

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

Interoperabilität von Metadaten innerhalb der NFDIⁱ

Konsortienübergreifender Metadaten-Workshop am 2./3. Juli 2020

Dorothea Iglezakis	Marc Fuhrmans	Susanne Arndt	Évariste Demandt
Stephan Hachinger	Daniela Hausen	Giacomo Lanza	Johannes Lipp
	Rainer Stotzka	Džulia Terzijska	

2021

Zitiervorschlag

Iglezakis, Dorothea, Fuhrmans, Marc, Arndt, Susanne, Demandt, Évariste, Hachinger, Stephan, Hausen, Daniela, Lanza, Giacomo, Lipp, Johannes, Stotzka, Rainer und Džulia Terzijska. 2021. Interoperabilität von Metadaten innerhalb der NFDI. Konsortienübergreifender Metadaten-Workshop am 2./3. Juli 2020. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2021: S. 124-135. DOI: [10.17192.bfdm.2021.2.8313](https://doi.org/10.17192.bfdm.2021.2.8313).

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

ⁱDorothea Iglezakis (ORCID: [0000-0002-8524-0569](https://orcid.org/0000-0002-8524-0569)), Marc Fuhrmans (ORCID: [0000-0002-9826-018X](https://orcid.org/0000-0002-9826-018X)), Susanne Arndt (ORCID: [0000-0002-1019-9151](https://orcid.org/0000-0002-1019-9151)), Évariste Demandt (ORCID: [0000-0002-5705-0071](https://orcid.org/0000-0002-5705-0071)), Stephan Hachinger (ORCID: [0000-0001-8341-1478](https://orcid.org/0000-0001-8341-1478)), Daniela Hausen (ORCID: [0000-0001-9083-0670](https://orcid.org/0000-0001-9083-0670)), Giacomo Lanza (ORCID: [0000-0002-2239-3955](https://orcid.org/0000-0002-2239-3955)), Johannes Lipp (ORCID: [0000-0002-2639-1949](https://orcid.org/0000-0002-2639-1949)), Rainer Stotzka (ORCID: [0000-0003-3642-1264](https://orcid.org/0000-0003-3642-1264)), Džulia Terzijska (ORCID: [0000-0002-1698-6826](https://orcid.org/0000-0002-1698-6826)).

1 Zusammenfassung

Auf einem virtuellen Metadatenworkshop diskutierten Vertreterinnen und Vertreter von 17 verschiedenen NFDI-Konsortien Konzepte zur Interoperabilität von Metadaten in Bezug auf inhaltliche Überschneidungen, Herangehensweisen zur Metadatenmodellierung, technische Aspekte und Infrastruktur bis hin zu organisatorischen Aspekten und Zusammenarbeit. Die vertretenen Konsortien möchten auch in Zukunft eng im Bereich Metadaten zusammenarbeiten und Standards schaffen, die es Forschenden in der Zukunft möglich machen, ihre Forschungsergebnisse gleichzeitig spezifisch und interoperabel zu beschreiben. Die Dynamik der Formierung der Nationalen Forschungsdateninfrastruktur (NFDI) soll für Einigung über allgemein bedeutsame Fragen der Metadatierung genutzt werden. Zudem sollen mit Unterstützung des NFDI-Direktorats Strukturen zur gemeinsamen Standardisierungsarbeit geschaffen und international vernetzt werden.

2 Einleitung

Am 2. und 3. Juli 2020 trafen sich Vertreterinnen und Vertreter von 17 verschiedenen NFDI-Konsortien in einem virtuellen Workshop, um Möglichkeiten der Zusammenarbeit und der Interoperabilität zwischen ihren unterschiedlichen Metadaten-Konzepten auszuloten. Die Teilnehmenden tauschten sich sehr rege und engagiert auf allen Kanälen der Plattform aus. Die Arbeitsgruppe Metadata4Ing von NFDI4Ing initiierte und moderierte den Workshop. Vertreter waren sowohl Konsortien aus den Lebenswissenschaften (GHGA¹, DataPlant², NFDI4Health³, NFDI4Medicine, NFDI4Agri⁴, NFDI4Earth⁵, NFDI4BioDiversity⁶) wie auch aus den Sozial- und Wirtschaftswissenschaften (KonsortSWD⁷, BERD@NFDI⁸), den Geisteswissenschaften (Text+⁹, NFDI4Culture¹⁰), den Naturwissenschaften (NFDI4Chem¹¹, NFDI4Cat¹², DAPHNE¹³) und den Material- und Ingenieurwissenschaften (NFDI4MobilTech¹⁴, NFDI4MatWerk¹⁵ und NFDI4Ing¹⁶).

¹<https://ghga.dkfz.de/>. Letzter Abruf für alle URLs: 25.03.2021

²<http://nfdi4plants.de/?page=consortium>

³<https://www.nfdi4health.de/>

⁴<https://www.nfdi4agri.de/>

⁵<https://www.nfdi4earth.de/>

⁶<https://www.nfdi4biodiversity.org/>

⁷<https://www.konsortswd.de/>

⁸<https://www.berd-nfdi.de/>

⁹<https://www.text-plus.org/>

¹⁰<https://nfdi4culture.de/consortium.html>

¹¹<https://www.nfdi4chem.de/>

¹²<http://nfdi4cat.org/>

¹³<https://www.sni-portal.de/de/daphne-nfdi>

¹⁴[https://www.dfg.de/download/\[...\]/nfdi_konferenz_2020/nfdi4mobiltech_abstract.pdf](https://www.dfg.de/download/[...]/nfdi_konferenz_2020/nfdi4mobiltech_abstract.pdf)

¹⁵<https://nfdi-matwerk.de/>

¹⁶<https://nfdi4ing.de/>

Die Konsortien streben im Rahmen der Umsetzung der FAIR-Prinzipien¹⁷ an, mit Hilfe von standardisierten Metadaten einerseits die Auffindbarkeit von Daten zu verbessern und zum anderen Möglichkeiten zu schaffen, verschiedene Datensätze (auch interdisziplinär) miteinander in Beziehung zu setzen.

Während in manchen Bereichen, insbesondere in den Lebenswissenschaften, bereits eine Vielzahl von Terminologien und (Metadaten-)Standards sowohl durch internationale Standardisierungsorganisationen wie auch durch grassroots-Bewegungen in den Fachcommunities existieren, gibt es in anderen Bereichen, wie beispielsweise in den Ingenieurwissenschaften, wenige von der Community akzeptierte Standards. Die Herausforderung besteht darin, vorhandene Standards miteinander in Beziehung zu setzen und gemeinsame Beschreibungsstandards für bestehende Lücken zu entwickeln. Weiterhin diskutierten die Teilnehmenden Aspekte der Zusammenarbeit in vier verschiedenen Diskussionsräumen: (1) Inhaltliche Überschneidungen der mit Hilfe von Metadaten zu beschreibenden Entitäten, (2) Fragen der Metadatenmodellierung, (3) technische Aspekte der existierenden und geplanten Infrastrukturen für Metadaten und schließlich (4) organisatorische Aspekte und die weitere Zusammenarbeit.

3 Diskussionsraum *Inhaltliche Überschneidungen*

Ein Ziel des Workshops war die Identifikation von inhaltlichen Überschneidungen und damit Potentialen zur Zusammenarbeit zwischen den Konsortien. Aus im Vorfeld des Workshops gesammelten Informationen zu genutzten Methoden, Tools und Forschungsgegenständen entstand eine Überlapp-Matrix¹⁸, die während des Workshops und im Nachgang gemeinsam weiterbearbeitet wurde. Sie gibt einen guten ersten Überblick darüber, wo inhaltliche Gemeinsamkeiten bestehen und Zusammenarbeit an Metadatenschemata und -standards sinnvoll wäre. Um aussagekräftige Information zu liefern, muss die Matrix aber im Nachgang innerhalb der Konsortien abgestimmt, weiter ausgearbeitet und klarer definiert werden. Ein Überblick über bestehende, in Entwicklung befindliche und geplante Schemata und Standards soll Parallelentwicklungen vorbeugen und Kooperationspotentiale aufzeigen.

Konsens war, dass eine Zusammenarbeit auf verschiedenen Ebenen in unterschiedlichen Zusammensetzungen notwendig ist und auch im internationalen Kontext erfolgen muss. Wichtigstes Ziel der Zusammenarbeit war für die Diskussionsteilnehmenden, Forschenden zu ermöglichen, ihre Forschungsdaten gleichzeitig spezifisch und interoperabel zu beschreiben. Entscheidend dafür ist eine Beschreibung der Provenienz der Daten mit ihren Methoden und Werkzeugen, dem Forschungsgegenstand, der zeitlichen und räumlichen Bestimmung ihrer Erhebung, aber auch der Bestimmung des Kontextes.

¹⁷Findable, Accessible, Interoperable, Reusable, siehe „The FAIR Data Principles“. <https://www.force11.org/group/fairgroup/fairprinciples>

¹⁸<https://docs.google.com/spreadsheets/d/14NppjHh7slo4Weo7kgYhARjFF4iVIYTQnVzovk3doj0/edit?usp=sharing>

Als konkrete nächste Schritte wurde zunächst eine klare Definition und eine systematische Zusammenstellung der Ziele und Anwendungsspektren gefordert, die mit der gemeinsamen Arbeit an den Metadaten erreicht werden sollen (z. B. verbesserte Auffindbarkeit und Verknüpfbarkeit von Daten innerhalb von Fachbereichen und über Fachbereiche hinweg). Ein Überblick über Bestehendes und die Identifikation von Lücken kann zur inhaltlichen Eingrenzung der gemeinsamen Arbeit dienen. Die eigentliche Zusammenarbeit kann dann auf verschiedenen Ebenen (allgemein genutzte vs. fachspezifische Metadatenkategorien) mit verschiedenen Beteiligten in unterschiedlichen Rollen (aktive Entwicklung vs. Nachnutzung der Ergebnisse) stattfinden. Für diese Zusammenarbeit müssen innerhalb und auch außerhalb der NFDI Strukturen geschaffen werden. Bereits bestehende fachspezifische Beschreibungskulturen und Standards sollen respektiert und über Mappings (in Form von Crosswalks oder auch eines gemeinsamen Ontologie-Baums) miteinander verknüpft werden; unterschiedliche Auslegungen ähnlicher Begriffe in unterschiedlichen Fachgebieten sollten ebenso adressiert werden.

Konkret genannte inhaltliche Bereiche einer baldigen Zusammenarbeit waren eine schnelle Einigung über von allen benötigte Metadatenkategorien (wie z. B. Personenidentifikation), die spezifische Entwicklung von mehrfach genutzten Informationen wie einem Ortsnamenverzeichnis und die Entwicklung einer maschinenlesbaren Kennzeichnung von Zugriffsrechten, um den Umgang mit lizenzierten Informationen zu erleichtern.

4 Diskussionsraum *Metadatenmodellierung*

Ziel der Session zum Thema Metadatenmodellierung war, sich bezüglich bekannter Ansätze zur Modellierung von fachspezifischen Metadaten in Form von standardisierten Schemata oder Terminologien (im Folgenden als Standardisierung bezeichnet) auszutauschen. Zunächst wurden die geplanten Vorhaben bzgl. Metadaten der vertretenen NFDI Konsortien abgefragt. Eine Übersicht findet sich in [Tabelle 1](#).

4.1 Anforderungen und Ziele

Bezüglich der zu beachtenden Anforderungen und Ziele bestand Konsens, dass eine Standardisierung zugleich Interoperabilität, Spezifität, Flexibilität sowie eine möglichst hohe Nachnutzbarkeit bzw. eine weite Anwendbarkeit aufweisen muss. Weiter muss die Anwendung einfach und vorteilhaft für die Wissenschaftler:innen sein, und die Standardisierungen müssen kostenfrei verfügbar sein. Außerdem sollte der gewählte Ansatz einen iterativen Standardisierungsprozess unterstützen.

Konsortium	Entwicklung		Anpassung		Verwendung	
	M	T	M	T	M	T
BERD@NFDI	x	x	(x)		x	x
DAPHNE	x	x	x	x		
DataPlant	x	x	x	x		
GHGA	x	x	x	x		
KonsortSWD		x	x	x	x	x
NFDI4Agri	(x)	(x)	x	x	x	x
NFDI4BioDiversity						
NFDI4Cat	x	x	x	x	x	x
NFDI4Chem						
NFDI4Culture	x	x	x	x	x	x
NFDI4Earth		x	x	x	x	x
NFDI4Health	x	(x)	x	x	x	x
NFDI4Ing	x	x	x	x		
NFDI4MED	x		x	x	x	x
NFDI4MobilTech	x	x	x	x		
NFDI-MatWerk	x	x	x	x		
Text+	x	x	x	x	x	x

Tabelle 1: Übersicht über die verschiedenen Konsortien und ihre Aktivitäten zur Gestaltung von Metadatenschemata (Unterspalte M) und Terminologien (Unterspalte T). Entwicklung beinhaltet dabei neben Neuentwicklung auch Weiterentwicklung zur Erhöhung der Domänenspezifität (Hinzufügen neuer Klassen oder Eigenschaften), wohingegen Anpassungen Arbeit an Standardisierungen bedeutet, die nicht (primär) das Ziel haben, deren Fachspezifität zu erhöhen.

4.2 Modellierungsansätze

Ebenso bestand Einigkeit, dass, um die oben genannten Anforderungen und Ziele zu erreichen, bestimmte Vorgehensweisen grundsätzlich geeignet sind (sog. „Best Practices“). So bietet es sich zum einen an, Metadaten modular aufzubauen, indem man getrennt beschreibbare Domänen in separat definierten Terminologie-Klassen und Schemata erfasst. Eine Trennung der Domänen Versuchsgegenstand, Methode und Werkzeug wurde dabei von den meisten der vertretenen Wissenschaftsgebiete als geeignet betrachtet. Ebenfalls wurde ein modulares Vorgehen auf Terminus-Ebene als sinnvoll erachtet, in dem bei der Definition neuer Standardisierungen auf Termini aus vorhandenen Terminologien zurückgegriffen wird. Darüber hinaus empfiehlt es sich, Bezug zu bestehenden Terminologien auch durch Verweise explizit anzugeben, z. B. durch semantische Instrumente wie owl:equivalentClass/Property. Allerdings ist zu beachten, dass echte semantische Äquivalenz nur selten gegeben ist und die Gefahr eines Missverständnisses durch eingeführte Unschärfen besteht. Zur Modellierung von Metadatenschemata und Terminologien ist dabei auf gängige Standards wie RDF(S) und OWL (Ontologiemodellierung), ISO 25964, iso-thes, SKOS (Thesaurusmodellierung) und ISO 704 sowie ISO 1087 (Terminologiemodellierung) zurückzugreifen. Diese enthalten eine Vielzahl an Möglichkeiten, um die gewünschten Eigenschaften präzise und für andere verständlich auszudrücken. Außerdem kann von *Ontology Design Patterns* ausgegangen werden.

Als weiteres empfehlenswertes Designprinzip wurden hierarchische Strukturen gesehen. So können zum einen über Vererbung wie in der Objekt-orientierten Programmierung spezifischere Terminologie-Klassen und Schemata als Kinder von weniger spezifischen Eltern abgeleitet werden, wobei die Kinder alle Eigenschaften der Eltern erben, und zusätzlich neue hinzufügen. Dadurch kann auch die oben beschriebene Modularität der Standardisierung weiter erhöht werden, indem die Ableitung in Form einer sich verzweigenden Baumstruktur erfolgt. Hier ist aber zwischen den Vorteilen einer flexiblen Beschreibung durch Kombination kleinteiliger modularer „Mikro“-Standardisierungen und der resultierenden Komplexität abzuwägen. Bezug zwischen den getrennten Teilen kann hergestellt werden, indem etwa eine Eigenschaft die Instanz einer an anderer Stelle im Baum beschriebenen Klasse referenziert. Die Baumstruktur kann auch durch Verschmelzen von Zweigen aufgehoben werden, indem Kinder erstellt werden, die von mehreren Eltern gleichzeitig abstammen. Dies bewirkt gewisse Vorteile z. B. für semantische Schlussfolgerung, steht aber in Konflikt mit der ebenfalls vorteilhaften Modularität der Modellierung.

Weiter können hierarchische Strukturen auch über semantische Relationen wie rdfs:subClass/PropertyOf abgebildet werden. Diese Option birgt in Bezug auf semantische Schlussfolgerung großes Potential, muss aber gerade deshalb als komplex und nur eingeschränkt laientauglich bezeichnet werden.

Eine Versionierung von entwickelten und überarbeiteten Schemata und Vokabularen wurde klar als notwendig befunden. Sie unterstützt sowohl eine sichere Nachnutzung als auch eindeutige Identifizierung der erstellten Inhalte. Die Umsetzung ist bisher nicht eindeutig gelöst und stellt einen aktuellen Gegenstand der Forschung dar. Ver-

breitete Methoden umfassen neben der Integration der aktuellen Versionsnummer in die Ontologie-URI (Beispiel [https://\[...\]-dcat-2-20200204/](https://[...]-dcat-2-20200204/)) auch die Verwendung von entsprechenden Properties innerhalb der Ontologien selbst, wie etwa owl:versionInfo.¹⁹

4.3 Entscheidender Einfluss von Randbedingungen

Auch wenn bezüglich der grundsätzlichen Eignung der oben dargestellten Ansätze zur Modellierung interoperabler und flexibler fachspezifischer Metadaten große Einigkeit bestand, zeigte sich auch deutlich, dass die letztendliche Entscheidung für eine Strategie stark von den jeweiligen Randbedingungen abhängt. Ein wesentlicher Unterschied zwischen den Konsortien ist dabei, ob bereits etablierte Standards zur Verfügung stehen. Wenn letzteres der Fall ist, nehmen diese Standards eine zentrale Rolle ein und bestimmen durch ihre Ausformung das weitere Vorgehen. Ebenso hat die Heterogenität und Variabilität der innerhalb eines Konsortiums zu beschreibenden Entitäten einen Einfluss auf die Wahl des Ansatzes.

Ein weiteres Ergebnis der Diskussion war, dass die Konsortien bezüglich ihrer Pläne zur Modellierung auf sehr unterschiedlichen Fortschrittsständen standen, welche von ersten Untersuchungen möglicher Metadaten bis hin zu jahrelangem Einsatz von Standards für Inhalt und Anwendungen reichen.

5 Diskussionsraum *Technische Aspekte und Infrastruktur*

Metadatenhaltung ist niemals ganz unabhängig vom technischen Verwendungszweck und -kontext der Daten. Oftmals benötigen z. B. Benutzerschnittstellen (wie Datenabruf- oder Suchportale) sehr spezielle Angaben zur Natur der Daten, zu ihren Formaten und der beabsichtigten Sichtbarkeit. Aus diesem Grund ist es auf dem Weg zu gemeinsamen Metadaten-Strategien sinnvoll zu eruieren, welche technischen Plattformen, Dienste und Werkzeuge verschiedene NFDI-Konsortien beim Forschungsdatenmanagement einzusetzen gedenken. Neben diesen praktischen Fragen spielen für das Metadaten-Management technische Standards und Schnittstellendefinitionen, wie sie in der jeweiligen Community verbreitet sind, eine große Rolle.

Die Session „Technische Aspekte und Infrastruktur“ trug demnach von allen teilnehmenden Konsortien die einzusetzenden Spezifikationen, technischen Funktionalitäten, Systeme und Dienste, Komponenten und Werkzeuge zusammen – soweit dies zum derzeitigen Zeitpunkt schon möglich ist. Hierzu wurde ein „Brainstorming“ durchgeführt, bei dem die Teilnehmenden aufgefordert wurden, die Komponenten ihres Konsortiums in den Bereichen

¹⁹Für eine Entwicklung mit Änderungshistorie ist zusätzlich auch ein kollaboratives Git-Repository zu empfehlen.

- Portale / Benutzerschnittstellen,
- semantische Dienste,
- Spezifikationen,
- Protokolle / Schnittstellen / APIs,
- technische Basisdienste / Repositorien,
- Werkzeuge, und
- Sonstiges

zu nennen. Hieraus ließen sich einige gemeinsame Tendenzen extrahieren.

Es besteht allgemein die Vision, mit eher fachspezifischen Portalen zu Forschungsdatensuche und -abruf (z. B. auf Basis von CKAN²⁰, SEEK, etc.) und modular einsetzbaren Werkzeugen den gesamten Forschungsprozess zu begleiten. Dies betrifft Metadatenerfassung, systemübergreifendes Management von persistenten Identifikatoren (PIDs), und Erfassung der Provenienz von Datenprodukten und ihren Teilen. Ein Problem hierbei sind fehlende „einfach verwendbare“ Werkzeuge zur Metadatenerfassung, aber auch – eine Ebene höher – zur Erstellung und Kuratierung von Vokabularen, Terminologien und Ontologien (z. B. Protégé²¹). Insbesondere die kollaborative Erstellung von Metadatenschemata und Metadatensätzen im Team ist schwierig, auch adäquate Automatisierung ist ein Problem. Im experimentellen Bereich werden teils gute Ergebnisse mit der Erfassung von Metadaten in elektronischen Laborbüchern (ELN, z. B. Chemotion²²) oder Labor-Informations- und Management-Systemen (LIMS) erzielt.

Die Daten- und Metadaten-Speicherebene stützt sich auf Repositoriumsdienste, welche von o. g. Portalen und Werkzeugen adressiert werden. Diese wurden von allen Konsortien als wichtig erachtet. Während einige Konsortien auf zentrale Plattformen wie RADAR²³ setzen, wurde eine Zentralisierung im Allgemeinen oft auch skeptisch gesehen, und eine Zusammenarbeit dezentraler Repositorien auf Interoperabilitäts-Basis als vielversprechend erachtet. Hierfür werden Schnittstellen / APIs und Standardspezifikationen allgemein als wichtig erkannt; mehrfach wurde das bewährte OAI-PMH²⁴ genannt. Oftmals sind gewisse Standards, die auch Metadaten betreffen, durch das Arbeitsumfeld vorgeschrieben – z. B. im kommerziellen Ingenieurbereich oder im Bereich von Behördendaten.

Ein wichtiges Problem beim Zugriff auf (v. a. zugangsgeschützte) Daten ist der Wildwuchs in der Landschaft der Authentifizierungs- und Autorisierungsinfrastrukturen, für den praktisch alle Konsortien eine bessere und allgemeinere, aber dennoch feingranulare Lösung erwarten.

Im Bereich der semantischen Dienste planen etliche Konsortien, Wissensgraphen einzusetzen, und auch das Framework Wikibase²⁵ wird in und von diversen Konsortien zur Semantifizierung vorgeschlagen.

²⁰<https://ckan.org/>

²¹<https://protege.stanford.edu/>

²²<https://chemotion.net/>

²³<https://www.radar-service.eu/de>

²⁴<https://www.openarchives.org/pmh/>

²⁵<https://www.wikiba.se/>

Es besteht ein weitreichender Konsensus, dass die Heterogenität der Anforderungen für die technischen Aspekte und Infrastruktur eine große Herausforderung darstellt. Es sollten so viele Aspekte wie möglich harmonisiert werden, die allerdings auch ausreichende Freiheiten für fachspezifische Lösungen ermöglichen: „*Harmonisieren so viel wie möglich, Diversität wenn nötig*“. Ein weiterer Informationsaustausch zwischen den Konsortien zu diesen Themen wurde ausdrücklich gefordert.

6 Diskussionsraum *Organisatorische Aspekte und Zusammenarbeit*

Ziel dieser Session war es, Ansätze zur Kommunikation und Zusammenarbeit zwischen Konsortien sowie weiteren Stakeholdern darzustellen und Konzepte hierzu zu erfassen. Dazu wurden die Teilnehmenden zu ihrer Zusammenarbeit innerhalb ihres Konsortiums, zu Möglichkeiten für die Zusammenarbeit zwischen den Konsortien, ihren Vorschlägen bzw. Anregungen an das NFDI-Direktorat und ihren (inter-)nationalen Netzwerken befragt, wobei sich alle Fragen auf die (Weiter-) Entwicklung von Metadaten/Ontologien/Standards bezogen.

Einige Konsortien nutzen Use Cases für den Umgang mit Metadaten und den geplanten Entwicklungen. Sie eignen sich als konkrete und anschauliche Beispiele in der Kommunikation zwischen den Forschenden; ein Einbeziehen von Gruppen mit unterschiedlich weit entwickelten Kenntnisständen und Fähigkeiten im Bereich Metadaten wird erleichtert. Eine Übertragung auf eine konsortienübergreifende Kommunikation wäre denkbar, um z. B. gemeinsame Anknüpfungspunkte in der Bearbeitung von Ontologien und Möglichkeiten der Interoperabilität zu identifizieren und nutzen. So könnten gemeinsam Basiskriterien entwickelt werden, aber auch Technologien für die Qualitätsüberprüfung (wie z. B. OBO Dashboard²⁶). Zusätzlich bilden Use Cases die technische Landschaft exemplarisch ab und dienen als Test für die Infrastruktur.

Neben notwendigen Neuentwicklungen im Bereich Metadaten sind Nachnutzung und Weiterentwicklung von etablierten Standards und Konventionen essentiell, insbesondere um eine breite Akzeptanz und Anwendung in den Communities zu gewährleisten. Die Konsortien setzten sich das Ziel, die Vorzüge einer koordinierten Metadatenerfassung zu vermitteln, z. B. durch Schulungen für den wissenschaftlichen Nachwuchs und im engen Kontakt mit Fachverbänden und Communities, mit Informationsständen auf Konferenzen sowie Fragebögen und Workshops zu Metadaten und Best Practices. Auch ein kreativer Umgang in der Sensibilisierung der Metadatenproblematik kann den Erfolg fördern: Nachwuchsforschende standen einer Veranstaltung zu „Fragen an Forschungsdaten“ offener gegenüber, als wenn zunächst explizit von Metadaten gesprochen wurde.

Die Etablierung von Fachpersonal wie z. B. Data Stewards kann die Kommunikation mit den Fachbereichen und letztendlich auch Änderungen im Wissenschaftssystem

²⁶<http://obo-dashboard-test.ontodev.com/about.html>

im Forschungsdatenmanagement befördern. Hier planen Konsortien unterschiedliche Ansätze: Data Stewards o. ä. Personen unterstützen in den Instituten vor Ort die Forschenden und bilden ein Bindeglied zu entsprechenden Gremien für die Entwicklung von Metadatenstandards, Ontologien etc. oder es gibt spezielle Datenzentren mit Expertinnen und Experten, an die sich Forschende wenden können.

Trotz der großen Diversität der NFDI-Konsortien gibt es Überschneidungen im Bereich Metadaten, wodurch konsortienübergreifende Kooperationen sinnvoll sein können und auch geplant werden. Bereits die DFG-Vorlage zu Metadaten hat verschiedentlich zu Abgleichen zwischen den Konsortien geführt und es entstand die Leipzig-Berlin-Erklärung.²⁷ Konsortienübergreifend sollte eruiert werden, was im Bereich Metadaten realistisch angegangen werden kann, besonders mit Blick auf internationale Initiativen. Die Zusammenarbeit auf möglichst vielen Ebenen, vom Promovierenden bis zur Leitung wurde von den Teilnehmenden als sinnvoll eingeschätzt.

Eine Zusammenarbeit mit dem und Unterstützung durch das Direktorat wird begrüßt. Die Teilnehmenden benannten mehrere Möglichkeiten der Unterstützung, wie z. B. ein allgemeiner Einstiegspunkt für eine NFDI-Austauschplattform, um einen Überblick über vorhandene Kompetenzen und mögliche Kontaktpersonen zu erhalten und sich untereinander über bewährte, aber auch abgelehnte Konzepte und Tools auszutauschen. Eine Übersicht der Fachdisziplinen und deren Relevanz für die Konsortien kann gewährleisten, dass jeweils relevante Fachbereiche eingebunden werden. High level guidelines können eine Basis für den Aufbau der fachspezifischen Systeme der Konsortien bilden, z. B. wird eine Unterstützung in allgemeinen Bereichen (RDF²⁸, SKOS²⁹, OWL³⁰ etc.), die Nutzung eines allgemeinen Metadatenschemas (z. B. Dublin Core³¹) und Empfehlungen oder konsortiumsübergreifende Festlegungen von gewissen Standards (z. B. bei Schnittstellen) gewünscht. Neben gemeinsam entwickelten Empfehlungen ist ein ausreichender Freiraum mit einem pragmatischen Ansatz im Umgang mit Metadaten auf der Arbeitsebene für die Konsortien essentiell. Der Aufbau der NFDI ist ein iterativer Prozess und dies sollte sich auch in einer flexiblen Kommunikationsstruktur aus Arbeitsgruppen, Workshops etc. spiegeln. Das Direktorat kann die erarbeiteten Ergebnisse im Bereich Metadaten auf (inter-)nationaler Ebene verbreiten und damit zu einer Sensibilisierung für diese Thematik im Wissenschaftssystem beitragen.

²⁷Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung: <https://zenodo.org/record/3895209>

²⁸<http://www.w3.org/RDF/>

²⁹<https://www.w3.org/2004/02/skos/>

³⁰<https://www.w3.org/TR/owl-semantic/>

³¹<http://dublincore.org/>

7 Wie könnte es weitergehen

In der abschließenden Diskussion kam einerseits der Wunsch auf, vorhandene Handlungsspielräume zu nutzen, andererseits auch das Bedürfnis, zunächst einmal systematisch die Anwendungsspektren einer Metadatierung und die vorhandenen Rollen zusammenzustellen und daraus die weitere Zusammenarbeit zu entwickeln.

Gerade zu Beginn der Formierung der NFDI eröffnen sich Freiheitsgrade, die genutzt werden können und sollen, um Einigung über die Fragen der Metadatierung zu erreichen, die für alle Konsortien von Interesse sind. Konkret genannt wurden in diesem Bereich die Personenidentifikation und die maschinenlesbare Dokumentation von rechtlichen Informationen wie Lizenzen. Hier gilt es die Balance zu halten zwischen der Einbeziehung zukünftiger Konsortien und der Wahrung der Handlungsfähigkeit.

Gleichzeitig besteht die Gefahr, das Thema Metadaten primär informationswissenschaftlich zu behandeln und das eigentliche Ziel aus dem Auge zu verlieren: Die Unterstützung der fachwissenschaftlichen Forschung. Mindestens genauso wichtig wie schnelles Handeln ist also die klare Definition der Ziele und Anwendungen der Metadatierung und die notwendigen Rollen im Prozess der Weiterentwicklung, hier insbesondere die Zusammenarbeit von Metadaten-expert:innen und Fachwissenschaftler:innen über Use Cases.

Strukturen zu schaffen, die die Zusammenarbeit an interoperabel nutzbaren Dokumentationsschemata, -standards und -mappings in verschiedenen Zusammensetzungen und auf unterschiedlichen Ebenen befördern, wird eine große, aber entscheidende Aufgabe der nächsten Monate und Jahre sein.

Das Thema Metadaten mit seinen einerseits inhaltlichen wie auch infrastrukturellen Facetten spielt nicht nur in der NFDI eine zentrale Rolle, sondern auch in anderen internationalen wie auch nationalen Initiativen wie z. B. in der Research Data Alliance (RDA)³², European Open Science Cloud (EOSC)³³, Australian Research Data Commons (ARDC)³⁴, Helmholtz Metadata Collaboration Platform (HMC)³⁵ sowie in unzähligen Forschungsinfrastrukturprojekten und Forschungsdatenmanagementteams der Institutionen. Neben der perspektivischen Einbettung der NFDI in EOSC erscheint eine internationalisierte Diskussion zu Metadatenaktivitäten mit Ausrichtung auf Policies und Standards schon am jetzigen Startpunkt der NFDI dringend notwendig. Als entsprechend ausgerichtete Plattform, um Kommunikation und harmonisierte sowie interoperable Standards international zu fördern, bietet sich hier die Research Data Alliance mit ihren ca. 11.000 Forschungsdatenexpertinnen und -experten an.³⁶

³²<https://www.rd-alliance.org/>

³³<https://www.eosc-portal.eu/>

³⁴<https://ardc.edu.au/>

³⁵<https://helmholtz-metadaten.de/>

³⁶Als Beispiel lässt sich hier die online-Seminarreihe hervorheben, die von der Interest Group (IG) *Research Data Management in Engineering* (RDMinEng) organisiert wurde und praktische Beispiele und Tools zum Thema Metadaten in den Ingenieurwissenschaften zeigte. Auf dem Plenary P16 der RDA wird eine Joint Session der IG RDMinEng zusammen mit der IG Metadata angeboten (<https://www.rd-alliance.org/joint-meeting-ig-rdmineng-and-ig-metadata>).

Die konkrete Zusammenarbeit am Thema Metadaten innerhalb der NFDI soll in Form von weiteren Workshops ausgestaltet werden. Der nächste Workshop wird von NFDI4Culture, DataPLANT und NFDI4Health organisiert.