Das Poster stellt das Projekt UniV-FDM der Universität Vechta vor, in dessen Rahmen basierend auf dem Fachwissen der Forschenden an der Universität Vechta und unter Einbezug ihrer Bedarfe ein fächerübergreifendes, institutionelles Forschungsdatenmanagement (FDM) etabliert werden soll, das sich an fachspezifischen sowie nationalen und internationalen Standards orientiert. Es wird zunächst die Ausgangslage an der Universität Vechta dargestellt, aus der sich die Projektziele für die erfolgreiche Etablierung eines institutionellen FDM ergeben. Des Weiteren werden die Maßnahmen für die konkrete Umsetzung dieser Ziele mithilfe eines Bottom-up-Ansatzes sowie die Besonderheiten des Projekts aufgezeigt.


Projekt UniV-FDM
Bottom-up-Managementmodell zur Etablierung eines institutionellen Forschungsdatenmanagements
Ausgangslage an der Universität
Breit aufgestellt in qualitativer Forschung
FDM-Aktivitäten unterschiedlich ausgeprägt
Bedarf an FDM-Know-how und Unterstützungsangeboten
Bedarf an universitärem FDM-Gesamtkonzept
Ziele des Projekts
Stärkung der FDM-Kultur
Etablierung eines Governance-Konzepts
Bereitstellung und Vermittlung von Infrastrukturen und Services
Kompetenzausbau
Textmining
Simulationen
Interviewtranskripte
Annotationen
Digitalisate
Bilder
Gensequenzen
3-D-Modelle
Forschungsdaten
Messdaten
Videoaufzeichnungen
Chromatogramme
Umfragedaten
Statistikdaten
gescannte Manuskripte
Audioaufnahmen
Sensordaten
Wirtschaftsmodelle
Klimamodelle
Konkrete Umsetzung
BOTTOM-UP
Interviews und Fragebogenerhebung durchführen
Diskussionsveranstaltungen organisieren und etablieren
Publikations- und Forschungsdatenserver bedarfsorientiert weiterentwickeln
Fördersystem und FDM-Regelungen erarbeiten und verankern
Datenmanagementplan entwickeln und FDM-Workflows etablieren
Informationssysteme, Schulungs- und Beratungskonzepte verfügbar machen
Besonderheiten des Projekts
Bottom-up-Managementmodell:
- Ermittlung der FDM-Kultur und -Bedarfe im Mixed-Methods-Design
- Enger Austausch mit Forschenden der Universität Vechta über regelmäßige Dialoge und Informationsveranstaltungen
- Einbindung der Universitätsleitung sowie aller universitären FDM-Akteure
Vertiefte Auseinandersetzung mit FDM-bezogenen Rechtsfragen und Entwicklung juristischer Informationsangebote
Konkretisierung der FDM-Anforderungen anhand typischer Use Cases
Universität Vechta
University of Vechta
Kontakt: Projekt UniV-FDM
Universität Vechta
Driverstraße 26
D-49377 Vechta
E-Mail FDM@uni-vechta.de
Internet www.bibliothek.uni-vechta.de/fdm
Das Projekt UniV-FDM vertreten durch
Armin Harry Wolf & Ramona Bouillon
Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Institutionelles Forschungsdatenmanagement an der Universität zu Köln als gemeinschaftliche Herausforderung

Jens Dierkes, Constanze Curdt, Sonja Kloppenburg

Universität Köln

Abstract

Universitätsweite Forschungsdateninfrastrukturen werden in der Regel von zentralen Einrichtungen wie Universitätsbibliotheken und Rechenzentren aufgebaut und unterhalten. An deutschen Hochschulen gibt es vielfach auch Ansätze das zentrale Forschungsmanagement und/oder die Leitung der Hochschule zu beteiligen. Aktuelle Erfahrungen mit mehreren Forschungsdatenmanagement (FDM)-Initiativen auf Universitätsebene zeigen, dass eine koordinierte Vernetzung auf dem Campus dazu beiträgt, eine lokale „Community of Practice“ zum FDM aufzubauen und „Good Practices“ in den Forschungsalltag zu integrieren. Neben der Vernetzung bleiben entscheidende Fragen, welches Kooperationsmodell und welcher Grad der Zentralisierung oder Dezentralisierung gewählt werden sollen. Es zeigen sich dabei deutliche Unterschiede zwischen einzelnen Hochschulstandorten.

Die Universität zu Köln (UzK) weist historisch gewachsene dezentrale Strukturen auf. Neben den zentralen Infrastrukturen wie Bibliothek und Rechenzentrum gibt es bereits eine Reihe von FDM-Aktivitäten sowohl auf Fakultätsebene (z.B. Datenzentrum an der Philosophischen Fakultät) als auch auf Forschungsverbundebene (z.B. innerhalb von Sonderforschungsbereichen, Exzellenzclustern etc.).

Die UzK hat 2017 eine Leitlinie zum Umfang mit Forschungsdaten verabschiedet und kürzlich das Cologne Competence Centre for Research Data Management (C3RDM) gegründet. Das C3RDM wird durch die zentralen Einrichtungen Dezernat Forschungsmanagement, das Regionale Rechenzentrum und die Universitäts- und Stadtbibliothek gebildet. Ziele sind die Bündelung der FDM-Angebote für den Campus und die Bildung eines Netzwerkes von e-Research-Expert*innen. Neben den klassischen FDM-Themen, sehen wir die Etablierung eines Dialogs vorrangig zwischen den lokalen Stakeholdern aus Forschung, Lehre, Verwaltung und Informationsinfrastrukturen. Dieser soll die bedarfsorientierte aber auch ressourcenschonende Entwicklung einer Datenkultur und -infrastruktur an der UzK ermöglichen und stärken.

Wir berichten über erste Schritte unserer Arbeit auf dem Campus. Hierbei liegt der Fokus auf der Ein- und Anbindung der FDM-Praktiken an der UzK und der Entwicklung einer Struktur der Zusammenarbeit zwischen lose gekoppelten Akteuren der Informationsinfrastruktur. Es hat sich deutlich gezeigt, dass FDM eine Team-Aktivität ist. Dazu gehören u. a. die Etablierung des C3RDM als Ansprechpartner für die Beratung zum FDM bei Drittmittelanträgen, der etablierte Jour fixe mit dem Data Center for the Humanities und die zentral organisierte Übertragung des DH-NRW Jour fixe.

Universität zu Köln

Cologne Competence Center for Research Data Management (C3RDM)

Institutionelles Forschungsdatenmanagement an der Universität zu Köln als gemeinschaftliche Herausforderung

C. Curdt [2], J. Dierkes [3], R. Depping [3], S. Kloppenburg [1], M. Linne [3], M. Riese [1], M. Röder [1], J. Schenk [1], M. Valencia-S. [2], V. Winkelmann [2]

[1] Dezernat Forschungsmanagement, [2] Regionales Rechenzentrum, [3] Universitäts- und Stadtbibliothek

Einleitung

Bestandteil aktueller FDM-Initiativen in deutschen Hochschulen sind u.a.

- Informationsinfrastrukturanbieter (Rechenzentrum, Bibliothek)
- Abteilung Forschungsmanagement, Forschungsförderung
- Universitätsleitungen, FDM-Stabsstellen
- FDM-Experten von Fakultäten, Projekten, …

Koordinierte Vernetzung auf dem Campus trägt zum Aufbau einer lokalen „Community of Practice" zum FDM bei und hilft „Good Practices" in den Forschungsalltag zu integrieren. Neben der Vernetzung bleiben entscheidende Fragen offen: (1) welches Kooperationsmodell und (2) welcher Grad der De-/Zentralisierung mit vielen Stakeholdern gewählt werden sollte. Hierbei wird es deutliche Unterschiede zwischen den einzelnen Hochschulstandorten geben.

Aufbau FDM-Kompetenzzentrum & -netzwerk

- 2017 Verabschiedung der **Leitlinie zum Umfang mit Forschungsdaten**
- 2018 Gründung des **Cologne Competence Center for Research Data Management (C3RDM)** durch die Universitäts- u. Stadtbibliothek, das Regionale Rechenzentrum und das Dezernat Forschungsmanagement
- **Ziel 1:** Aufbau von umfangreicher FDM-Unterstützung und Beratung in allen Phasen des Forschungsvorhabens
- **Ziel 2:** Aufbau eines FDM-Expert*innen-Netzwerks und Forum zur Vernetzung der gewachsenen FDM-Strukturen bzw. verschiedener Stakeholder an der UzK

Abb 2. Aufbau einer FDM-Infrastruktur und Expert*innen-Netzwerk an der UzK

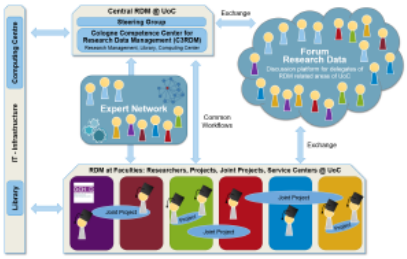
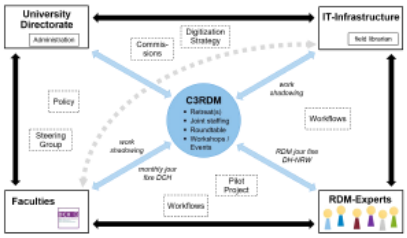


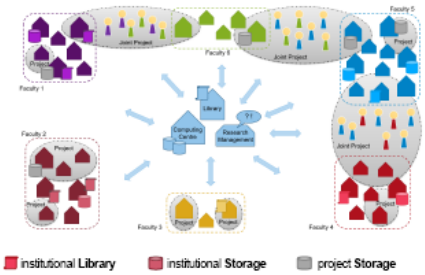
Abb 3. Vernetzung zwischen den FDM-Stakeholdern der UzK



Universität zu Köln (UzK)

- forschungsorientierte Universität, Erstgründung 1388
- **Key Facts - Personal:** ~ 50.000 Studierende, ~ 650 Professor*innen, ~ 5.800 wissenschaftliche Mitarbeiter*innen, ~ 6.000 technische und administrative Mitarbeiter*innen
- **Key Facts - Forschung:** u.a. 4 Exzellenz Cluster, 13 Sonderforschungsbereiche/Transregios (inklusive 4 INF-Projekte), 20 ERC Grants, 7 Forschungsgruppen, 34 Graduiertenschulen, 18 Stiftungsprofessuren, 22 An-Institute
- **breites Spektrum an Disziplinen::** u.a. Geistes-, Human-, Rechts-, Lebens-, Natur-, Wirtschafts- und Sozialwissenschaften, Medizin und Mathematik, organisiert in **sechs Fakultäten**
- historisch gewachsene, starke dezentrale Strukturen (z.B. zweistufige Bibliotheksstruktur)
- **dezentrale FDM-Aktivitäten** gibt es auf Fakultätsebene (z.B. Data Center for the Humanities der Philosophischen Fakultät) und Forschungsverbundebene (z.B. Exzellenzcluster, SFBs, etc.)

Abb 1. Forschungslandschaft an der UzK und de-/zentrale (FDM-)Strukturen



Zusammenfassung

- institutionelles FDM ist eine gemeinschaftliche Herausforderung in einem komplexen sozio-technologischen Umfeld mit vielen Stakeholdern
- Berücksichtigung und Anerkennung gewachsener, bestehender FDM-Strukturen der UzK (zentral und dezentral)
- neben generischen Diensten sind die Vernetzung und Etablierung von Kooperationen mit den FDM-Aktivitäten in den einzelnen Fachbereichen zentrale Anliegen des Aufbauprojektes des C3RDM
- dialogische Vorgehensweise zusammen mit Pilotprojekten
- vorgeschlagene Netzwerk-/Kooperationsstruktur ist der erste Schritt zur Herstellung eines breiteren Konsenses über die Anforderungen an eine nachhaltige FDM-Infrastruktur

Referenzen

Dierkes, J., und Curdt, C. (2018): Von der Idee zum Konzept – Forschungsdatenmanagement an der Universität zu Köln, O-bib 5 (2), 28-46.

University of Cologne (UoC) (2018): Leitlinie zum Umgang mit Forschungsdaten an der Universität zu Köln. Amtliche Mitteilungen 07/2018.

Kontakt: Cologne Competence Center for Research Data Management (C3RDM) | **E-Mail:** fdm-support@uni-koeln.de | **Web:** http://fdm.uni-koeln.de

re3data - Offene Infrastruktur für Open Science

Maxi Kindling[1], Heinz Pampel[2], Robert Ulrich[3], Paul Vierkant[4], Frank Scholze[5], Martin Fenner[6], Michael Witt[7], Kirsten Elger[8]

Abstract
re3data (https://www.re3data.org) ist das globale Verzeichnis von Forschungsdatenrepositorien (Pampel et al. 2013). Im Dezember 2018 weist der Dienst über 2.240 digitale Repositorien für Forschungsdaten auf Basis eines umfassendes Metadatenschemas (Rücknagel et al.2015) nach. Eine Vielzahl von Förderorganisationen, Verlagen und wissenschaftlichen Einrichtungen aus der ganzen Welt empfehlen die Nutzung des Dienstes zur Identifikation geeigneter Repositorien in Leit- und Richtlinien zum Forschungsdatenmanagement. Betrieben wird der Dienst unter dem Dach von DataCite an der Bibliothek des Karlsruher Instituts für Technologie (KIT) im Dialog mit dem Helmholtz Open Science Koordinationsbüro am Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ, der Bibliothek der Purdue University, USA und dem Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin. Das Poster stellt den aktuellen Stand von re3data vor und widmet sich den aktuellen Weiterentwicklungen des Dienstes im Kontext aktueller und zukünftiger Anforderungen. Dabei stehen die Offenheit des Dienstes, seiner Daten sowie seiner Schnittstellen im Fokus
(Dasler, 2018, Witt et al., 2019).

Literatur

Dasler, R. (2o18). Data sharing made easier: use Repository Finder to find the right repository for your data. DataCite Blog, 19.12.2018. DOI: https://doi.org/10.5438/wday-8958.

Pampel, H. et al. (2013). Making Research Data Repositories Visible: The re3data.org Registry. PLOS ONE, 8(11), e78080. DOI: https://doi.org/10.1371/journal.pone.0078080.

Rücknagel, J. et al. (2015). Metadata Schema for the Description of Research Data Repositories. Version 3.0. DOI: https://doi.org/10.2312/re3.008.

Witt, M. et al. (2019). Connecting Researchers to Data Repositories in the Earth, Space, and Environmental Sciences. Libraries Faculty and Staff Scholarship and Research. Paper 209. https://docs.lib.purdue.edu/lib_fsdocs/209.

[1] https://orcid.org/0000-0002-0167-0466, maxi.kindling@hu-berlin.de
[2] https://orcid.org/0000-0003-3334-2771, heinz.pampel@os.helmholtz.de
[3] https://orcid.org/0000-0001-9063-2703, robert.ulrich@kit.edu
[4] https://orcid.org/0000-0003-4448-3844, paul.vierkant@os.helmholtz.de
[5] https://orcid.org/0000-0003-3404-1452, frank.scholze@kit.edu
[6] https://orcid.org/0000-0003-1419-2405, martin.fenner@datacite.org
[7] https://orcid.org/0000-0003-4221-7956, mwitt@purdue.edu
[8] https://orcid.org/0000-0001-5140-8602, kirsten.elger@gfz-potsdam.de

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

lists

2332 research data repositories

from

75 countries

covering

74 subjects from acoustics to zoology

COREF Community Driven Open Reference for Research Data Repositories

is a proposed project that seeks to **improve**, **extend** and **update re3data**

Use unique identification in machine-to-machine (M2M) communication to foster the usage of re3data as an open reference for repositories.

Import and export of metadata fragments, controlled vocabularies and database entries of external sources.

Adress funding bodies and publishers to recommend re3data in their respective policies to foster the usage of trustworthy repositories complying with the FAIR-principles.

PARTNERS

HELMHOLTZ Open Science, GFZ, KIT, DataCite, PURDUE UNIVERSITY, FAIRSFAIR, DFG

AUTHORS

Sponsoren

und

Aussteller

Wir danken den Ausstellern und Sponsoren

Hauptsponsor der WissKom 2019

Ausstellende Firmen:

Concat

De Gruyter

EBSCO Information Services

Elsevier

F1000

GALE - Cengage Company

IET

Programmfabrik

SAM – Standards And More

Schweitzer Fachinformationen / Goethe + Schweitzer GmbH

Springer Nature

UKW Innenarchitekten

Firmenpräsentationen im Ausstellerforum:

EBSCO Information Services

Elsevier

F1000

IET

Programmfabrik

SAM – Standards And More

Springer Nature

UKW Innenarchitekten

Die Konferenz wird gesponsert von:
Digital Science, b.i.t. Verlag, IET, Springer Nature, UKW Innenarchitekten

Schriften des Forschungszentrums Jülich
Reihe Bibliothek / Library

Band / Volume 18
WissKom 2007: Wissenschaftskommunikation der Zukunft
4. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,
6. – 8. November 2007, Beiträge und Poster
hrsg. von R. Ball (2007), 300 pp
ISBN: 978-3-89336-459-6

Band / Volume 19
Bibliometrische Verfahren und Methoden als Beitrag zu Trendbeobachtung und -erkennung in den Naturwissenschaften
von D. Tunger (2009), 311 pp
ISBN: 978-3-89336-550-0

Band / Volume 20
WissKom 2010: eLibrary – den Wandel gestalten
5. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,
8. – 10. November 2010, Proceedingsband
hrsg. von B. Mittermaier (2010), 389 pp
ISBN: 978-3-89336-668-2

Band / Volume 21
WissKom 2012: Vernetztes Wissen – Daten, Menschen, Systeme
6. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,
5. – 7. November 2012, Proceedingsband
hrsg. von B. Mittermaier (2012), 379 pp
ISBN: 978-3-89336-821-1

Band / Volume 22
WissKom 2016: Der Schritt zurück als Schritt nach vorn – Macht der Siegeszug des Open Access Bibliotheken arbeitslos?
7. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,
14. – 16. Juni 2016, Proceedingsband
hrsg. von B. Mittermaier (2016), ca. 270 pp
ISBN: 978-3-95806-146-0

Band / Volume 23
WissKom 2019: Forschungsdaten – Sammeln, sichern, strukturieren
8. Konferenz der Zentralbibliothek, Forschungszentrum Jülich,
4. – 6. Juni 2019, Proceedingsband

Bibliothek / Library
Band / Volume 23
ISBN 978-3-95806-405-8

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
Forschungszentrum