

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

Das muss noch in Wikidata rein

Lukas C. Bossertⁱ

Magdalene Cyraⁱⁱ

Évariste Demandtⁱⁱⁱ

Matthias Fingerhuth^{iv}

Ceren Yildiz^v

2023

Zitiervorschlag

Bossert, Lukas C., Cyra, Magdalene, Demandt, Évariste, Fingerhuth, Matthias und Ceren Yildiz.
2023. Das muss noch in Wikidata rein. *Bausteine Forschungsdatenmanagement.
Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und
-managern* Nr. 5/2023: S. 2-18. DOI: [10.17192/bfdm.2023.5.8580](https://doi.org/10.17192/bfdm.2023.5.8580).

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

ⁱRWTH Aachen. ORCID: [0000-0003-3076-3968](https://orcid.org/0000-0003-3076-3968)

ⁱⁱLandesinitiative für Forschungsdatenmanagement – fdm.nrw. ORCID: [0000-0001-7738-2703](https://orcid.org/0000-0001-7738-2703)

ⁱⁱⁱRWTH Aachen. ORCID: [0000-0002-5705-0071](https://orcid.org/0000-0002-5705-0071)

^{iv}Landesinitiative für Forschungsdatenmanagement – fdm.nrw. ORCID: [0000-0002-0248-8914](https://orcid.org/0000-0002-0248-8914)

^vLandesinitiative für Forschungsdatenmanagement – fdm.nrw. ORCID: [0009-0007-6160-8836](https://orcid.org/0009-0007-6160-8836)

Abstract

Mit wachsender Zahl der Akteure wird die Landschaft des Forschungsdatenmanagements (FDM) zunehmend unübersichtlich. Obwohl es einige einschlägige Ressourcen und Anlaufstellen gibt, sind Informationen zu FDM-Strukturen bisher zumeist verstreut und unstrukturiert. Wikidata bietet eine Infrastruktur, in der diese Informationen als Linked Open Data strukturiert gespeichert werden können. Der Beitrag illustriert die Möglichkeiten, wie man Einträge in Wikidata anlegt, editiert und auswertet, am Beispiel von FDM-Servicestellen und den Konsortien der Nationalen Forschungsdateninfrastruktur.

1 Problemstellung und Ziel

Das Handlungsfeld Forschungsdatenmanagement (FDM) erfährt seit einigen Jahren ein stetiges Wachstum. An Universitäten sind Servicestellen oder Kontaktpunkte für FDM inzwischen die Regel, auch an Hochschulen für Angewandte Wissenschaften oder außeruniversitären Forschungseinrichtungen etablieren sich solche Stellen zusehends. Begleitet wird diese Entwicklung von der wachsenden Betonung der Bedeutung von Forschungsdaten im wissenschaftspolitischen Diskurs bzw. vom ausgelobten Ziel, sie FAIR zu machen.¹ Damit geht eine gezielte Förderung des Handlungsfelds einher, etwa durch Ausschreibungen des Bundesministeriums für Bildung und Forschung (BMBF), aber auch durch DFG-Projekte im Bereich wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS), Informationsinfrastruktur-Projekte (INF) in Sonderforschungsbereichen (SFB) und die Nationale Forschungsdateninfrastruktur (NFDI). Dies führt zu einer wachsenden Zahl involvierter Akteure und folglich einer komplexeren FDM-Landschaft. Auf persönlichem Austausch basierende Netzwerke können diese Komplexität nur noch bedingt erfassen. Zwar bestehen verschiedene Einstiegsmöglichkeiten und Quellen, um strukturelle Informationen über FDM – z. B. Kontaktstellen an Einrichtungen, Drittmittelprojekte, eingesetzte Software oder angebotene Services – zu erhalten. Speziell – aber nicht nur – für Neueinsteigende ist dies eine Herausforderung.

Bisher sind Informationen zu Akteuren und Initiativen verstreut. In GEPRIS² sind Informationen zu Projekten, die von der Deutschen Forschungsgesellschaft (DFG) gefördert werden, hinterlegt; vom BMBF geförderte Projektinformationen sind im Förderkatalog des Bundes³ einzusehen. Hinzu kommen Informationsressourcen wie forschungsdaten.org

¹Wilkinson, Mark D., Dumontier, Michel, Aalbersberg, IJsbrand Jan, Appleton, Gabrielle, Axton, Myles, Baak, Arie, Blomberg, Niklas, Boiten, Jan-Willem, Bonino da Silva Santos, Luiz, Bourne, Philip E. et al. „The FAIR guiding principles for scientific data management and stewardship“. *Scientific Data* 3, Nr. 1 (2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.

²<https://gepris.dfg.de/gepris/OCTOPUS> (zugegriffen 13. Mai 2023).

³<https://foerderportal.bund.de/foekat/jsp/SucheAction.do?actionMode=searchmask> (zugegriffen 13. Mai 2023).

oder forschungsdaten.info, die aus der FDM-Praxis heraus Informationen aufbereiten. Sie können als Indizien dafür gewertet werden, dass ein Bedarf besteht, weitergehende Informationen einzusehen. Einen Überblick über die Angebote und Aktivitäten an einzelnen Standorten oder in bestimmten Bundesländern zu bewahren, stellt eine Herausforderung für Interessierte wie auch Involvierte dar. In diesem Beitrag skizzieren wir, wie Wikidata⁴ eingesetzt werden kann, um eine öffentliche Informationsbasis zu Einrichtungen und Projekten im Bereich FDM zu pflegen. Dazu wird zunächst ein Überblick darüber gegeben, wie Informationen zu FDM-Strukturen bislang nur verteilt zugänglich sind. Dabei werden auch Herausforderungen der Pflege und Nutzung von Informationen zu FDM illustriert. Darauf basierend illustrieren wir anhand von zwei Beispielen, wie der Einsatz von Linked Open Data, umgesetzt in Form von Wikidata, dazu beitragen kann, diese verteilten Informationen strukturiert zusammenzuführen, was einen zentralen Zugang sowie eine gemeinschaftliche Pflege der Datenbasis ermöglicht. Die anschließende Diskussion bemüht sich um eine Abwägung der Potentiale und Risiken dieses Ansatzes, dem sich ein abschließendes Fazit und Ausblick anschließt.

2 Informationsquellen zu FDM-Strukturen – Status Quo

Gegenwärtig existiert eine Vielzahl von Ressourcen, in denen Informationen zu FDM-Strukturen verfügbar sind, die sich aber stark voneinander unterscheiden.

Das Informationsportal forschungsdaten.info wird durch ein im deutschsprachigen Raum verteiltes Redaktionsteam getragen. Dort sind neben praktischen Informationen, die der Orientierung im Thema dienen sollen, auch Informationen zu strukturellen Aspekten wie FDM-Projekten oder regionalen Aktivitäten bereitgestellt. Es gibt Informationen zur NFDI oder zu FDM-Initiativen und -Projekten in den Bundesländern. Die redaktionelle Betreuung hält die Daten dort aktuell und strukturiert, allerdings ist das Informationsangebot nicht auf eine umfassende Darstellung von laufenden Projekten oder Kontakten an einzelnen Standorten ausgelegt. Entsprechend bietet forschungsdaten.info hier allenfalls punktuelle Informationen.

Eine weitere Informationsressource für strukturelle Aspekte ist forschungsdaten.org. Das seit 2013 betriebene Wiki stellt eine offene Plattform dar, die vielfältige Informationen zum FDM bündelt. Neben der Dokumentation von Community-Aktivitäten wie etwa aus dem Kontext von Arbeitsgruppen der Deutschen Initiative für Netzwerkinformation (DINI) wird es z. T. auch genutzt, um die Arbeit von Drittmittelprojekten darzustellen. Die Informationen bieten dabei zum Teil einen hohen Detailgrad, eine engere redaktionelle Betreuung der Beiträge geschieht innerhalb des Wikis nicht. Die übergreifende Struktur erscheint in vielen Teilen als organisch gewachsen, und Informationen werden nur bedingt systematisch erfasst. Für einen strukturellen Einblick in die institutionelle Verankerung von FDM-Servicestellen stellen die „FDM-Kontakte“⁵ auf forschungsdaten.org

⁴<https://wikidata.org> (zugegriffen 13. Mai 2023).

⁵<https://www.forschungsdaten.org/index.php/FDM-Kontakte> (zugegriffen 13. Mai 2023).

die bisher wohl umfassendste Übersicht mit institutionellen Kontakten dar, die neben Hochschulen auch Anlaufstellen in außeruniversitären Forschungseinrichtungen oder Serviceangebote im Rahmen von SFB berücksichtigen.

Das Ziel der Bündelung von Kontakten und Diensten verfolgt auch das Serviceverzeichnis Forschungsdaten.⁶ Dieses wurde aus der DINI-AG Forschungsdaten heraus konzipiert und verfolgt eine strukturierte Sammlung von Kontaktstellen und Diensten für das FDM. Bislang existiert das Verzeichnis jedoch nur als Prototyp und wurde noch nicht umfänglich mit Informationen gefüllt, so dass es gegenwärtig kaum als praxistaugliches Werkzeug gelten kann.

Neben diesen communitynahen oder -betriebenen Anlaufstellen zum Thema FDM existieren auch die bereits erwähnten Datenbanken von Drittmittelgebern. Diese zielen nicht eigentlich auf die Bereitstellung von Informationen zu FDM-Strukturen, bieten aber über die enthaltenen Daten zu Projekten dennoch wesentliche strukturelle Informationen. Zu nennen sind hier vor allem GEPRIS der DFG und der Förderkatalog des BMBF als Verzeichnisse der im deutschen Kontext wohl wesentlichsten Fördergeber. Die DFG nimmt eine zentrale Rolle in der Drittmittelförderung von FDM ein. Neben der NFDI und der Förderung im Bereich wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS) leisten auch INF-Teilprojekte in SFB potentiell wesentliche Beiträge zur Entwicklung von FDM-Infrastrukturen oder -Anwendungen. In GEPRIS werden sämtliche von der DFG geförderten Projekte inklusive ergänzender Informationen etwa zu den Projektbeteiligten dargestellt. Der Zugriff auf GEPRIS ist nur über die Webmaske möglich. Aufgrund fehlender Schnittstellen ist eine automatisierte Verarbeitung der Daten nicht ohne weiteres möglich. Auch das BMBF tritt durch verschiedene Ausschreibungen als struktureller Förderer von FDM auf.⁷ Der Förderkatalog des Bundes stellt für diese Projekte eine Referenzressource dar. Auch der Förderkatalog bietet keine Schnittstellen, allerdings lassen sich die Suchergebnisse im csv-Format exportieren, was ihre Verarbeitung erleichtert.

Diese Übersicht zeigt: Es existieren gegenwärtig bereits zahlreiche Informationsquellen, aus denen sich strukturelle Informationen zum Thema FDM entnehmen lassen. Diese greifen aber nicht ineinander, so dass der Versuch, sich einen Gesamtüberblick etwa zu Aktivitäten an einem Standort zu verschaffen, bisher umfassende Recherchen in verschiedenen Quellen benötigt. Eine an einer Stelle zusammengeführte Informationsbasis ist vor diesem Hintergrund ein Desiderat für alle Personen, die Interesse an strukturellen Informationen haben. Dies betrifft Personen, die Anlaufstellen für kollegialen Austausch suchen, oder auch Akteure aus der Wissenschaftspolitik. Ganz im Sinne von Open Knowledge können die entsprechenden Informationen in unterschiedlichen

⁶<https://serviceverzeichnis-forschungsdaten.org/search.html> (zugegriffen Februar 2023; vgl. <https://doi.org/10.11588/heidok.00026850>).

⁷<https://www.bmbf.de/foerderungen/bekanntmachung-1233.html> (zugegriffen 13. Mai 2023).
<https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2021/08/2021-08-17-Bekanntmachung-Fachhochschulen.html> (zugegriffen 13. Mai 2023).

Kontexten von Interesse sein – dass sie bislang nicht den FAIR-Prinzipien entsprechen, steht dem jedoch im Weg. Wir präsentieren an dieser Stelle deshalb:

3 Wikidata als Lösungsansatz

Angesichts dieser Ausgangslage möchten wir aufzeigen, welche wesentlichen Potentiale in der Zusammenführung von Informationen zu FDM-Strukturen liegen. Dabei greifen wir auf Wikidata zurück. Im Folgenden illustrieren wir, wie sich die dort vorhandene Datenbasis erweitern lässt, um relevante Strukturen im FDM abzubilden, und wie sich die so gesammelten Daten für gezielte Abfragen zu FDM-Strukturen nutzen lassen. Grundlage von Wikidata ist das Konzept von Linked Open Data, das wir an dieser Stelle kurz umreißen möchten.

Das Konzept einer Verknüpfung von offen verfügbaren Daten miteinander ist so alt wie das Internet selbst und geht auf Tim Berners-Lee zurück.⁸ Im Gegensatz zum World Wide Web, welches ein Gefüge aus Internetseiten ist, zielt das Konzept von Linked Open Data auf ein strukturiertes Inbeziehungsetzen von Wissen ab. Damit dieses Wissensnetz entsteht, bedarf es vierer Grundregeln, die man nicht-technisch wie folgt zusammenfassen kann: 1) Die Daten müssen mit einer offenen Lizenz versehen sein. 2) Die Beschreibung der Daten muss in einer standardisierten, maschinenverwertbaren Form erfolgen (im sog. *Resource Description Framework*). 3) Daten, die online verfügbar sind, verwenden die standardisierte Schnittstelle HTTP. Damit haben sie auch 4) eine einheitliche Bezeichnung und einen eindeutigen Identifikator, den *Uniform Resource Identifier*.⁹ Anhand dieser vier Grundregeln können Daten miteinander verknüpft werden, um so mitunter neue Wissensbereiche zu erschließen, deren Beziehungsgeflecht vorher nicht bekannt war.

Wikidata setzt dieses Konzept in einer Weise um, die eine niedrighschwellige Bearbeitung und Erweiterung der Datenbasis ermöglicht. Die Dateneingabe erfolgt per Browser und kann jederzeit und unabhängig von einer Anmeldung durchgeführt werden. Mit einer Registrierung gehen eine Reihe von Vorteilen einher in Form zusätzlicher Privilegien: Alle Beiträge werden auf der jeweiligen Nutzerseite angezeigt, was eine Zuordnung ermöglicht. Zudem verlieren registrierte Nutzer sukzessive Bearbeitungseinschränkungen, die einer missbräuchlichen Bearbeitung vorbeugen.¹⁰ Nutzende können sowohl Einträge bearbeiten als auch gänzlich neue Einträge erstellen. Die Erstellung neuer Einträge kann auf der Startseite über *Create a new Item* erfolgen. Jedoch sollte vorher stets

⁸Pomerantz, Jeffrey. *Metadata*. Cambridge: MIT Press, 2015. <http://doi.org/10.7551/mitpress/10237.001.0001>, S. 173–178.

⁹Vgl. <https://pro.europeana.eu/page/linked-open-data> (zugegriffen 13. Mai 2023) und die mehr technische Ausführung auf <https://www.w3.org/DesignIssues/LinkedData.html> (zugegriffen 13. Mai 2023).

¹⁰Siehe dazu ausführlich https://www.wikidata.org/wiki/Wikidata:User_access_levels (zugegriffen 13. Mai 2023).

label — **NFDI4Ing** (Q98380344) — **item identifier**

description — NFDI consortium for Engineering Sciences
National Research Data Infrastructure for Engineering Services — **aliases**

▼ In more languages
Configure

Language	Label	Description	Also known as
English	NFDI4Ing	NFDI consortium for Engineering Sciences	National Research Data Infr...
German	NFDI4Ing	Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften	NFDI Konsortium
French	No label defined	No description defined	
Bavarian	No label defined	No description defined	

All entered languages

Statements

property — **instance of** — **accepted NFDI consortium** — **value**

rank — 1 reference — **collapsed reference**

+ add value

statement group — **affiliation** — **RWTH Aachen University** — **edit**

object has role — **applicant** — **qualifiers**

▼ 1 reference
reference URL — **opened references**

https://gepris.dfg.de/gepris/projekt/442146713

+ add reference
+ add value

Abbildung 1: Übersicht der Strukturelemente in Wikidata

recherchiert werden, ob bereits ein entsprechender Eintrag existiert, um Redundanzen zu vermeiden.¹¹ Nach diesen kann über *Search Wikidata* gesucht werden.

Für die Arbeit mit Wikidata ist es unerlässlich, die Struktur der enthaltenen Daten nachzuvollziehen. Diese illustrieren wir deshalb hier anhand des NFDI-Konsortiums NFDI4Ing. Die zentrale Eigenschaft der Daten in Wikidata ist die Verknüpfung der Datenobjekte in Form sogenannter Triples. Jedes Triple stellt eine Verbindung von Elementen dar, die sich in der Form Subjekt, Prädikat und Objekt abbilden lässt und die damit im Wesentlichen der Struktur eines einfachen Aussagesatzes entspricht, z. B. „NFDI4Ing ist ein akzeptiertes NFDI-Konsortium“.

Abbildung 1 gibt wieder, wie sich die verschiedenen Komponenten von Wikidata Besuchenden der Website darstellen. Grundlage für die verknüpfte Struktur von Wikidata sind eindeutig identifizierbare *Items* (Datenobjekte) wie NFDI4Ing. Die Rolle des Subjekts in einem Triple wird immer von einem Datenobjekt gefüllt. Jedes Datenobjekt erhält einen individuellen *Item Identifier* (QXXX) – in unserem Fall Q98380344. Neben diesem verfügen Datenobjekte auch über ein Label, das die Objekte für Menschen

¹¹Für eine detailliertere Anleitung zum Anlegen neuer Einträge siehe https://upload.wikimedia.org/wikipedia/commons/9/94/How_to_Edit_Wikidata.pdf (zugegriffen 13. Mai 2023).

identifizierbar macht. Im Gegensatz zu den Identifikatoren sind Labels jedoch nicht einzigartig, sondern können mehrfach verwendet werden. Tatsächlich können auch unterschiedliche Labels für unterschiedliche Sprachen verwendet werden. Eine *Description* (Beschreibung) und *Aliases* (alternative Namen) ergänzen diese Informationen.

Die Rolle des Prädikats wird in Wikidata von *Properties* (Eigenschaften) übernommen. Diese teilen viele Eigenschaften mit Datenobjekten und lassen sich ebenfalls mittels Statements weiter beschreiben. Sie haben jedoch eine semantisch andere Rolle und ihr Identifikator beginnt in Abgrenzung zu den Datenobjekten mit einem P, wie z. B. P31 „ist ein“.

Das Objekt in einem Triple wird schließlich mit einem *Value* (Wert) belegt. Werte können in Wikidata sowohl die Form von Datenobjekten (z. B. „akzeptiertes NFDI-Konsortium“) als auch von quantitativen Werten (z. B. „5“) annehmen, und auch unbekannte Werte (z. B. wenn ein Datum unbekannt ist) und der Wert „kein Wert“ (z. B. wenn ein Planet über keinen Mond verfügt) lassen sich in Wikidata abbilden. Es ist auch ohne weiteres möglich, dass ein Statement mehrere Werte für die gleiche Eigenschaft hat (z. B. wenn ein Planet mehrere Monde besitzt). Die Werte lassen sich zudem mit *Qualifiern* (Qualifikatoren) weiter spezifizieren. Im Beispiel von NFDI4Ing ist dies für die Affiliation mit der Rheinisch-Westfälischen Technischen Hochschule Aachen (RWTH Aachen) umgesetzt. Diese hat im Statement „NFDI4Ing – (hat) Affiliation – RWTH Aachen“ zunächst die Rolle des Objekts. Sie wird jedoch umgehend zum Ausgangspunkt des Statements „RWTH Aachen – object has role – applicant“. Dabei ist wichtig, sich diesen Zusammenhang mit dem ursprünglichen Statement bewusst zu halten: Diese Aussage ist keine globale Aussage über die RWTH Aachen (sie hat also nicht an sich die Eigenschaft, „applicant“ zu sein), sondern nur über die RWTH Aachen im Kontext des Statements über NFDI4Ing.

Sämtliche Statements in Wikidata können durch *References* (Verweise) mit externen Quellen verknüpft werden und gewährleisten damit die Nachvollziehbarkeit der Informationen in Wikidata. Schließlich lassen sich die Statements in Wikidata auch durch einen *Rank* (Rang) gewichten. Dies erfolgt beispielsweise beim Vorliegen historischer Werte, Werten aus unterschiedlichen Messmethoden oder widersprüchlichen Quellen.

Die folgenden Beispiele in Tabelle 1 sollen die Anwendung des Prinzips der Triples illustrieren:

Subject	Predicate	Object	Specification	Specification Value
<i>Item</i>	<i>Property</i>	<i>Value</i>	<i>Qualifier</i>	<i>Qualifier Value</i>
„NFDI4Ing“ (Q98380344)	„instance of“ (P31)	„accepted NFDI consortium“ (Q98270496)	-	-
„NFDI4Ing“ (Q98380344)	„affiliation“ (P1416)	„RWTH Aachen“ (Q273263)	„object has role“ (P3831)	„applicant“ (Q54875338)

Tabelle 1: Beispielhafte Darstellung der Triple-Struktur von Einträgen in Wikidata

Im Folgenden stellen wir anhand von zwei Beispielen vor, wie sich einige FDM-relevante Inhalte in Wikidata abbilden lassen und zeigen, wie diese Daten nach bestimmten Kriterien abgefragt werden können.

4 Fallbeispiel 1: FDM-Servicestellen

Im ersten Fallbeispiel erstellen wir eine Liste von FDM-Servicestellen, die neben deren Namen auch deren Email-Kontakt, ihre URL und den Namen der beherbergenden Einrichtung enthält. Die Einträge in Wikidata basieren auf der eingangs diskutierten Liste von FDM-Kontakten auf [forschungsdaten.org](https://www.forschungsdaten.org), die wir im Laufe des Jahres 2022 übertragen und in der Folge aktualisiert haben.

Damit die Abfrage alle relevanten Einträge berücksichtigen kann, müssen diese vier zentrale *Properties* berücksichtigen: „instance of“ (P31), „part of“ (P361), „email address“ (P968) und „official website“ (P856).¹² Tabelle 1 gibt einen Überblick, wie diese *properties* mit den entsprechenden Servicestellen als Objekten zu Statements in Form von Triples kombiniert werden. Zum besseren Verständnis führen wir an dieser Stelle ergänzend aus, wie diese Statements zu lesen sind.

Das erste *Property* ist „instance of“ (P31) und trifft eine Aussage darüber, um was für eine Art von Entität es sich handelt. Die FDM-Servicestellen sind eine Organisationseinheit, sodass als *Value* „organizational unit“ (Q679206) eingetragen wird. Der *Value* von „part of“ (P361) ist die Forschungseinrichtung, an der die FDM-Servicestelle angesiedelt ist. Einrichtungen wie Hochschulen sind in vielen Fällen bereits als *Items* in Wikidata angelegt und mit weiteren Informationen versehen. Im Sinne der Nachvollziehbarkeit

¹²Detaillierte Informationen zu den jeweiligen *Properties* können über die jeweiligen Webseiten gemäß des Schemas <https://www.wikidata.org/wiki/Property:<PROPERTY-ID>> (beispielsweise <https://www.wikidata.org/wiki/Property:P31>) aufgerufen werden.

kann hier auch eine Referenz (*Reference*) angegeben werden. Die *Property* „email address“ (P968) ermöglicht die Angabe einer Kontaktadresse für die Servicestelle. Auch hier kann eine Referenz angegeben werden. Bei der letzten *Property* „official website“ (P856) wird die URL der Webseite als *Value* angegeben, hinzu kommt jedoch auch ein *Qualifier*, der über „language of work or name“ (P407), also die Sprache, in der die Webseite verfasst ist, Auskunft gibt. Da die URL selbst schon ein Verweis ist, ist eine Referenz hier nicht unbedingt nötig, aber möglich.

	Property	Value	Qualifier	Qualifier Value
<i>Item</i>	„instance of“ (P31)	„organizational unit“ (Q679206)	-	-
	„part of“ (P361)	Einrichtung	-	-
	„mail address“ (P968)	mailto:E-Mail-Adresse	-	-
	„official website“ (P865)	Link + „language of work or name“ (P407)	„language of work or name“ (P407)	Sprache, in der die Webseite verfasst ist

Tabelle 2: Übersicht der Statements zu den FDM-Servicestellen, die in der folgenden Abfrage verwendet werden

Auf Grundlage dieser Eintragungen lassen sich nun Abfragen stellen, die entsprechende Einrichtungen listen. Der in Listing 1 nachfolgende Code in der Abfragesprache SPARQL führt im Wikidata Query Service eine entsprechende Abfrage aus.¹³ Die mit Rauten versehenen Zeilen erklären jeweils die Funktion des Codes in der folgenden Zeile.

```
# Festlegung der in der Ausgabe angezeigten Variablen
SELECT DISTINCT ?Organisationseinheit ?OrganisationseinheitLabel ?Email
    ↪ ?Website ?EinrichtungLabel

# Festlegung der Kriterien
WHERE {
  # Die Variable "Organisationseinheit" hat die Property (wdt:) "ist ein" (P31)
    ↪ mit dem Wert (wd:) "Organisationseinheit" (Q679206)
  ?Organisationseinheit wdt:P31 wd:Q679206.

  # Die Variable "Organisationseinheit" hat die Property (wdt:) "Arbeitsgebiet"
    ↪ (P101) mit dem Wert (wd:) "Forschungsdatenmanagement" (Q30089794)
  ?Organisationseinheit wdt:P101 wd:Q30089794.
# Die Variable "Einrichtung" wird mit dem Wert der Property (wdt:) "ist Teil
    ↪ von" (P361) der Variable "Organisationseinheit" belegt
  ?Organisationseinheit wdt:P361 ?Einrichtung .

# Die Variable "Email" wird mit dem Wert der Property (wdt:) "Email" (P968) der
    ↪ Variable "Organisationseinheit" belegt. Das Vorhandensein dieses Werts
    ↪ ist optional
  OPTIONAL { ?Organisationseinheit wdt:P968 ?Email . }

# Die Variable "Email" wird mit dem Wert der Property (wdt:) "offizielle
    ↪ Webseite" (P856) der Variable "Organisationseinheit" belegt
  OPTIONAL { ?Organisationseinheit wdt:P856 ?Website . }

# Der Wikibase Label-Service wird eingebunden und die Bezeichnungen der Items
    ↪ auf Englisch genutzt
SERVICE wikibase:label{bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en".}
}
```

Listing 1: SPARQL-Abfrage zu FDM-Servicestellen

Diese Abfrage lässt sich als ein ständig aktualisiertes Verzeichnis der FDM-Servicestellen verwenden, das die in Wikidata hinterlegten Informationen bei jeder Abfrage erneut abrufen. Werden neue Servicestellen angelegt, werden sie unmittelbar in den Suchabfragen gelistet. Die Abfrage ist um weitere Informationen wie die Geokoordinaten der

¹³<https://query.wikidata.org/> (zugegriffen 13. Mai 2023). Die konkrete Abfrage ist unter dem Link <https://w.wiki/6NSL> (zugegriffen 13. Mai 2023) erreichbar.

Einrichtungen erweiterbar, wodurch sich neue Darstellungsmöglichkeiten ergeben. Dies illustrieren wir am zweiten Anwendungsfall.

5 Fallbeispiel 2: NFDI

Unser zweites Beispiel behandelt die Konsortien der NFDI. Die folgende Abfrage erfasst die im Rahmen der NFDI geförderten Konsortien anhand der beteiligten Standorte. Im Falle der NFDI-Konsortien existierte zu Beginn unserer Arbeit bereits eine lebhafte Community, die bereits in Teilen Daten zu den Konsortien in Wikidata eingearbeitet hatte. Hier wurde die Datenbasis im Laufe des Jahres 2022 durch uns ergänzt. Für die Konsortien der ersten und zweiten Ausschreibungsrunde wurden dabei die in GEPRIS verzeichneten Daten zu Applicant, Co-Applicant und Participant-Institutionen als Referenzpunkt verwendet. Da die Konsortien der dritten Ausschreibungsrunde noch nicht in GEPRIS verzeichnet waren, wurde hier auf die durch die DFG veröffentlichten Absichtserklärungen zur Antragstellung gemachten Angaben zurückgegriffen. Über den jeweiligen Ursprung eines Datums kann im Zweifel die Referenz Auskunft geben.

Für die Abfrage sind nur zwei *Properties* eines NFDI-Konsortiums zentral: „instance of“ (P31) sowie „affiliation“ (P1416). Die Zusammensetzung der Statements ist in Tabelle 3 aufgeführt. Tabelle 4 gibt ergänzend dazu aber einen Überblick über weitere Statements, die NFDI-Konsortien beschreiben. Zu ihnen gehören Informationen wie die GEPRIS-Projekt-ID, die Einordnung in der NFDI-Konferenzklassifikation oder das Datum des Förderbeginns.

Item	Property	Value	Qualifier	Qualifier Value(s)
Item	„instance of“ (P31)	„accepted NFDI consortium“ (Q98270496)	-	-
	„affiliation“ (P1416)	alle zugehörigen Institutionen und Organisationen	„object has role“ (P3831)	„applicant“ (Q54875338) „represented by“ (P1875) + Name vorsitzende Person „co-applicant“ (Q105906738) „participant“ (Q105906860)

Tabelle 3: Übersicht der Statements zu den NFDI-Konsortien, die in der folgenden Abfrage zentral sind

	Property	Value	Qualifier	Qualifier Value(s)
Item				
	„sponsor“ (P859)	„German Research Foundation“ (Q707283)	-	-
	„participant“ (P710)	alle zugehörigen Personen	„subject has role“ (P2868)	„spokesperson“ (Q17221) „represents“ (P1268) + Institution „co-spokesperson“ (Q111353657), „participant“ (Q105906860)
	„email address“ (P968)	mailto:E-Mail-Adresse	-	-
	„official website“ (P856)	Link	„language of work or name“ (P407)	Sprache(n)
	„GEPRIS project ID“ (P4870)	ID	-	-
	„chairperson“ (P488)	Name vorsitzende Person	-	-
	„inception“ (P571)	Datum Beginn	-	-
	„field of work“ (P101)	„research infrastructure“ (Q1438053) und die anderen Bereiche	-	-
	„part of“ (P361)	„National Research Data Infrastructure“ (Q61658497)	-	-
	„parent organization“ (P749)	„German National Research Data Infrastructure“ (Q105757481)	-	-

Tabelle 4: Übersicht von weiteren Statements zu NFDI-Konsortien, die zwar nicht in der folgenden Abfrage verwendet werden, jedoch in den Einträgen zu den NFDI-Konsortien enthalten sind

Die folgende Abfrage in Listing 2 generiert eine Kartendarstellung der Standorte, die an den geförderten NFDI-Konsortien beteiligt sind. Die Abfrage ist unter dem Link <https://w.wiki/72bE> erreichbar.¹⁴ Diese werden über Markercluster abgebildet, die sich beim Heranzoomen zunehmend aufteilen, bis einzelne Standorte sichtbar werden. Durch Klicken auf die Marker werden Informationen zu Standort, zum NFDI-Konsortium, an dem die Beteiligung erfolgt, sowie zur Rolle innerhalb des Konsortiums angezeigt. Grundlage dieser Darstellung sind die Geokoordinaten der an den Konsortien beteiligten Einrichtungen, die unabhängig von den Einträgen der Konsortien existieren und die im Zuge der Abfrage erfasst werden. Sämtliche abgefragten Informationen lassen sich jedoch ebenso in tabellarischer Form anzeigen.

¹⁴<https://w.wiki/72bE> (zugegriffen 13. Juli 2023).

```

# Festlegung der Standarddarstellung der Ergebnisse als Karte unter Verwendung
  ↳ von Markerclustern
#defaultView:Map{"layer":"?affiliationTypeLabel", "markercluster": "true"}
# Festlegung der in der Ausgabe angezeigten Variablen
SELECT ?Konsortium ?KonsortiumLabel ?AffiliationTypeLabel ?Institution
  ↳ ?InstitutionLabel ?geoLoc
# Festlegung der Kriterien
WHERE {
# Die Variable "Konsortium" hat die Property (wdt:) "ist ein" (P31) mit dem
  ↳ Wert (wd:) "angenommenes NFDI-Konsortium" (Q98270496)
  ?Konsortium wdt:P31 wd:Q98270496.
# Die Variable "statement" wird mit der Statement-Node (p:) "Zugehörigkeit zu"
  ↳ (P1416) belegt
  ?Konsortium p:P1416 ?statement .
# Die Variable "Institution" wird mit dem Wert der Statement-Node (ps:) "Zugehö
  ↳ rigkeit zu (P1416) belegt
  ?statement ps:P1416 ?Institution .
# Die Variable "AffiliationType" wird mit dem Wert des Qualifiers (pq:) "Objekt
  ↳ fungiert als" belegt
  ?statement pq:P3831 ?AffiliationType .
# Die Variable "geoLoc" wird mit dem Wert "geographische Koordinaten" (P625)
  ↳ der Variable "Institution" belegt. Das Vorhandensein dieses Wertes ist
  ↳ optional
  OPTIONAL { ?Institution wdt:P625 ?geoLoc. }
# Die Variable "geoHead" wird mit dem Wert "geographische Koordinaten" (P625)
  ↳ der Hauptverwaltung (P159) der Variable "Institution" belegt. Das
  ↳ Vorhandensein dieses Wertes ist optional
  OPTIONAL { ?Institution (wdt:P159/wdt:P625) ?geoHead. }
# Die Variablen "geoLoc" und "geoHead" werden in der Variable "geoLoc"
  ↳ zusammengeführt. Wenn "geoLoc" bereits belegt ist, bleibt dieser Wert
  ↳ erhalten, wenn er leer ist, wird mit dem Wert von "geoHead" aufgefüllt
  BIND(COALESCE(?geoLoc, ?geoHead) AS ?geoLoc)
# Der Wikibase Label-Service wird eingebunden und die Bezeichnungen der Items
  ↳ auf Englisch genutzt
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
# Die Ergebnisse werden in der Tabellendarstellung aufsteigend nach dem Wert
  ↳ der Variable "InstitutionLabel" sortiert
ORDER BY ASC (?InstitutionLabel )

```

Listing 2

Die Kartendarstellung bietet einen interaktiven Zugang, um die Beteiligung an der NFDI aus ihrer räumlichen Verteilung heraus zu erkunden. Die in ihrer Gesamtheit sehr

umfangreiche Beteiligung an der NFDI lässt sich zudem über spezifischere Abfragen einschränken und so etwa auf bestimmte Fragestellungen hin darstellen – beispielsweise die Beteiligung aus verschiedenen Wissenschaftsbereichen. Dies möchten wir an dieser Stelle jedoch nicht weiterverfolgen, sondern in die Diskussion über die Potentiale wie auch die Grenzen dieses Ansatzes zur Sammlung und Darstellung von Informationen zu FDM-Strukturen übergehen.

6 Chancen und Grenzen des Einsatzes von Wikidata

Die vorgestellten Beispiele verdeutlichen, dass sich relevante Strukturen im Bereich FDM gut mittels Wikidata abbilden lassen. Einige Chancen und Herausforderungen dieses Ansatzes sollen an dieser Stelle hervorgehoben werden.

Grundsätzlich sind die Daten in Wikidata nutzendengepflegt. Sie sind öffentlich zugänglich und können von jeder Person geändert werden. Die Missbrauchspotentiale eines solchen Ansatzes sind offensichtlich: Daten können grundsätzlich beliebig verfälscht und entfernt werden. Das Risiko, dass dies geschieht, sollte jedoch abgewogen werden. Für den Bereich FDM scheint uns das Risiko von kontroversen Daten oder von Sabotageakten insgesamt überschaubar. Zudem wird jede Änderung in den Daten in Wikidata dokumentiert und es existieren Mechanismen, um die Editierbarkeit einzuschränken und böswilligen Eingriffen in den Datenbestand entgegenzuwirken.

Durch die vielfältigen Beitragenden kann es in Wikidata dazu kommen, dass Daten heterogen abgebildet werden, so dass der Aspekt der Datenqualität besonderer Diskussion bedarf. In unseren Beispielen hat sich dies im Falle der Geokoordinaten in den Kartendarstellungen gezeigt, für die auf zwei unterschiedliche Werte zurückgegriffen wurde: zum einen auf den Wert für die Koordinaten der Einrichtungen und zum anderen auf den Wert für den Sitz der Hauptverwaltung. Zudem erfassen die Abfragen auch Institutionen, die über mehrere Einträge für Geokoordinaten verfügen. Umgekehrt gibt es Einrichtungen, deren Einträge keine Geokoordinaten enthalten. Es ist offensichtlich, dass Mehrfacheinträge oder Datenlücken zu einer Verzerrung führen können. Zudem sind unter den verknüpften Informationen falsche Daten möglich. Deshalb gilt es, im Umgang mit den Daten ein Bewusstsein für mögliche Fehlerquellen und Lücken zu bewahren. Die Fehlerquellen unterscheiden sich dabei allerdings nicht grundsätzlich von denen bei Abfragen geschlossener und von Grund auf selbst generierter Datenbestände. Eine Abwägung, inwieweit ein Datensatz eine gegebene Frage adäquat beantworten kann, ist stets zu treffen.

Besonders hervorheben möchten wir in diesem Zusammenhang die Potentiale der nutzendengetriebenen Pflege des Datenbestands. Denn im Gegensatz zu zentral gepflegten Datenbeständen ist es gerade die verteilte Überarbeitung und Erweiterung der Datenbasis, die die Datengrundlage ohne einzelne zentral verantwortliche Stelle aus verschiedenen Quellen zusammenträgt und aktuell hält. Hierdurch kann auch schneller

auf Änderungen der Daten eingegangen werden, da der Weg über eine einzubindende Redaktion entfällt. Eine kontinuierliche Nutzung der Daten und ein von verschiedenen Nutzenden geteiltes Interesse an den Daten sind dabei Faktoren, die die Datenqualität stützen. Fällt bei Nutzung der Daten auf, dass Informationen fehlen oder falsch sind, lassen sich diese umgehend ergänzen und korrigieren. Bezogen etwa auf das Verzeichnis der FDM-Kontaktstellen lässt sich dieses mit geringen Hürden von den Einrichtungen ergänzen und aktuell halten, wenn beispielsweise an einer Einrichtung eine Kontaktstelle neu eingerichtet wird oder eine Veränderung in den Kontaktdaten erfolgt. Dies setzt einen engagierten Nutzendenkreis voraus, der in der Lage ist, Daten in Wikidata zu bearbeiten. Die Hürden für die Bearbeitung von Daten oder das Erstellen von neuen *Items* nach einem etablierten Muster erscheinen uns überschaubar. Und wo für das Erstellen von komplexen Abfragen zumindest grundlegende SPARQL-Kenntnisse nötig sind, sind diese nicht nötig, um zur Pflege des Datenbestands beizutragen. Wir sind daher der Ansicht, dass die Potentiale einer nutzendenbasierten Datenverwaltung in Wikidata, die auch den Prozess der Datenpflege dokumentiert, die Risiken zumindest in unseren Anwendungsszenarien deutlich überwiegen.

Hinsichtlich der weiteren Nutzung der Daten ist deren Strukturierung sowie vorhandene Schnittstellen in Wikidata ein wesentliches Plus. So lassen sich die SPARQL-Abfragen leicht etwa in Jupyter Notebooks integrieren und dort für komplexere Auswertungen verwenden. Dass die Nachvollziehbarkeit, die solche Auswertungsansätze bieten, für den Bereich FDM mehr als eine Zier sind, soll dabei trotz aller Offensichtlichkeit kurz erwähnt sein.

Den Vorteilen in der Auswertbarkeit steht die Seite der Dateneingabe entgegen, die einen nicht unerheblichen Aufwand darstellt. Zwar ist der Einsatz von Bots in Wikidata möglich, geht jedoch mit der nicht trivialen Hürde der Programmierung einher. Im FDM stellen zudem die Datenquellen eine weitere Hürde dar: Wie unsere Beispiele zeigen, kann die Bereitstellung unterschiedlich erfolgen. Im Falle der FDM-Servicestellen gibt es keine zuständige Stelle, die diese Daten gezielt erfasst. Wikidata selbst könnte hier, eine entsprechende Pflege aus der Nutzendenschaft heraus vorausgesetzt, zur hauptsächlichen Datenquelle werden. Anders ist es bei den Förderdaten gelagert. Der Förderkatalog des Bundes bietet immerhin einen Export von csv-Dateien an, der eine automatisierte Verarbeitung ermöglicht. In GEPRIS hingegen sind die Daten nur über eine Webansicht zugänglich. In beiden Fällen sind die Daten strukturiert, wünschenswert wären aber zweifelsohne dokumentierte Schnittstellen, die einen niedrigschwelligen Zugriff ermöglichen. Dies würde Workflows erlauben, die die Vollständigkeit und die Aktualisierung der Daten in regelmäßigen Abständen sicherstellen. Denn gerade für große Strukturen wie die NFDI, deren Dynamik aufgrund der Vielzahl der Beteiligten schnell zu einer Zerfaserung führen kann, stellt die Aktualität der Daten eine große Herausforderung dar, bei der eine manuelle Überprüfung schnell an ihre Grenzen stößt.

7 Fazit und Ausblick

Wir hoffen, dass wir mit unserem Beitrag illustrieren konnten, wie sich Wikidata in der dynamischen FDM-Landschaft einsetzen lässt, um einen Überblick über Akteure und Aktivitäten zu behalten. Sicherlich erfordert das Erstellen von Abfragen an Wikidata eine gewisse Einarbeitung, das Erstellen von Einträgen nach einem etablierten Muster erscheint demgegenüber niedrigschwellig und ermöglicht so die nutzendengetriebene Pflege des Datenbestands.

Die vorgestellten Abfragen sind allerdings raumlos, solange sie nicht adäquat eingebunden werden. Wikidata kann deshalb sicherlich nicht die eingangs genannten bestehenden Informationsquellen wie forschungsdaten.info oder forschungsdaten.org ersetzen. Vielmehr denken wir, dass die Einbettung von Inhalten, die in Wikidata gepflegt werden, eine geeignete Ergänzung der dort vorhandenen Inhalte darstellen kann, die durch die verteilte Bearbeitung besonders einfach aktuell gehalten werden kann. Durch die Einbindung von Übersichten auf Grundlage von Wikidata lassen sich die vorhandenen Informationen in einen größeren Zusammenhang setzen und aufwerten. Die Ergänzungen dieser Informationen tragen zudem zur Vernetzung der verschiedenen Akteure bei.

Die gleiche Datengrundlage lässt sich dabei für unterschiedliche Zielgruppen abhängig von deren Interessenschwerpunkten aufbereiten. Möchte man wissen, in welchen Bundesländern Aktivitäten stattfinden oder welcher Typus von Einrichtung an einem Projekt beteiligt ist? Wikidata erlaubt es, auch komplexe Fragestellungen zu beantworten. Da bereits ein massiver Bestand an Daten, etwa zu Forschungseinrichtungen oder administrativen Zusammenhängen, vorhanden ist, werden sich neue Daten in vielen Fällen leicht ergänzen und verknüpfen lassen. Dies möchten wir in weiterer Arbeit ausführen. Wichtig ist bei allem, einen hinreichenden Konsens hinsichtlich der Abbildung von gleichartigen Entitäten zu finden, um nicht im Datensalat zu enden. Hier gilt es also, kontinuierlich den Austausch zwischen Interessierten zu suchen und musterhafte Lösungen zu adaptieren. Wir hoffen, dass unser Beitrag dazu einlädt, sich mit den Möglichkeiten von Wikidata auseinanderzusetzen und sich an der Pflege der Datenbasis zu FDM-Strukturen zu beteiligen – denn die Offenheit und Transparenz der Daten steht gerade einem Thema wie FDM gut.

8 Danksagung

Wir danken der gesamten Wikidata-Community für Ihre Arbeit, insbesondere aber den Mitwirkenden am Wikidata Project NFDI und unter diesen vor allem Cord Wiljes für vielfältige Anregungen. Dank gebührt ebenfalls Amar Suljkić für seine Unterstützung bei der Dateneingabe. Der Autor Évariste Demandt möchte sich bei Bund, Ländern und der Gemeinsamen Wissenschaftskonferenz (GWK) für die Förderung und Unterstützung

im Rahmen des Konsortiums NFDI4Ing (gefördert durch die Deutsche Forschungsgemeinschaft – Projektnummer 442146713) bedanken.