

Metadata is a Love Note to the Future^{1, 2}

Irina Sens

Einleitung

In den „FAIR Guiding Principles for scientific data management and stewardship“³ spielen Metadaten eine entscheidende Rolle, um Forschungsdaten effizient und effektiv nutzen zu können. Eine hohe und standardisierte Qualität bildet die Grundlage für Mehrwertdienste. Die Erschließung von Forschungsdaten rückt damit in den Fokus bibliothekarischer Dienstleistungen. Mit der Entwicklung hin zu datengetriebener Wissenschaft wird Forschungsdatenmanagement zum Bestandteil von Informationsinfrastrukturen wie beispielsweise den Fachinformationsdiensten (FIDs).

FID und Forschungsdatenmanagement

Für die FIDs hat die Deutsche Forschungsgemeinschaft (DFG) klare Ziele definiert:

- Auf- und Ausbau überregionaler, leistungsfähiger und bedarfsorientierter Informationsinfrastrukturen für die Wissenschaft
- Am Spezialbedarf der wissenschaftlichen Fächer orientierte standortunabhängige Informationsversorgung
- Ergänzung der Angebote der lokalen Informationsinfrastrukturen der Hochschulen und Forschungseinrichtungen
- Interessen der Forschung in den jeweiligen Fächern stehen im Mittelpunkt
- Abgrenzung von den Grundaufgaben wissenschaftlicher Bibliotheken und Informationseinrichtungen
- Mehrwert gegenüber lokalen Informationsangeboten, die sehr deutlich die Bedarfe der jeweiligen Fachcommunity in den Mittelpunkt stellen und aufbauend auf lokalen Angeboten der jeweiligen Hochschulen aufsetzen.

Daraus leitet sich ab, dass FID aufgrund dieser Zielsetzung auch Aufgaben im Forschungsdatenmanagement zu leisten haben. Bereits 2012 wies Eefke Smit in ihrem Beitrag „Avoiding a Digital Dark Age for Data: why data and publications belong together“ in dem ICSTI-Workshop „Delivering Data in Science“ am 5. März 2012 in Paris darauf hin, dass fehlende Infrastrukturen wie Datenzentren und Repositorien zu einer Nicht-Sichtbarkeit und einem Verlust von Forschungsdaten führen. Zum damaligen Zeitpunkt wurden circa 75 Prozent der Forschungsdaten ausschließ-

Abstract

Metadaten spielen eine entscheidende Rolle, damit Initiativen wie die Nationale Forschungsdateninfrastruktur ihre Wirksamkeit entfalten können. Dieser Beitrag betrachtet den aktuellen Stand der Metadatenqualität und die Notwendigkeit einer verbesserten Metadatenpraxis für die Auffindbarkeit und Nachnutzbarkeit von Forschungsdaten. Dabei werden die Herausforderungen hinsichtlich der Qualität und Standardisierung von Metadaten sowie die Rolle der Fachinformationsdienste (FID) bei der Bereitstellung qualitativ hochwertiger Metadaten für die Forschung diskutiert.

Metadata play a central role in the effectiveness of initiatives such as the National Research Data Infrastructure. This article describes the current state of metadata quality and the need for improved metadata practices for the findability and reusability of research data. It discusses the challenges of metadata quality and standardisation and the role of subject information services in providing high quality metadata for research.

lich auf privaten oder institutionellen Festplatten veröffentlicht.

Forschungsdaten – zwei Definitionen

Für Forschungsdaten existieren zahlreiche Definitionen, von denen zwei hier als Grundlage dienen sollen:

1. „Das Material, das in der wissenschaftlichen Gemeinschaft allgemein als notwendig für die Validierung von Forschungsergebnissen akzeptiert wird“ – U.S. Office of Management and Budget, <https://guides.lib.vt.edu/NSF-DMP/basics>
2. „Zu Forschungsdaten zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden“ – DFG, <https://www.dfg.de/resource/blob/172112/4ea861510ea369157afb499e96fb359a/leitlinien-forschungsdaten-data.pdf>.

Zu den Forschungsdaten gehören keine „vorläufigen Analysen“, Entwürfe wissenschaftlicher Arbeiten, Pläne für zukünftige Forschung, Peer Reviews oder Kommunikation mit Kolleg:innen. Nicht unberücksichtigt bleiben sollten auch Daten, die zu keinem Ergebnis geführt haben, sogenannte „Negativ-Daten“.

1 Basierend auf dem Vortrag „Und wohin mit den Forschungsdaten – FID, Repositories und Forschungsdaten“ auf der 111. BiblioCon am 23.5.2023 in der Session „FID-Netzwerk“

2 Jason Scott prägte diesen Ausdruck.

3 Publiziert 2016 in Scientific Data, <https://www.nature.com/articles/sdata201618>






 <p>Spezialisten-Daten</p> <p> Freier Zugriff</p> <p>FZI Forschungszentrum Informatik mCLOUD 2022</p> <p>Im Rahmen des Projektes OCROSS wurde durch Spezialisten mit einem Messfahrzeug der Straßenraum erfasst. Ein Auszug der Daten (A6/Sinsheim ...</p>	<p>Forschungsdaten</p>
 <p>Durchblick im Datensteckverbinder-Dschungel Jeder Datentyp und jeder Bus stellen unterschiedliche Anforderungen</p> <p>British Library Online Contents 2009</p>	<p>Aufsatz</p>
 <p>Daten VVS</p> <p> Freier Zugriff</p> <p>MobiData BW mCLOUD 2020</p> <p>Soll-Fahrplandaten (GTFS), Haltestellen und Linienverläufe des</p>	

Abb. 1: Suchergebnis im Portal des FID move, <https://www.fid-move.de> mit Daten und Texten

Dass Forschungsdaten eine immer zentralere Rolle spielen, wird durch die Gründung der Nationalen Forschungsdateninfrastruktur (NFDI) – Förderung der ersten Konsortien seit 2020 – sowie durch die Implementierung der European Open Science Cloud seit 2018 manifestiert. Bereits 2010 erschien der Report „Riding the Wave: How Europe can gain from the rising tide of scientific data“. Er ist das Ergebnis eines sechsmonatigen intensiven Brainstormings und verschiedener Konsultationen mit Expert:innen auf der ganzen Welt, um eine „Vision 2030“ für E-Infrastrukturen für wissenschaftliche Daten zu entwickeln. Der Bericht beschreibt langfristige Szenarien und die damit verbundenen Herausforderungen in Bezug auf den Zugang zu wissenschaftlichen Daten, deren Kuratierung und Bewahrung sowie die zur Verwirklichung der Vision erforderlichen Strategien und Maßnahmen.

Insbesondere mit der Gründung der Nationalen Forschungsdateninfrastruktur stellt sich für die FID die Frage, wie sie sich hier positionieren können und welche Dienste in Zusammenarbeit mit den NFDI-Konsortien unter dem Schlagwort „Synergetische Komplementarität“ den Forschenden Mehrwerte bringen. Dabei ist zu beachten, dass Forschungsdateninfrastrukturen der Dezentralität und Heterogenität der Forschung Rechnung tragen müssen. Interoperabilität zwischen Dateninfrastrukturen ist eine notwendige Voraussetzung.

FAIRe Forschungsdaten

Um dies zu erreichen, sind die Handlungsfelder um FAIRe Daten zu betrachten. FAIR steht dabei für Findable (auffindbar), Accessible (zugänglich), Interoperable (interoperabel) und Reusable (wiederverwendbar). Ziel ist es, dass

Forschungsdaten für Menschen und Maschinen optimal aufbereitet und zugänglich sind.

Dabei geht es unter anderem um:

- öffentliche Zugänglichmachung
- Zitierfähigkeit – zum Beispiel via DOI (Digital Object Identifier)
- Lizenzvergabe – idealerweise CC0/CC-BY
- Gewährleistung der Langzeitverfügbarkeit
- Angebot eines Fachrepositoriums, das alle o.g. Eigenschaften erfüllt.

Wo positionieren sich hier die Fachinformationsdienste? Mit ihren Fachportalen bieten sie ideale Voraussetzungen, um Informationsressourcen – Texte, Daten, AV-Materialien – nachzuweisen. Stellvertretend sei hier das Beispiel des FID move (Mobilitäts- und Verkehrsforschung) (Abb. 1) gezeigt. Notwendige Voraussetzung, um Forschungsdaten auffindbar zu machen, sind Metadaten, die aus ungenutzten, unverknüpften, unsichtbaren und nutzend nur für den Einzelnen, strukturierte Kollektionen, verknüpfte, findbare und nachnutzbare Forschungsressourcen machen.

FAIRe Metadaten

FAIR müssen also auch die Metadaten sein. Dass es hier noch einen erheblichen Nachholbedarf gibt, zeigt sich auf den ersten Blick, wenn man die von der TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften entwickelte Suchanwendung RESODATE – <https://resodate.org> – benutzt. Diese Suchanwendung ermöglicht es, Research Software, Data und Terminologies zu finden.

Die unterschiedliche Verwendung des Fachgebiets als Metadatum erlaubt es nicht, zu einem Fachgebiet, die Er-

Abb. 2:
Suchanwendung
RESODATE,
Filter: Fachgebiet

gebnisse zusammenzuführen, insbesondere, wenn Fachgebiet überhaupt nicht vergeben wird (s. Abb. 2).

Ausführlicher analysiert wurde dies im Beitrag⁴ „Wie FAIR sind unsere Metadaten? Eine Analyse der Metadaten in den Repositorien des TIB-DOI-Services“.

Verwendet werden seitens der Betreiber von Forschungsdatenrepositorien 51 Prozent der Metadatenfelder aus der Kategorie „Findable Essential“. Hier sind allerdings die Pflichtfelder aus dem DataCite-Metadatenchema enthalten und nur 18 Prozent der Metadatenfelder aus der Kategorie „Findable Support“. Eine Analyse der Repositorien deutscher DataCite-Datenzentren vom 17. Februar 2023 ergab, dass 88 Prozent der untersuchten 65 Repository-Konten Schlagworte vergeben, allerdings nur 54 Prozent der 6.500 Datensätze mindestens ein Schlagwort haben und in nur 464 Fällen (2 Prozent) das OECD Field Schema angewendet wurde. Hier besteht offensichtlich erheblicher Nachholbedarf.

Mit besseren Metadaten Forschungsdaten sichtbarer machen

Daraus leiten sich Aufgaben für die FID ab, um die Find- und Sichtbarkeit der Daten zu ermöglichen:

- Gewährleistung qualitativ hochwertiger Metadaten (formal und inhaltlich), um Forschungsressourcen zu vernetzen
- aktive Beteiligung an der Entwicklung internationaler fachlich anerkannter Metadaten schemata
- Vernetzung mit NFDI-Konsortien und Betreibern institutioneller Repositorien mit dem Ziel der Standardisierung.

Damit können die FID entscheidend und wirksam zur von Vincent S. Smith in seinem Beitrag „Data publication: towards a database of everything, BMC Research Notes 2009, 2, 113“ formulierten Zielsetzung beitragen:

“Make it citable – motivating data publication through peer recognition

Make it useful – moving beyond data archival”

Fazit

Die Investition in gute Metadaten wird sich für die Forschung auszahlen. Heterogene Daten aus verschiedenen Domänen können effektiv (auch in NFDIs und EOSC) integriert werden. Verschiedene Akteur:innen können agil und effizient zusammenarbeiten, ohne in eine zentrale Plattform gezwungen zu werden („cooperate on standards, compete on implementations“). Es werden sich neue Perspektiven für die Wissenschaften ergeben: automatisierte Hypothesengenerierung, Maschinelles Lernen, Open Science und vieles mehr. Auch im geplanten Forschungsdatengesetz spielen Metadaten eine wichtige Rolle: Eine der zentralen Maßnahmen könnte in der Schaffung von Metadatenkatalogen für öffentliche Forschungseinrichtungen bestehen. Diese Kataloge sollen die Auffindbarkeit von Forschungsdaten signifikant verbessern. **I**



Dr. Irina Sens

ist seit 1999 stellvertretende Direktorin der TIB – Leibniz-Informationzentrum Technik und Naturwissenschaften und Universitätsbibliothek, wo sie den Bibliotheksbetrieb sowie den Programmbereich „Bestandsentwicklung und Metadaten“ leitet. Nach dem Studium (Chemie und Mathematik) sowie der Promotion in Chemie in Marburg absolvierte sie ein Bibliotheksreferendariat. Sie ist in verschiedenen Gremien zum Forschungsdatenmanagement aktiv.

irina.sens@tib.eu

<https://orcid.org/0000-0001-9190-8628>

4 Burger, M., Cordts, A., & Habermann, T. (2021). Wie FAIR sind unsere Metadaten? Eine Analyse der Metadaten in den Repositorien des TIB-DOI-Services. Bausteine Forschungsdatenmanagement, (3), 1-13. <https://doi.org/10.17192/bfdm.2021.3.8351>