

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 erarbeiten¹. 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 Alliance², 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 (Rfll 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 (Rfll 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://eoscipilot.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](https://www.forschungsdaten.org) gelistet³. 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 Dariah⁴ und in der Biologie GFBio⁵. 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 Forschungsdaten⁷, GOFAIR⁸ 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/arbeitschwerpunkte/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 singularär 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 Watson⁹ 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: Addressing Disciplinary Data Management Needs. In: International Journal of Digital Curation 13 (1), S. 141–149. DOI: 10.2218/ijdc.v13i1.604.