

METADATENSCHEMATA FÜR FORSCHUNGSDATEN – GENERISCHE STANDARDS UND SPEZIFIKA IN DER BIOLOGIE UND DEN INGENIEURWISSENSCHAFTEN

Katharina Markus

ORCID: 0000-0002-9316-8982

ZB MED – Informationszentrum Lebenswissenschaften¹

markus@zbmed.de

1. Einleitung

Die Digitalisierung verändert die Forschungslandschaft grundlegend, indem Forschungsmethoden und Ergebnisse zunehmend digitale Form annehmen.² Neuerdings sollen die Forschungsergebnisse, auch Forschungsdaten (FD), als Teil einer Open-Science-Strategie in großem Umfang für die Öffentlichkeit und Wissenschaft zugänglich und nachnutzbar gemacht werden³ und so zu innovativer, effizienter und transparenter Forschung beitragen.⁴ Für langfristige Zugänglichkeit und Nachnutzbarkeit werden beschreibende Informationen zu den FD, die Metadaten, sowie zugehörige Standards wie Metadatenschemata als essenziell angesehen.⁵ Sie sollen FD entsprechend den FAIR-Prinzipien auffindbar (Findable), zugänglich (Accessible), interoperabel (Interoperable) und wiederverwendbar (Reusable) machen, besonders im Hinblick auf die

¹ Dieser Artikel basiert auf einer Hausarbeit, die während des Bibliotheksreferendariats an der Universitätsbibliothek Stuttgart und der Bibliotheksakademie der Bayerischen Staatsbibliothek entstanden ist.

² Howe u. a. (2008, S. 47).

³ Rat für Informationsinfrastrukturen (2019, S. 38), Whyte & Pryor (2011, S. 207).

⁴ Wilkinson u. a. (2016, S. 1), Alexander Ball, Darlington, Howard, McMahon & Culley (2012).

⁵ Wilkinson u. a. (2016, S. 5).

maschinelle Verarbeitung.⁶ In der Biologie können FD beispielweise Genomdaten⁷ oder Arten-Erhebungsdaten⁸ sein. Der Austausch von Daten, die Verwendung von Datenbanken für die Veröffentlichung dieser Daten, sowie eigene Metadatenstandards sind in dieser Disziplin fest etabliert.⁹ Darüber hinaus haben sich verschiedene biologische Institutionen herausgebildet, die sich um Standardisierung bemühen.¹⁰ Besonders in der biologischen Teildisziplin der Genom-Forschung herrscht eine Kultur des Datenaustausches, ebenso wie in den Sozialwissenschaften, der Astronomie, den Geowissenschaften und einzelnen interdisziplinären Bereichen wie Polar-Wissenschaften.¹¹ In den Ingenieurwissenschaften findet dagegen traditionell weniger Datenaustausch statt.¹² Hinzu kommt, dass die Daten dieses Fachgebiets eine große Diversität aufweisen und oft aus interdisziplinärer Forschung stammen.¹³ Dies ist ersichtlich durch ingenieurwissenschaftliche Daten beispielsweise aus Computersimulationen¹⁴, Ergebnisse von Computertomografie oder von Zugfestigkeitstests.¹⁵

Eine qualitativ hochwertige Datendokumentation stellt aufgrund der Komplexität und des Aufwandes eine Herausforderung für Wissenschaftler¹⁶ dar.¹⁷ Von Fach-

⁶ Wilkinson u. a. (2016, S. 4–5) Die GO FAIR-Initiative beschäftigt sich mit der Implementierung der FAIR-Datenprinzipien (GO FAIR - GO FAIR Initiative).

⁷ Howe u. a. (2008, S. 47).

⁸ Hugo, Hobern, Kõljalg, Tuama & Saarenmaa (2017, S. 283–284).

⁹ Scott u. a. (2014, S. 38), Howe u. a. (2008).

¹⁰ Howe u. a. (2008), McQuilton u. a. (2019).

¹¹ Austin u. a. (2017, S. 78).

¹² Alex Ball & Neilson (2010, S. 5), Alexander Ball u. a. (2012).

¹³ Alexander Ball u. a. (2012), Howard, Darlington, Ball, Culley & McMahon (2010a, S. 19–21), Scott u. a. (2014, S. 36–37).

¹⁴ Iglezakis & Schembera (2018, S. 48–49).

¹⁵ Scott u. a. (2014, S. 36) An den genannten Beispielen lässt sich die Interdisziplinarität der Ingenieurwissenschaften erkennen. Sie entsteht, da ingenieurwissenschaftliche Forschung praxis-orientiert ist und dabei thematisch in einem Bereich wie beispielsweise Medizin verankert sein kann.

¹⁶ Mit „Wissenschaftler“ sind an dieser Stelle auch Wissenschaftlerinnen angesprochen. Für die bessere Lesbarkeit wird im Folgenden ausschließlich die maskuline Form verwendet. Trotzdem sind ausdrücklich Menschen aller Geschlechter in diesen Formulierungen inbegriffen.

¹⁷ Castro u. a. (2017, S. 183), Rat für Informationsinfrastrukturen (2019, S. 38), Hermann, Hahn, Gärtner & Fritze (2018, S. 33), Tristram u. a. (2015, S. 74–82).

zeitschriften und Forschungsförderern wird sie dagegen in Form von Datenveröffentlichung in Datenrepositorien und als Datenmanagementpläne¹⁸ zunehmend erwartet.¹⁹ Hier können Bibliotheken als die Wissenschaft unterstützende Infrastrukturen im Bereich Forschungsdatenmanagement einen wichtigen Beitrag leisten und durch Erfahrungen mit Metadaten²⁰ und Repositorien²¹ bei der Auswahl, Erarbeitung und Anwendung entsprechender Standards unterstützen.²²

Dieser Artikel befasst sich mit fachunabhängigen, generischen, Standards sowie Standards in den Disziplinen Biologie und Ingenieurwissenschaften. Ein Überblick über die behandelten Schemata und Standards ist in Tabelle 1 (Anhang, Tab. 1) zu finden. Darauf folgt ein Einblick in verschiedene Möglichkeiten der Integration von Metadatenstandards in Software-Anwendungen, die wiederum die Dokumentation von FD und ihrer Metadaten vereinfachen.

2. Forschungsdaten und Metadaten

Forschungsdaten lassen sich definieren als „Daten, die im Zuge wissenschaftlicher Vorhaben entstehen“²³. Sie sind äußerst divers:²⁴ sie stammen beispielsweise aus „Beobachtungen, Experimente[n], Simulationsrechnungen, Erhebungen, Befragungen, Quellenforschungen, Aufzeichnungen, Digitalisierung [und] Auswertungen“²⁵ sowie weiteren Forschungsaktivitäten²⁶ und können verschiedene Formen wie z. B. Bilder aus der Strömungsforschung, Messdaten aus der Chemie oder Computersimulationen²⁷ annehmen. Ihre langfristige Nutzbarkeit entsprechend der FAIR-Prinzipien

¹⁸ Datenmanagementpläne beschreiben den Umgang mit FD während und nach einem wissenschaftlichen Projekt und werden im Optimalfall vor Projektanfang erstellt. Oft werden strukturierte Vorlagen verwendet und ausgefüllt mit einem Projektantrag eingereicht, wie z.B. auf Forschungsdaten.info dargelegt (Forschungsdaten.info - Datenmanagementplan). In Datenmanagementplänen sollte die Langzeitarchivierung geplant sein und im Falle einer geplanten Veröffentlichung festgelegt werden, in welchem Repository diese erfolgen soll (Forschungsdaten.info - Datenmanagementplan).

¹⁹ Castro u. a. (2017, S. 182).

²⁰ Riley (2017, S. 5, 27–31), Greenberg, Swauger & Feinstein (2013, S. 141).

²¹ Simons & Richardson (2013, S. 3).

²² Rat für Informationsinfrastrukturen (2019, S. 71).

²³ Rat für Informationsinfrastrukturen (2019, S. 11).

²⁴ Castro u. a. (2017, S. 184).

²⁵ Rat für Informationsinfrastrukturen (2019, S. 11).

²⁶ Rat für Informationsinfrastrukturen (2019, S. 11).

²⁷ Johannes u. a. (2013, S. 17–24) Die genannten Bilder der Strömungsforschung werden bei Forschung am Windkanal aufgezeichnet und auf High Performance Computern (HPC) ausgewertet (Johannes

(siehe 2.1) erfordert Maßnahmen wie z. B. ihre Veröffentlichung. Diese Maßnahmen werden unter dem Begriff „Forschungsdatenmanagement“ (FDM) zusammengefasst.²⁸ Speziell die Veröffentlichung von Forschungsdaten lässt sich definieren als das Einstellen von Forschungsdaten und zugehörigen Metadaten ins Internet mit der Absicht, diese langfristig nachnutzbar zu machen.²⁹ Inzwischen wird zunehmend eine Veröffentlichung der FD in Datenrepositorien erwartet³⁰, in denen die zugehörigen beschreibenden Metadaten der FD auch strukturiert vorliegen können.³¹

Metadaten werden als „Daten über Daten“³² definiert. Sie haben die Funktion „der Beschreibung von Daten“³³ und der Unterstützung von „[...] Datennutzer[n] bei der Recherche nach Daten, bei der Bewertung der Eignung recherchierter Daten für die eigenen Zwecke und bei der Integration gefundener Daten in die eigene Systemumgebung“³⁴ zu dienen. Die Standardisierung der Datenstruktur unterstützt unter anderem die maschinelle Suche, indem Metadaten verschiedener Datensätze in Beziehung zueinander gesetzt, indexiert und so von Dritten, wie Datenbanken und Suchmaschinen, automatisiert aggregiert werden können.

Metadatenschemata helfen bei der Standardisierung von Metadaten, indem sie einen Satz an Metadatenelementen definieren, die Attribute einer Ressource repräsentieren³⁵ (z. B. das Attribut *Autor* der Ressource *Datensatz*; das Attribut wird in diesem Fall durch das gleichnamige Metadatenelement *Autor* repräsentiert). Metadatenschemata, die oft im XML-Format vorliegen³⁶, können wiederum in Datenrepositorien und

u. a. (2013, S. 17)), die Messdaten können aus der organischen Chemie stammen und in elektronischer Form vorliegen (Johannes u. a. (2013, S. 22)) und die Computersimulation können Proteinen und Bewegungsabläufen in der Biophysik darstellen (Johannes u. a. (2013, S. 22)). Ein weiteres Beispiel für FD sind Pulshöhenspektren der Neutronenspektroskopie (Johannes u. a. (2013, S. 23–24)).

²⁸ Rat für Informationsinfrastrukturen (2019, S. 11).

²⁹ Austin u. a. (2017, S. 82) Austin und ihre Mitarbeiter spezifizieren für die Definition von FD-Publikationen, dass bei diesen die zu den FD gehörende Dokumentation und, falls verwendet, der zugehörige Softwarecode zusammen mit den FD veröffentlicht werden. Dabei sollen durch die Nutzung von Daten-Repositorien und (Daten-)Journals die publizierten Objekte gut dokumentiert, kuratiert, langfristig archiviert interoperabel, zitierbar, auffindbar und von gesicherter Qualität sein (Austin u. a. (2017, S. 82)).

³⁰ Candela, Castelli, Manghi & Tani (2015, S. 1755).

³¹ Castro u. a. (2017, S. 187–189).

³² Rat für Informationsinfrastrukturen (2019, S. 16), Riley (2017, S. 1).

³³ Rat für Informationsinfrastrukturen (2019, S. 16).

³⁴ Rat für Informationsinfrastrukturen (2019, S. 16).

³⁵ Chamnongsri (2019, S. 244).

³⁶ Riley (2017, S. 16–17) Neben XML können Metadatenschemata auch als RDF modelliert werden, vgl. Riley (2017, S. 16–17) und, unter anderem, im RDF/XML-Format vorliegen (Riley (2017, S. 12–14)).

Datenbanken implementiert werden.³⁷ Definiert ist ein *Schema* nach ISO 23081.1³⁸ als ein logischer Plan, der die Beziehungen zwischen Metadatenelementen aufzeigt. Der Plan beinhaltet generell festgelegte Regeln der Nutzung und des Managements von Metadaten bezüglich der Semantik, der Syntax und der Optionalität (Pflicht-Level) von Werten.³⁹ Jedes *Element* kann dabei einen Wert oder mehrere Werte annehmen⁴⁰ (z. B. der entsprechende Name des Autors für das Element *Autor*).

Metadatenmodelle und Anwendungsprofile⁴¹ sind weitere Standardisierungs- und Dokumentationsmöglichkeiten. In *Metadatenmodellen* werden Metadatenstandards grafisch dargestellt.⁴² *Metadaten-Anwendungsprofile* (application profiles) dienen ursprünglich zur Zusammenstellung von Metadatenelementen aus einer oder mehreren Quellen zur optimierten lokalen Anwendung.⁴³ Sie können für die Elemente auch zugehörige Regelwerke, Richtlinien⁴⁴ und ein standardisiertes Vokabular bzw. die Semantik festlegen.⁴⁵ Die standardisierte Semantik kann für die Beschreibung von Inhalten verwendet werden und beinhaltet kontrollierte Vokabulare (z. B. Library of Congress Subject Headings⁴⁶) und Ontologien.⁴⁷ *Ontologien* sind Abstraktionen von Datenmodellen und Wissensdomänen (z. B. die Wissensdomäne *Genetik*), indem repräsentative Archetypen, typischerweise Klassen, Attribute und die Beziehungen dieser zueinander,

³⁷ Castro u. a. (2017, S. 187–189).

³⁸ ISO 23081.1 ist die ISO-Norm für „Information und Dokumentation - Metadaten für Verfahren der Schriftgutverwaltung – Teil 1: Grundsätze“ (Deutsches Institut für Normung e.V. (DIN) - ISO 23081.1.).

³⁹ ISO 23081.1, Teil 1 Terms and Definitions, 3., zitiert nach National Information Standards Organization o. J., S. 2]. Obwohl die ISO-Norm für die Schriftgutverwaltung erstellt wurde, lassen sich ihre grundlegende Komponenten auch für Forschungsdaten verwenden. Der Originaltext lautet: „*A schema is a logical plan showing the relationships between metadata elements, normally through establishing rules for the use and management of metadata specifically as regards the semantics, the syntax and the optionality (obligation level) of values.*“ (ISO 23081.1, Teil 1 Terms and Definitions, zitiert nach National Information Standards Organization o. J.]. Weitere ISO-Standards im Zusammenhang mit der Forschungsdaten-Qualität werden von dem Rat für Informationsinfrastrukturen evaluiert (Rat für Informationsinfrastrukturen (2019, S. A-6-A-8)). Mit Ausnahme der in diesem Artikel erwähnten Dublin Core ISO-Norm 15836 scheint keine dieser Normen auf generische Forschungsdaten direkt anwendbar. Spezifisch für Metadaten zu geografischen FD liegt eine ISO-Norm vor (International Standard Organisation (ISO) - ISO 19115-1:2014).

⁴⁰ Chamnongsri (2019, S. 244).

⁴¹ Thalhath, Nagamori, Sakaguchi & Sugimoto (2019, S. 116), Riley (2017, S. 24).

⁴² Coyle (2017, S. 4) Wie bei Coyle 2017 auf Seite 4 dargestellt, stellen Metadatenmodelle ihre Elemente (wiederum z. B. *Autor*) und deren Beziehungen untereinander visuell dar während Regeln für z. B. die Anwendung nicht enthalten sind. Schemata sind dagegen Text-basiert.

⁴³ Heery & Patel (2000).

⁴⁴ Thalhath u. a. (2019, S. 116).

⁴⁵ Thalhath u. a. (2019, S. 116), Taylor u. a. (2008, S. 890).

⁴⁶ Riley (2017, S. 17).

⁴⁷ Caracciolo, Aubin, Whitehead & Zervas (2019, S. 340–341).

definiert werden.⁴⁸ Werte (bei Ontologien „Instanzen“) sind generell nicht Teil der Ontologie.⁴⁹ Ihre Verwendung in Metadatenschemata erlaubt, Deskriptoren und deren Beziehungen untereinander maschinenverwendbar zu machen.⁵⁰ Durch beispielsweise die Nutzung der Gene Ontology⁵¹ mit der hinterlegten Genfunktion „DNA recombination“⁵² können zwei Gene mit dieser Funktion in unterschiedlichen Datensätzen sinnvoll gruppiert werden. Datensätze können so verlinkt und verglichen werden und durch die hinterlegten Beziehungen der Klassen untereinander auch in Beziehung zueinander gesetzt werden.

Eine der möglichen Ontologie-Sprachen ist das *Resource Description Framework (RDF)*⁵³, das in dieser Funktion die Modellierung von Informationen unterstützt. RDF ermöglicht Informationsstrukturierung im semantischen Web⁵⁴, indem es Informationen in Triplets (z. B. *Gen X* (Subjekt) *gehört zu* (Prädikat) *Gengruppe Y* (Objekt))⁵⁵ modelliert. Diese Informationen werden auf einer Webseite hinterlegt, die den Datensatz von z. B. *Gen X* darstellt. Als Resultat können die Informationen so von (Such-) Maschinen verwertet werden. Bei der Suche nach *Gen X* kann dann z. B. auch automatisch die *Gengruppe Y* berücksichtigt werden.

In diesem Artikel wird der Fokus auf Metadatenschemata gelegt, allerdings werden aufgrund der thematischen Überschneidung ebenfalls allgemein Metadatenstandards, unter anderem Metadatenmodelle, Metadaten-Anwendungsprofile und Ontologien, berücksichtigt.⁵⁶

⁴⁸ Gruber (2009, S. 1963).

⁴⁹ Arroyo & Siorpaes (2014, S. 147).

⁵⁰ Castro u. a. (2017, S. 186).

⁵¹ Ashburner u. a. (2000).

⁵² Ashburner u. a. (2000, S. 26).

⁵³ Arroyo & Siorpaes (2014, S. 149–150).

⁵⁴ Das Semantische Web ist eine Anreicherung des Internets mit Maschinen-verwendbarer Semantik durch die Verwendung von Metadaten als semantische Annotationen, Ontologien für die Beschreibung von Informationen im Internet und die Verbindungen von Annotationen mit Ontologien (Antoniou & Plexousakis (2009, S. 2579)). Innerhalb des semantischen Webs soll es möglich sein, die Sammlung von Metadaten und Ontologien zu prozessieren und abzufragen, indem Logik-basierte Techniken verwendet werden (Antoniou & Plexousakis (2009, S. 2579)).

⁵⁵ Riley (2017, S. 9–15) RDF wurde von dem World Wide Web Consortium (W3C) für die Spezifizierung von Metadaten im semantischen Web anerkannt (Riley (2017, S. 10)).

⁵⁶ Sansone und Mitarbeiter haben diese thematischen Gruppen ebenfalls unter Metadatenstandards zusammengefasst, indem sie Minimum-Einreichungsleitlinien (minimum reporting guidelines), Terminologie-Artefakte (terminology artefacts) wie Ontologien, Modelle und Formate (models and formats) wie Metadatenschemata sowie Identifikator-Schemata (identifier schemata) unterscheiden (Sansone u. a. (2019, S. 359)).

Standardisierungsbestrebungen finden sich nun in verschiedenen Wissenschaftsbereichen⁵⁷ und können bottom-up, beispielsweise von Forschergruppen, oder top-down, von nationalen Standardisierungsstellen, ausgehen.⁵⁸ Nicht nur Institutionen, die Standards erstellen, sondern auch Services, die Standards aggregieren und zugänglich machen, haben sich herausgebildet. Zu diesen gehören der *Metadata Standards Catalog (MSC)* und *FAIRsharing*, die sich aus der Notwendigkeit heraus entwickelt haben, eine Übersicht über eine große Anzahl⁵⁹ von Metadatenstandards und -schemata zu vermitteln.⁶⁰ Der *MSC* wurde von der Metadata Standards Directory Working Group der *Research Data Alliance (RDA)*⁶¹ aufgebaut und basiert selbst auf diesem Directory.⁶² Er enthält unter anderem Metadatenstandards und wird von Freiwilligen unter der Aufsicht der RDA Metadata Standards Catalog Working Group betreut.⁶³ Er lässt sich daher als Service mit der RDA assoziieren. Die RDA ist eine internationale Initiative, die von der Europäischen Union, den USA und der australischen Regierung mit der Absicht unterstützt wird, bottom-up und kollaborativ den offenen Austausch von FD zu fördern.⁶⁴ Sie entstand aus der Wissenschaft heraus⁶⁵ und bietet eine Plattform für Interessens- und Arbeitsgruppen, in denen Wissenschaftler, Bibliothekare und Interessierte aus weiteren Bereichen⁶⁶ Werkzeuge, Standards, Best Practices und ähnliche Themen im Zusammenhang mit FD bearbeiten.⁶⁷

Der zweite Service, *FAIRsharing*, begann als „Minimum Information about a Biomedical or Biological Investigation (MIBBI) portal“ und BioSharing.⁶⁸ Er wird

⁵⁷ Rat für Informationsinfrastrukturen (2019, S. 16–18).

⁵⁸ Rat für Informationsinfrastrukturen (2019, S. 27), Brooksbank & Quackenbush (2006, S. 96–97)

⁵⁹ Sansone u. a. (2019, S. 358).

⁶⁰ FAIRsharing: Sansone u. a. (2019, S. 359), MSC: Alexander Ball u. a. (2014, S. 143).

⁶¹ Research Data Alliance - <https://www.rd-alliance.org/>.

⁶² Alex Ball, Greenberg, Jeffery & Koskela (2016, S. 16), RDA Metadata Standards Directory - <http://rd-alliance.github.io/metadata-directory/>.

⁶³ Metadata Standards Catalog - Terms of use.

⁶⁴ Rat für Informationsinfrastrukturen (2019, S. 17–18), Research Data Alliance - About RDA.

⁶⁵ Rat für Informationsinfrastrukturen (2019, S. A-11).

⁶⁶ Parsons (2013), Research Data Alliance - About RDA.

⁶⁷ Alexander Ball u. a. (2014, S. 143), Parsons (2013), Berman (2019, S. 1).

⁶⁸ Sansone u. a. (2019, S. 366), FAIRsharing - Communities.

nach eigenen Angaben von Daten-Forschern („data researcher“⁶⁹), Forschungssoftware- und Wissensingenieuren („research software und knowledge engineers“⁷⁰) betrieben, die aus verschiedenen Wissenschaftsdisziplinen stammen.⁷¹ Im Februar 2020 verzeichnete FAIRsharing 1376 Standards.⁷² Hinzu kommen Datenbanken, Richtlinien und Sammlungen.⁷³ Da für FAIRsharing ebenfalls eine zugehörige RDA-Arbeitsgruppe existiert⁷⁴ und die Initiative in Austausch mit den großen Wissenschaftsverlagen (Elsevier, Springer Nature BioMed Central, Springer Nature Scientific Data, Taylor & Francis, Wiley und PLOS) steht⁷⁵, scheint dieser Service stark in die Wissenschaftsgesellschaft integriert zu sein.

2.1 Fachunabhängige Standards

Bestrebungen, Metadaten für FD zu standardisieren, zielen auf die Etablierung von Qualitätskriterien für Metadaten entsprechend den FAIR-Datenprinzipien ab. Diese sind vergleichsweise allgemein formuliert, wodurch sie im FDM-Bereich weitläufig angenommen wurden.⁷⁶ In den FAIR-Prinzipien ist in Bezug auf Metadaten unter anderem enthalten: umfangreiche Metadaten, die zutreffende und relevante Attribute enthalten, eine formale, zugängliche, öffentlich geteilte und breit anwendbare Sprache nutzen, den fachgebietsrelevanten Community-Standards folgen sowie ein Vokabular, das den FAIR-Prinzipien folgt.⁷⁷ Existierende Metadatenschemata setzen die FAIR-Prinzipien nicht unbedingt vollständig um, unterstützen aber auf jeden Fall durch die

⁶⁹ Data Readiness Group - People.

⁷⁰ Data Readiness Group - People.

⁷¹ FAIRsharing - Communities, Data Readiness Group - People, Sansone u. a. (2019, S. 358).

⁷² FAIRsharing - Standards: Zu beachten ist, dass in diesen auch nicht mehr aktive Standards inbegriffen sind.

⁷³ FAIRsharing - <https://fairsharing.org/>.

⁷⁴ Research Data Alliance - Arbeitsgruppe FAIRSharing Registry: connecting data policies, standards & databases.

⁷⁵ FAIRsharing - Communities, Sansone u. a. (2019, S. 262).

⁷⁶ Rat für Informationsinfrastrukturen (2019, S. 27).

⁷⁷ Wilkinson u. a. (2016, S. 4), Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek (IITB) - Die FAIR Data Prinzipien für Forschungsdaten. Wilkinson und seine Mitarbeiter nennen als Voraussetzungen für Daten, die den FAIR-Standards entsprechen, über die im obigen Text genannten Bedingungen hinaus: einen einzigartigen und persistenten Identifikator, Datennutzungs-Lizenzen, sowie Provenienz-Informationen und eventuelle Beziehungen zu anderen Daten oder Metadaten, Metadaten, die frei verfügbar sind, den eindeutigen Identifier der zugehörigen FD enthalten, sich in einem indizierten und durchsuchbaren Verzeichnis befinden und auf die mit einem standardisierten, offenen und frei implementierbaren Kommunikationsprotokoll, das auch die Authentifizierung und die Rechteverwaltung unterstützt, zugegriffen werden kann (Wilkinson u. a. (2016, S. 4)).

Standardisierung von Metadaten die maschinelle Interoperabilität. Zusätzlich enthalten sowohl FAIR-Prinzipien wie auch Datenmanagementpläne den Gedanken der Langzeitarchivierung. Diese wird allerdings oft allein als Abgabe der Daten an ein FD-Repository mit perspektivisch gesicherter Existenz verstanden.⁷⁸ Der Aspekt des aktiven Risikomanagements und die Ausrichtung auf zukünftige Datennutzer wird durch die neuen TRUST-Prinzipien (Transparency, Reponsibility, User Focus, Sustainability, Technology) für Repositorien eingebracht.⁷⁹ Speziell in Bezug auf Metadaten wird die Nutzung von Metadaten- und Kurationsstandards der jeweiligen Datennutzer-Community hervorgehoben.⁸⁰ Im Fall der Genomforschung kann das z. B. das Genbank Sequence Format sein (siehe auch 3.1 und Anhang, Tab. 4), das Vorgaben zur Formatierung von Sequenzierungsdaten und zugehörigen Metadaten enthält.⁸¹

Im Bereich der FD existieren verschiedene disziplinenunabhängige Metadatenstandards, die unterschiedlich stark etabliert sind. Ein sehr weitverbreitetes Metadaten-schema ist das *Dublin Core* Metadatenchema.⁸² Dublin Core definiert 15 Simple Dublin Core Elements, z. B. Titel einer Ressource (title) und Ersteller einer Ressource (creator)⁸³ (Anhang, Tab. 2), die auch in einer ISO-Norm hinterlegt sind⁸⁴ und als Mindestaustauschstandard für das OAI-PMH-Protokoll dienen.⁸⁵ Darüber hinaus können Bedingungen (qualifier, Qualified Dublin Core⁸⁶) benutzt werden, um die Bedeutung eines Elements einzugrenzen oder um mithilfe der Bedingungen ein Codierungsschema für Dublin Core festzulegen.⁸⁷ Für die Nutzung im Semantic Web liegt Dublin Core im RDF-Format (siehe Kapitel 2) vor.⁸⁸ Interoperabilität mit anderen Schemata

⁷⁸ Lindlar, Rudnik, Jones & Horton (2020, S. 19).

⁷⁹ Lin u. a. (2020).

⁸⁰ Lin u. a. (2020, S. 3).

⁸¹ Benson, Karsch-Mizrachi, Lipman, Ostell & Sayers (2011).

⁸² Rat für Informationsinfrastrukturen (2019, S. 17), Caverlee, Mitra & Laarsgard (2009, S. 947–949), Austin u. a. (2017, S. 90).

⁸³ Caverlee u. a. (2009, S. 948),

⁸⁴ Caverlee u. a. (2009, S. 947) Der Dublin Core ISO-Standard ISO 15836 ist auch bei der ISO-Organisation zu finden: International Organization for Standardization - ISO 15836-1:2017 und International Organization for Standardization - ISO 15836-2:2019.

⁸⁵ Riley (2017, S. 24) OAI-PMH ist das Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), das als Minimum-Elemente Simple Dublin Core erkennt, sodass Dublin Core in vielen Repositorien implementiert wurde (Simons & Richardson (2013, S. 61)).

⁸⁶ Caverlee u. a. (2009, S. 948).

⁸⁷ Caverlee u. a. (2009, S. 948) Für Dublin Core gibt es ebenfalls Anwendungsprofile, sodass die Dublin Core-Anwendung stärker spezifiziert und damit gegebenenfalls bei breiter Anwendung dieser Anwendungsprofile verstärkt interoperabel und maschinenverwendbar ist (Coyle (2017, S. 7–11)).

⁸⁸ Caverlee u. a. (2009, S. 949).

erreicht Dublin Core, indem es die Grundlage für diese bildet, z. B. für Darwin Core.⁸⁹ Dublin Core ist allerdings als Schema für die Beschreibung von Ressourcen im Internet wie z. B. elektronische Dokumente und Audiodateien⁹⁰ konzipiert, weniger für die Beschreibung von FD.

Ähnliches gilt für *Schema.org*⁹¹, das als weiterer generischer Metadatenstandard und RDF-Vokabular vorliegt. Es erlaubt die Auszeichnung von Webseiten-Text anhand Schema.org-Types und -Properties. So kann diese strukturierte Information von Internet-Suchmaschinen weiterverwendet werden.⁹² Interoperabilität wird nicht nur erreicht, indem führende Suchmaschinen das Schema.org-Vokabular unterstützen⁹³, sondern auch durch das Mapping von Schema.org-Properties auf Metadatenelemente anderer Schemata und Standards.⁹⁴ Die Popularität von Schema.org ist in FD-Kreisen relativ hoch, ersichtlich durch die Zugehörigkeit zu den meist aufgerufenen (2018)⁹⁵ und meist implementierten (2019)⁹⁶ Standards gemäß der FAIRsharing-Datenbank.

Ein weiteres generisches Metadatenschema, das zunehmend an Wichtigkeit für Forschungsdaten gewinnt⁹⁷, ist *PREMIS* (PREservation Metadata Implementation Strategies). PREMIS stammt aus dem Bereich der digitalen Langzeitarchivierung und wurde 2020 gemäß des PREMIS Implementation Registries bereits von 50 Institutionen, unter anderem FD-Repositoryn, implementiert.⁹⁸ Das Schema besitzt vier Entitäten: Objekte (Objects), Ereignis (Event), Akteur (Agent), Rechte (Rights) und Umwelt (Environment) (Anhang, Abb. 1). Unter die Entitäten fallen 88 Semantic Units⁹⁹, die im Data Dictionary beschrieben werden, und auf die Informationen über ein digitales Objekt sowie Handlungen, die an diesem Objekt durchgeführt wurden, gemappt

⁸⁹ Wiczorek u. a. (2012).

⁹⁰ Caverlee u. a. (2009, S. 948), Simons & Richardson (2013, S. 59).

⁹¹ Guha, Guha, Brickley & Macbeth (2016).

⁹² Riley (2017, S. 19).

⁹³ Riley (2017, S. 19).

⁹⁴ Research Data Alliance - Arbeitsgruppe Research Metadata Schemas: Schema crosswalks visualizations.

⁹⁵ Sansone u. a. (2019, S. 359).

⁹⁶ Sansone u. a. (2019, S. 360).

⁹⁷ Lindlar u. a. (2020, S. 16–17).

⁹⁸ Lindlar u. a. (2020, S. 17), Library of Congress - PREMIS Implementation Registry.

⁹⁹ Lindlar u. a. (2020, S. 16), Zu den 88 Semantic Units gehören Core Semantic Units, die wahrscheinlich für die meisten Archive von Relevanz sind, aber trotzdem nicht als Pflichtelemente auftreten (PREMIS Editorial Committee (2015, S. 3)).

werden können. Dabei sind die Semantic Units im strikten Sinn keine Metadatenelemente, können aber ähnlich verwendet werden.¹⁰⁰ Sie lassen sich für die jeweiligen Anwendungen flexibel zusammenstellen¹⁰¹, PREMIS erlaubt aber auch die Verwendung eigener Metadatenelemente. Konformität wird in dann erreicht, indem diese Elemente auf die PREMIS Semantic Units gemappt werden.¹⁰² PREMIS ermöglicht Informationen abzulegen, die notwendig für die langfristige Verwendung von digitalen Objekten sind.¹⁰³ Dazu können z. B. Informationen zur Hardware und Software gehören, mit der das digitale Objekt dargestellt werden kann, oder Informationen zu Veränderungen des archivierten Objekts im Zuge dessen langfristiger Erhaltung.¹⁰⁴

Spezifischer zugeschnitten auf die Anforderungen von FD ist im Gegensatz zu Dublin Core, Schema.org und PREMIS das *DataCite* Metadatenchema.¹⁰⁵ Das Schema legt sechs Pflichtangaben fest: Identifier (DOI¹⁰⁶), Creator, Title, Publisher, Publication Year und ResourceType¹⁰⁷ (Anhang, Tab. 3). Zu den Pflichtangaben kommen weitere empfohlene und optionale Angaben.¹⁰⁸ Es wird primär von generischen Repositorien genutzt¹⁰⁹, während fachspezifische Repositorien tendenziell zusätzlich auch fachspezifische Metadatenstandards verwenden. Da DataCite 2018 zu den 12 meist besuchten Daten- und Metadatenstandards in FAIRsharing gehörte¹¹⁰, scheint es weit verbreitet zu sein. Hinzu kommt, dass DataCite für die Vergabe von DOIs für FD die Nutzung des eigenen Metadatenchemas, bzw. Elemente des Schemas, zur Bedingung macht¹¹¹,

¹⁰⁰ Caplan (2017, S. 5), PREMIS Editorial Committee (2015, S. 3).

¹⁰¹ Caplan (2017, S. 13–14), Lindlar u. a. (2020, S. 16), PREMIS Editorial Committee (2015, S. 22–24).

¹⁰² Caplan (2017, S. 13–14), Lindlar u. a. (2020, S. 16), PREMIS Editorial Committee (2015, S. 22–14).

¹⁰³ Lindlar u. a. (2020, S. 16), PREMIS Editorial Committee (2015, S. 1).

¹⁰⁴ Lindlar u. a. (2020, S. 16).

¹⁰⁵ DataCite Metadata Working Group (2019), Wilkinson u. a. (2016, S. 5).

¹⁰⁶ Als Wert des Elements *Identifier* soll der Digital Object Identifier (DOI) genutzt werden (DataCite Metadata Working Group (2019, S. 12)). DataCite selber vergibt DOIs (DataCite - Assign DOIs).

¹⁰⁷ DataCite Metadata Working Group (2019, S. 7, 12–15) Die 6 Pflicht-Angaben bzw. Eigenschaften besitzen zum Teil weitere Pflicht-Untereigenschaften (DataCite Metadata Working Group (2019, S. 7)).

¹⁰⁸ DataCite Metadata Working Group (2019, S. 8).

¹⁰⁹ Austin u. a. (2017, S. 88).

¹¹⁰ Sansone u. a. (2019, S. 359).

¹¹¹ DataCite weist darauf hin, dass DOIs mit einem Metadatenchema einher gehen („DataCite DOIs come with a metadata schema that includes a controlled vocabulary of 15 different resource types to describe the content being shared.“ (DataCite - Getting Started). Die Registrierung eines Datensatzes beinhaltet das Hochladen der zugehörigen Metadaten unter Nutzung des DataCite Metadatenchemas (Neumann & Brase (2014, S. 1038)) und beispielsweise ein Metadatenchema aus den Geowissenschaften hat aus diesem Grund DataCite-Elemente integriert (Specka u. a. (2019, S. 34)).

was ein großer Anreiz für die Verwendung dieses Schemas ist. Interoperabilität zwischen Dublin Core und dem DataCite Metadatenchema wird durch die Entwicklung eines DataCite Dublin Core Anwendungsprofils (DC2AP) unterstützt.¹¹²

Richtlinien bezüglich Metadaten werden zunehmend auch von Forschungsförderern und Verlagen von Fachzeitschriften vorgegeben¹¹³, sodass bei FAIRsharing Februar 2020 bereits 130 entsprechende Richtlinien eingetragen sind.¹¹⁴ Forschungsförderer und politische Akteure geben tendenziell eher generische Vorgaben zur Veröffentlichung von Forschungsdaten, primär die Einhaltung der FAIR-Prinzipien und die Nutzung angemessener Repositorien, vor.¹¹⁵ Wissenschaftliche Fachzeitschriften, die zum Teil Forschungsdaten im Zusammenhang mit Artikeln erwarten¹¹⁶, verweisen generell auf fachspezifische öffentliche Repositorien mit entsprechend fachspezifischen Standards.¹¹⁷ Daraus wird ersichtlich, dass Verlage und Zeitschriften grundsätzlich die Standards und Datenrepositorien nutzen, die von den Wissenschaftsdisziplinen erarbeitet werden. Um effektiven Datenaustausch zu ermöglichen, der auf den FAIR-Prinzipien und maschinenverwendbaren, detaillierten Metadaten beruht, ist folglich eine starke Involvierung der jeweiligen Wissenschaftsdisziplinen notwendig.

3. Forschungsdaten und Metadaten in der Wissenschaftscommunity

Die Wissenschaftscommunity ist äußerst vielfältig, was sich auch im FDM widerspiegelt. Neben Naturwissenschaften (im Sinne von Sciences), zu denen die Biologie gehört, führt beispielsweise der MSC fünf weitere übergeordnete Disziplinen an, wie z. B.

¹¹² DataCite Metadata Working Group (2019, S. 4)

¹¹³ Castro u. a. (2017, S. 182).

¹¹⁴ FAIRsharing - Policies: Zu beachten ist, dass sich unter diesen ausgewählten Richtlinien auch nicht mehr aktive befinden können.

¹¹⁵ Deutsche Forschungsgemeinschaft (DFG) (2019, S. 18–19).

¹¹⁶ Castro u. a. (2017, S. 182) Beispiele sind Plant Physiology (Howe u. a. (2008, S. 48)), Nature (Castro u. a. (2017, S. 182)) und PLOS ONE (Castro u. a. (2017, S. 182)).

¹¹⁷ Rat für Informationsinfrastrukturen (2019, S. 50–51) Informationen zur Veröffentlichung von Daten im Zusammenhang mit Artikeln in Plant Physiologie sind unter „Quick guide: Submission requirements“ und „Large-scale datasets and other supplemental data“ (Plant Physiology - Instructions for authors) zu finden. Für Artikel in Nature sind diese Informationen bei Supplementary Information (Nature - Supplementary information) und für Artikel in PLOS ONE unter Data Availability (PLOS - Data availability) hinterlegt.

Sozialwissenschaften oder Politik.¹¹⁸ Trotz der Vielfalt lassen sich bei der Einstellung von Wissenschaftlern gegenüber FDM zunehmend Tendenzen in Richtung Datenaustausch erkennen, wenn auch teilweise in disziplinspezifischer Ausprägung. Zunehmend ist nicht nur ein Interesse an der Weiterverwendung von veröffentlichten Daten wahrnehmbar, sondern auch ein Verständnis für die Notwendigkeit, Daten zu teilen, besonders im Zuge der Big-Data-Forschung.¹¹⁹ Gleichzeitig lastet ein hoher Publikationsdruck auf Forschern, sodass Zeit, die für das Erstellen von Datenmanagementplänen und für die arbeitsintensive Datendokumentation während des Forschungsprozesses und der Veröffentlichung der FD aufgewendet werden muss, als Bürde gesehen wird.¹²⁰

Trotzdem haben sich mit der Zeit in einzelnen Disziplinen gut kuratierte, stark integrierte Strukturen herausgebildet, die FDM ermöglichen und unterstützen,¹²¹ während sie in anderen Forschungsbereichen weniger stark etabliert sind (Kapitel 3.1 und 3.2). Bei sehr diversifizierten und auch bei fehlenden Strukturen haben Bibliotheken und andere forschungsnahe Einrichtungen die Möglichkeit, ihre Expertise einzubringen, indem sie die Entwicklung von Metadatenstandards fördern¹²², diese Standards der Forschung empfehlen¹²³ und sie in institutionellen Repositorien implementieren. Entsprechende Kooperationen mit Bibliotheken finden bereits sowohl für die Biologie als auch die Ingenieurwissenschaften statt.

3.1 Biologie

Die Biowissenschaften verändern sich ebenso wie viele andere Disziplinen in Richtung verstärkten Informationsaustausch, begründet in der Produktion von großen Datenmengen und dem zunehmenden Interesse der Forscher an diesen.¹²⁴ Diese Tendenz

¹¹⁸ Metadata Standards Catalog - Index of subjects: Die sechs Disziplinen im MSC sind „Culture“, „Education“, „Information and Communication“, „Politics“, „Law and economics“, „Science“, „Social and human sciences“. Die Gruppen „Multidisciplinary“ und „Countries and country groupings“ wurden nicht mitgezählt (Metadata Standards Catalog - Index of subjects).

¹¹⁹ Candela u. a. (2015, S. 1747–1748), Reilly, Schallier, Schrimpf, Smit & Wilkinson (2011, S. 21–22), Howe u. a. (2008, S. 47).

¹²⁰ Rat für Informationsinfrastrukturen (2019, S. 71, 73–74), Tristram u. a. (2015, S. 74–75).

¹²¹ Wilkinson u. a. (2016, S. 2).

¹²² Rat für Informationsinfrastrukturen (2019, S. 71).

¹²³ Iglezakis & Schembera (2018, S. 58).

¹²⁴ Howe u. a. (2008, S. 47).

ist vor allem in der molekularen Biologie erkennbar, im Zusammenhang mit der Genomsequenzierung, bei Daten zu Proteinen und Stoffwechselwegen, sowie in weiteren Bereichen der Molekularbiologie.¹²⁵ Neben der molekularen Biologie verfolgt auch die Biodiversitätsforschung intensive Datenauswertung¹²⁶, indem beispielsweise Verbreitungsgebiete von Arten anhand der Datenaggregation einzelner Fundorte softwarebasiert modelliert werden.¹²⁷

Die biologische Forschung lässt sich am Beispiel der biologischen Institute niedersächsischer Hochschulen 2001 in die stark vertretenen Forschungsfelder Biochemie und Biotechnologie, Genetik, Mikrobiologie, Botanik und Zoologie¹²⁸ gliedern.¹²⁹ Weitere, zum Teil neue Forschungsfelder waren 2001 die organismische Biologie mit Ökologie und Systematik¹³⁰, die Biodiversität¹³¹ und die Bioinformatik.¹³² Interdisziplinäre Überschneidungen mit anderen Disziplinen wie Medizin, Chemie und Agrarwissenschaften¹³³ sind ebenfalls möglich. Für diesen Artikel wird der Fokus auf die Biologie entsprechend der molekularen Forschung im weiteren Sinne (oben genannte primäre Forschungsfelder exklusive des nicht-molekularen Anteils der Botanik und Zoologie) und die Biodiversität gelegt.

Die molekulare Biologie besitzt entsprechend der verzeichneten Standards im MSC Februar 2020 deutlich mehr Metadatenstandards als die Biodiversität (Molekularbiologie: 18¹³⁴, Biodiversität: 5 Metadatenstandards¹³⁵). Diese Tendenz lässt sich vermutlich durch eine lange Tradition des Datenaustausches in der molekularen Biologie begründen¹³⁶, die durch die Förderung von molekularbiologischen Datenbanken für Forschungsergebnisse aus öffentlichen Geldern begünstigt wurde,¹³⁷ sowie durch die

¹²⁵ Howe u. a. (2008, S. 47).

¹²⁶ Hugo u. a. (2017, S. 263–264).

¹²⁷ Hugo u. a. (2017, S. 270–271), Graham, Loisele, Velásquez-Tibata & Cuesta (2011).

¹²⁸ Wissenschaftliche Kommission Niedersachsen (2001a, S. 15, 19, 26, 38).

¹²⁹ Zu den primären Forschungsfeldern kommen vereinzelte Schwerpunkte wie Anthropologie und Entwicklungsbiologie (S. 19), sowie Biophysik (S. 26, 38), Geobotanik (S. 26) und Geo- und Umweltwissenschaften (S. 34) in Wissenschaftliche Kommission Niedersachsen (2001a).

¹³⁰ Wissenschaftliche Kommission Niedersachsen (2001a, S. 49).

¹³¹ Wissenschaftliche Kommission Niedersachsen (2001a, S. 19, 24, 34).

¹³² Wissenschaftliche Kommission Niedersachsen (2001a, S. 24, 48).

¹³³ Wissenschaftliche Kommission Niedersachsen (2001a, S. 19).

¹³⁴ Metadata Standards Catalog - Biology

¹³⁵ Metadata Standards Catalog - Biological diversity

¹³⁶ Leonelli (2016, S. 17–24).

¹³⁷ Leonelli (2016, S. 21–24) Beispiele für Modellorganismus-Datenbanken sind FlyBase, WormBase, The Arabidopsis Information Resource (TAIR) und weitere (Leonelli (2016, S. 21)).

Selbstverpflichtung der Disziplin zum Datenaustausch.¹³⁸ Generell ist auch die Tatsache, dass FD bzw. Rohdaten keinen urheberrechtlichen Beschränkungen unterliegen, für den Datenaustausch von Vorteil.¹³⁹ Heute sind es Biologen gewohnt, eine große Menge an Daten in frei zugänglichen Datenbanken aufzufinden¹⁴⁰ und es haben sich verschiedene fachspezifische Institutionen, die Standards erarbeiten, etabliert.¹⁴¹ Dabei sind vor allem Datenbanken für Daten über standardisierte Versuchsorganismen¹⁴² (Modellorganismen, wie z. B. die Maus in der Medizin) sowie allgemeine Genom- und Protein-Datenbanken¹⁴³ und Ontologien¹⁴⁴ zu erwähnen. Als besonders erfolgreich hat sich die *Gene Ontology* erwiesen¹⁴⁵, die im Februar 2020 von 173 Datenbanken gemäß FAIRsharing implementiert wurde.¹⁴⁶ Datenkuratoren, auch Biokuratoren genannt¹⁴⁷, die selbst in erster Linie aus der Wissenschaft stammen¹⁴⁸, bereiten Datensätze und Literatur für die molekularbiologische und medizinische Forschung mit standardisierten Metadaten auf.¹⁴⁹ Auch ein Repository, das ursprünglich für naturwissenschaftliche und medizinische Daten aufgebaut wurde¹⁵⁰, Dryad, hat sich global etabliert.¹⁵¹ Aber auch in der Biodiversitätsforschung haben sich inzwischen Infrastrukturen herausgebildet, die Daten aggregieren und Standards etablieren, wie die Group on Earth

¹³⁸ Austin u. a. (2017, S. 78) Austin und Mitarbeiter nennen die Bermuda Prinzipien und die Vereinbarung des Fort Lauderdale Treffens. Besonders die Bermuda Prinzipien von 1996 (Maxson Jones, Ankeny & Cook-Deegan (2018)) hatten und hatten einen großen Impact auf das zeitnahe und öffentliche Teilen von Genom-Daten.

¹³⁹ Enke u. a. (2013), forschungsdaten.info - Urheberrecht.

¹⁴⁰ Brooksbank & Quackenbush (2006, S. 94).

¹⁴¹ Brooksbank & Quackenbush (2006, S. 94), McQuilton u. a. (2016). Zu diesen Institutionen lassen sich auch die Konsortien zählen, die sich im Zusammenhang mit den Ontologie gebildet haben (Leonelli (2016, S. 47), Ashburner u. a. (2000)) und den Austausch zwischen der Wissenschaft und politischen sowie internationalen Akteuren fördern (Leonelli (2016, S. 48)).

¹⁴² Leonelli (2016, S. 21).

¹⁴³ Scott u. a. (2014, S. 38) Die in Scott und Mitarbeiter 2014 auf Seite 38 genannten Beispiele sind die Genbank und die European Molecular Biology Laboratory Nucleotide Sequence Database (EMBL).

¹⁴⁴ Leonelli (2016, S. 26–28) Biologische und biomedizinische Ontologien sind in der Open Biological and Biomedical Ontology (OBO) Foundry hinterlegt (Open Biological and Biomedical Ontology Foundry - <http://www.obofoundry.org/>). Die OBO Foundry bringt Gruppen zusammen, die sich nach gemeinsamen Prinzipien der Ontologie-Entwicklung richten (Sansone & Rocca-Serra (2016, S. 9)).

¹⁴⁵ Leonelli (2016, S. 27), Ashburner u. a. (2000).

¹⁴⁶ FAIRsharing - GO; Gene Ontology; letzte Veränderung: 23. 09. 2019, 10:11 a.m..

¹⁴⁷ Salimi & Vita (2006).

¹⁴⁸ Leonelli (2016, S. 33)

¹⁴⁹ Leonelli (2016, S. 24–26).

¹⁵⁰ Akers & Green (2014, S. 121).

¹⁵¹ Wilkinson u. a. (2016, S. 2).

Observations (GEO)¹⁵², die Global Biodiversity Information Facility (GBIF)¹⁵³, die RDA-Interessensgruppe Biodiversity Data Integration¹⁵⁴, die Organisation Biodiversity Information Standards¹⁵⁵ und weitere.¹⁵⁶

Hinsichtlich der Popularität von allen in FAIRsharing verzeichneten Metadatenstandards gehören 2019 molekularbiologienahe Standards, neben generischen Standards, zu den meist empfohlenen.¹⁵⁷ Insgesamt sind im Februar 2020 für Biologie und Lebenswissenschaften im MSC 28 Standards hinterlegt¹⁵⁸, bei FAIRsharing für Biologie dagegen allein 398.¹⁵⁹ Diese große Anzahl zeigt, dass im Zuge dieser Arbeit eine Behandlung aller Standards nicht möglich ist. Daher wurden einzelne Standards mit großflächiger Anwendung, die sich zwar nicht Metadaten schemata nennen, aber entsprechende Funktion übernehmen, ausgewählt: das *GenBank Sequence Format* (implementiert von 30 Datenbanken¹⁶⁰, darunter die Datenbank Genbank), die *Systems Biology Markup Language (SBML)* (implementiert von 17 Datenbanken¹⁶¹) und *Darwin Core* (implementiert von 20 Datenbanken¹⁶²).

Das *Genbank Sequence Format*¹⁶³ ebenso wie die Auszeichnungssprache *SBML*¹⁶⁴ geben fast ausschließlich fachspezifische Metadaten elemente vor. Die meisten Felder des *GenBank Sequence Formats* sind genetikspezifisch: es enthält acht vorgegebene Felder

¹⁵² Group On Earth Observation - GEO Community, Lautenbacher (2006, S. 8–9).

¹⁵³ Global Biodiversity Information Facility - What is GBIF?, Gaiji u. a. (2013).

¹⁵⁴ Research Data Alliance - Interessensgruppe Biodiversity Data Integration.

¹⁵⁵ Biodiversity Information Standards - <https://www.tdwg.org/>, Hugo u. a. (2017, S. 265–266).

¹⁵⁶ Hugo u. a. (2017, S. 260–261).

¹⁵⁷ Sansone u. a. (2019, S. 359–360) Bei der Sortierung von Standards anhand der Anzahl der Seitenaufrufe auf FAIRsharing finden sich molekularbiologische Standards auf Platz 2, 3, 4, 6, 8 und 9. Der medizinische Standard auf Platz 1 kann auch für biologisch-medizinische Forschung verwendet werden (Sansone u. a. (2019, S. 359)). Bei Sortierung der Standards auf FAIRsharing anhand der Anzahl der sie implementierenden Datenbanken und Repositorien werden alle 10 höchsten Plätze außer die Plätze 4 und 8 von molekularbiologischen Standards belegt (Sansone u. a. (2019, S. 360)). Dabei ist zu beachten, dass FAIRsharing aus einer gezielt molekularbiologisch ausgerichteten Datenbank für Standards, BioSharing, entstand (Sansone u. a. (2019, S. 366)), wodurch diese starke Präsenz von molekularbiologischen Metadatenstandards beeinflusst sein kann. Die Existenz und der Bedarf nach BioSharing selbst weist allerdings auch bereits auf eine hohe Anzahl und intensive Nutzung von molekularbiologischen Standards hin.

¹⁵⁸ 18 Metadaten schemata unter Biologie (im Metadata Standards Catalog - Biology) und 10 Metadaten schemata unter Lebenswissenschaften (im Metadata Standards Catalog - Natural sciences).

¹⁵⁹ FAIRsharing - Standards; Biology.

¹⁶⁰ FAIRsharing - GenBank Sequence Format; letzte Veränderung: 05. 02. 2020, 12:45 p.m..

¹⁶¹ FAIRsharing - SBML; Systems Biology Markup Language; letzte Veränderung: 10. 04. 2019, 10:49 a.m..

¹⁶² FAIRsharing - DwC; Darwin Core; letzte Veränderung: 08. 01. 2019, 1:38 p.m..

¹⁶³ Benson u. a. (2011, S. D33).

¹⁶⁴ Michael Hucka u. a. (2019).

(z. B. Locus, Definition, Identifikator (bzw. Accession), Version, Stichwörter, Herkunft, Referenz, Eigenschaften, Ursprung) mit verschiedenen Unterfeldern¹⁶⁵ (Anhang, Tab. 4). Mit „Locus“ ist der Genlocus gemeint, repräsentiert durch einen einzigartigen Genlocus-Namen. Zum Genlocus gehört auch eine Genbank-Bereichsangabe, die Bereiche der Genbank-Datenbank widerspiegelt, und den Eintrag anhand 18 vorgegebener möglicher Abkürzungen charakterisiert. Damit wird der Genbank-Datenbank die maschinelle Gruppierung von Einträgen erlaubt. Als Referenz sind die allgemeinen Daten Autor, Titel, Zeitschrift, in der Daten veröffentlicht werden, Pubmed-Identifikator und Kontaktdaten des Einreichenden gefragt.¹⁶⁶

SBML dagegen ist eine Auszeichnungssprache für die Annotation von Computermodellen für biologische Prozesse, z. B. biochemische Reaktionen.¹⁶⁷ Sie enthält keine Vorgaben für allgemeine Angaben, sondern nur fachspezifische Elemente wie mathematische Funktionen, Messeinheiten und weitere¹⁶⁸ (Anhang, Tab. 5). *SBML* wird von verschiedenen Datenbanken implementiert und ist mit mehreren anderen Standards assoziiert.¹⁶⁹ Daher eignet sie sich gut als Austauschformat innerhalb der Disziplin.

Darwin Core ist eines der bekanntesten Standards des Biodiversitäts-Fachbereichs¹⁷⁰ und leitet sich unter anderem von Dublin Core ab.¹⁷¹ Dadurch sind allgemeine Angaben (Anhang, Abb. 2, Tab. 6) zu beispielsweise Ressourcen-Typ und Sprache¹⁷² sowie Veränderungsdatum, Zugriffsberechtigung usw.¹⁷³ über den Fachbereich hinaus

¹⁶⁵ National Center for Biotechnology Information - Sample GenBank record. GenBank flat file format.

¹⁶⁶ National Center for Biotechnology Information - Sample GenBank record. GenBank flat file format.

¹⁶⁷ Michael Hucka u. a. (2019, S. 9), M. Hucka u. a. (2003).

¹⁶⁸ Michael Hucka u. a. (2019, S. 9).

¹⁶⁹ FAIRsharing - SBML; Systems Biology Markup Language; letzte Veränderung: 10. 04. 2019, 10:49 a.m..

¹⁷⁰ Global Biodiversity Information Facility - Data Standards, Castro u. a. (2017, S. 185).

¹⁷¹ Wieczorek u. a. (2012, S. 2).

¹⁷² Die Darwin Core Elemente (Biodiversity Information Standards - Darwin Core quick reference guide) lassen sich bei den Dublin Core Elementen wiederfinden (Dublin Core Metadata Initiative - DCMI usage board: DCMI metadata terms. Elements).

¹⁷³ Die Darwin Core Elemente (Biodiversity Information Standards - Darwin Core quick reference guide) lassen sich bei den Dublin Core Begriffen wiederfinden (Dublin Core Metadata Initiative - DCMI usage board: DCMI metadata terms. Terms).

standardisiert. Begriffe aus weiteren Standards, z. B. des Access to Biological Collections Data (ABCD) Schemas,¹⁷⁴ wurden ebenfalls integriert.¹⁷⁵ Auch wurde die fachspezifische Beschreibung von Daten ermöglicht (z. B. geologischer Kontext¹⁷⁶).

Darüber hinaus gibt es neben den oben erwähnten Ontologien weitere übergreifende Projekte der Metadaten-Standardisierung. Ein prominentes ist das 2008 eingeführte Projekt *MIBBI*¹⁷⁷, dessen Portal für biologische Standards zu FAIRsharing geführt hat (Kapitel 2). Neben MIBBI hat sich außerdem *Bioschemas* herausgebildet. Bioschemas unterstützt die Entwicklung von Auszeichnungsprofilen für verschiedene Daten-Typen in den Lebenswissenschaften, die auf dem Schema.org-Standard (Kapitel 2.1) basieren.¹⁷⁸ Die Diversität der Standards in der Biologie spiegelt zum einen die Vielfalt der biologischen Forschungsdaten wider, zum anderen das Bedürfnis der Fachcommunity, passgenaue Standards in den einzelnen Fachbereichen zur Verfügung zu haben. Verstärkt wurde dies durch hochspezialisierte fachspezifische Repositorien, die entsprechend spezifische Metadatenstandards und Kuratierung anbieten.¹⁷⁹ Die große Anzahl an etablierten Standards und Organisationen, die diese Standards bearbeiten, weist darauf hin, dass Bibliotheken keine führende Rolle in der Standard-Etablierung für einen Großteil dieser Disziplin übernehmen können oder sollten. Auf der anderen Seite zeigt diese Fragmentierung, dass die Herstellung von Beziehungen zwischen Standards, bzw. Mappings, die Integration von generischen Standards in fachspezifische oder ähnliche Bestrebungen in Richtung Interoperabilität zunehmend nötig werden. Hier können Bibliotheken mit Wissenschaftlern und Daten-Kuratoren innerhalb der bestehenden Strukturen, wie der RDA, zusammenarbeiten. Um bestimmte wichtige, grundlegende Standards, wie die Zitierbarkeit über persistente Identifikatoren¹⁸⁰ oder die Provenienz von Metadatenschemata selbst,¹⁸¹ zu fördern, können sich

¹⁷⁴ FAIRsharing - ABCD; Access to Biological Collection Data; letzte Veränderung: 29. 10. 2019, 1:27 p.m., Holetschek, Dröge, Güntsch & Berendsohn (2012)

¹⁷⁵ Wiczorek u. a. (2012, S. 2).

¹⁷⁶ Wiczorek u. a. (2012, S. 3).

¹⁷⁷ Taylor u. a. (2008) MIBBI steht für Minimum Information for Biological and Biomedical Investigations.

¹⁷⁸ Bioschemas - What Is Bioschemas?, Gray, Goble & Jimenez (2017).

¹⁷⁹ Austin u. a. (2017, S. 88).

¹⁸⁰ Wilkinson u. a. (2016, S. 4).

¹⁸¹ Sugimoto, Li, Nagamori & Greenberg (2016, S. 47–48), Sansone & Rocca-Serra (2016, S. 11–15) Metadatenstandards, inklusive der Metadatenschemata, sind ebenso wie FD dynamisch und unterliegen einem Lebenszyklus (Formulierung, Entwicklung und Pflege) (Sansone u. a. (2019, S. 358)),

Bibliothekare in die Weiterentwicklung von allgemeinen Schemata wie Dublin Core und DataCite einbringen. Falls die Entwicklung eines neuen Metadatenschemas nötig ist, sollte auf bestehende Schemata aufgebaut werden. Bibliotheken können in diesem Fall fachübergreifende Grundlagen beisteuern und Hilfestellungen geben.¹⁸² Eine stärker gestaltende Rolle bei der Standard-Entwicklung können Bibliotheken gegebenenfalls auch disziplinspezifisch einnehmen, sofern in einer Disziplin wenig Datenaustausch und etablierte Standards existieren. Ein Beispiel für einen solchen Fachbereich sind die Ingenieurwissenschaften.

3.2 Ingenieurwissenschaften

In den Ingenieurwissenschaften (IW) ist, im Gegensatz zur Biologie, Datenaustausch weitaus weniger etabliert.¹⁸³ Hinzu kommt eine starke Auffächerung der Disziplin, sichtbar auch in der Evaluation der ingenieurwissenschaftlichen Forschung 2001 in Niedersachsen.¹⁸⁴ In dieser wurden die IW grundsätzlich unterteilt in „Bauingenieurwesen und Architektur“, „Elektrotechnik und Informationstechnik“, sowie „Maschinenbau“.¹⁸⁵ Zu diesen Bereichen kommen gemäß der DFG Fachsystematik¹⁸⁶ „Produktionstechnik“, „Verfahrenstechnik, Technische Chemie“, der Bereich thermische IW und „Strömungsmechanik“, die Bereiche „Werkstofftechnik“, „Materialwissenschaft“, „Systemtechnik“ und „Informatik“ hinzu.¹⁸⁷ Unter diese Forschungsfelder

sodass sie als eigenständige digitale Objekte ebenfalls den FAIR-Prinzipien entsprechen sollten (Sansone u. a. (2019, S. 360)), inklusive Versionierung und Zitierbarkeit.

¹⁸² Iglezakis & Schembera (2018, S. 58–59).

¹⁸³ Alexander Ball u. a. (2012), Iglezakis & Schembera (2018, S. 47).

¹⁸⁴ Wissenschaftliche Kommission Niedersachsen (2001b).

¹⁸⁵ Wissenschaftliche Kommission Niedersachsen (2001b, S. 3–4).

¹⁸⁶ Deutsche Forschungsgemeinschaft (DFG) - Fachsystematik.

¹⁸⁷ Deutsche Forschungsgemeinschaft (DFG) - Fachsystematik Die DFG unterteilt Ingenieurwissenschaften in „Produktionstechnik“, „Mechanik und Konstruktiver Maschinenbau“, „Verfahrenstechnik, Technische Chemie“, „Strömungsmechanik, Technische Thermodynamik und Thermische Energietechnik“, „Werkstofftechnik“, „Materialwissenschaft“, „Systemtechnik“, „Elektrotechnik und Informationstechnik“, „Informatik“ und „Bauwesen und Architektur“. Von diesen Bereichen wurde „Bauwesen und Architektur“, „Elektrotechnik und Informationstechnik“ und „Mechanik und Konstruktiver Maschinenbau“ mit der Einteilung der Wissenschaftlichen Kommission Niedersachsen (Wissenschaftliche Kommission Niedersachsen (2001b, S. 3–4)) gleichgesetzt, der Rest wurde hinzugefügt aufgezählt mit Ausnahme von „Strömungsmechanik, Technische Thermodynamik und Thermische Energietechnik“. Dieser fachliche Bereich wurde für die bessere Lesbarkeit als der Bereich thermische IW und „Strömungsmechanik“ zusammengefasst.

kann z. B.: Gebäudetechnik¹⁸⁸ und Informationstechnologie¹⁸⁹, sowie Überschneidungen mit Mathematik, Medizin, Naturwissenschaften, Betriebswirtschaft, Geisteswissenschaften und Soziologie¹⁹⁰ fallen. In den IW ist Forschung und Lehre äußerst anwendungsorientiert¹⁹¹ und oft werden Kooperationen mit der Wirtschaft eingegangen.¹⁹² Als Folge vermuten Forscher selbst wenig Interesse in der Fachcommunity an Ergebnissen ihrer eigenen spezifischen Projekte¹⁹³ und selbst bei vorhandenem Interesse können Ergebnisse gegebenenfalls auf Grund von Vereinbarungen mit der Industrie nicht veröffentlicht werden.¹⁹⁴ Hinzu kommen große Datenvolumina und das schnelle Veralten von Ergebnissen, wodurch Austausch erschwert wird.¹⁹⁵ Aber auch die starke Interdisziplinarität von IW¹⁹⁶ und die damit einhergehende Diversität von FD¹⁹⁷ verkompliziert das Erstellen von Standards, die die Disziplin weitreichend abdecken. Beispielsweise beziehen sich die Forschungsdaten in den IW oft auf Software und Code¹⁹⁸, können aber auch als Videoausschnitte, XSLT-Dateien, Interview-Mitschriften oder Daten in Excel-Dateien¹⁹⁹ vorliegen. Speziell Software und Code war in den Anfängen des FDM nicht explizit als Forschungsdaten präsent, sodass in diesem Bereich weniger Standards für den Datenaustausch vorhanden sind als für etablierte FD-Typen.²⁰⁰

¹⁸⁸ Wissenschaftliche Kommission Niedersachsen (2001b, S. 12).

¹⁸⁹ Wissenschaftliche Kommission Niedersachsen (2001b, S. 42).

¹⁹⁰ Wissenschaftliche Kommission Niedersachsen (2001b, S. 64).

¹⁹¹ Wissenschaftliche Kommission Niedersachsen (2001b, S. 11, 41, 61).

¹⁹² Wissenschaftliche Kommission Niedersachsen (2001b, S. 12, 41, 61).

¹⁹³ Howard, Darlington, Ball, Culley & McMahon (2010b, S. 5), Iglezakis & Schembera (2018, S. 50).

¹⁹⁴ Howard u. a. (2010b, S. 9–10), Iglezakis & Schembera (2018, S. 52).

¹⁹⁵ Iglezakis & Schembera (2018, S. 51–52).

¹⁹⁶ Die Interdisziplinarität zeigt sich auch bei der Darstellung der Disziplin durch FDM-Web-Angebote: Im MSC werden die IW nicht als eigene Disziplin aufgeführt, sondern z. B. „Politics, Law and economics“ (darin: „civil, military and mining engineering“, „manufacturing and transport engineering“) oder den Wissenschaften (darin: „Environmental sciences and engineering“) zugeordnet (Metadata Standards Catalog - Index of subjects). In FAIRsharing resultiert die Einschränkung auf Standards der „Engineering Science“ in Standards, die auch „Computer Science“, „Informatics“ und „Natural Science“ zugeordnet werden (FAIRsharing - Standards; Engineering Science). Für re3data zeigen die IW in Abbildung 4 des Artikels von Kindling et al. vergleichsweise wenige Datenrepositorien, die nur den IW zugeordnet wurden (Kindling u. a. (2017)).

¹⁹⁷ Alexander Ball u. a. (2012), Howard u. a. (2010a, S. 19–21), Scott u. a. (2014, S. 36–37).

¹⁹⁸ Iglezakis & Schembera (2018, S. 48–51).

¹⁹⁹ Howard u. a. (2010a, S. 20, 21), Scott u. a. (2014, S. 36).

²⁰⁰ Li, Lin & Greenberg (2016, S. 1–3).

Nur wenige Metadaten schemata sind, dem MSC zufolge, primär ingenieurwissenschaftlich (drei Schemata im Ingenieurwesen (engineering)²⁰¹), während FAIRsharing mit 478 Standards für Ingenieurwissenschaften eine große Menge der interdisziplinären Standards zu diesem Fachbereich zählt.²⁰² Da FAIRsharing aus dem biologisch ausgerichteten BioSharing entstanden ist²⁰³, ist zu erwarten, dass die verzeichneten Daten vornehmlich biologisch ausgerichtet sind. Der MSC dagegen wurde unter anderem von Alexander Ball aufgebaut²⁰⁴, der auch im FDM im Bereich IW involviert war.²⁰⁵ Daher wurden für diesen Artikel repräsentative IW-Standards entsprechend der im MSC hinterlegten Standards im Ingenieurwesen („engineering“)²⁰⁶ ausgewählt: *Core Scientific Metadata model (CSMD)*, *Crystallographic Information Framework (CIF)* und *NeXus*²⁰⁷. Zusätzlich wurde das Metadaten schema *EngMeta*²⁰⁸ als Beispiel für ein ingenieurwissenschaftliches Schema, das Elemente anderer Schemata inkorporiert, betrachtet. Es wurde erstellt, um die vorhandenen, aber für IW, speziell die Simulationswissenschaften, nicht ausreichenden Metadaten schemata zu ergänzen.²⁰⁹

Das *CSMD* ist ein vergleichsweise allgemeines Metadatenmodell für die Beschreibung von wissenschaftlichen Aktivitäten, sodass die Vorgaben auf ein Minimum beschränkt wurden²¹⁰. Zu den hauptsächlichen Elementen (Anhang, Abb. 3) gehören allgemeine Angaben, wie Studien, Untersuchung, Probe, Apparat, untersuchende Person und weitere²¹¹; hinzu kommen Elemente, die genauere Angaben zu z. B. dem Namen des Nutzers (user_name) erlauben.²¹² Daran zeigt sich, dass wenige fachspezifische Angaben erwartet werden, was die allgemeine, fachübergreifende Nutzbarkeit des Schemas unterstreicht.

²⁰¹ Metadata Standards Catalog - Engineering.

²⁰² FAIRsharing - Standards; Engineering Science.

²⁰³ Sansone u. a. (2019, S. 366).

²⁰⁴ Alex Ball u. a. (2016).

²⁰⁵ Alex Ball & Neilson (2010), Alexander Ball u. a. (2012), Howard u. a. (2010b), (2010a).

²⁰⁶ Metadata Standards Catalog - Engineering.

²⁰⁷ Metadata Standards Catalog - Engineering.

²⁰⁸ Schembera & Iglezakis (2019).

²⁰⁹ Iglezakis & Schembera (2018, S. 47).

²¹⁰ Matthews & Fisher (2013, S. 1).

²¹¹ Matthews & Fisher (2013, S. 2).

²¹² Matthews & Fisher (2013, S. 3).

Das Metadata-Standard *CIF* dagegen ist äußerst fachspezifisch, indem sich die Elemente (Anhang, Tab. 7) auf kristallografische Beschreibungen (z. B. atom, cell, chemical²¹³) oder auf die Datenpublikation (journal und publ²¹⁴) beschränken, während zum Autor keine Angaben erwartet werden. Ähnliches gilt für *NeXus*, ein Datenformat für den Austausch und die Archivierung von Neutronen-, Röntgen- und Myonen-Daten.²¹⁵ Das Format besteht aus Basisklassen²¹⁶ (Anhang, Tab. 8) und Applikationen²¹⁷, die bis auf *NXcite* und *NXuser*²¹⁸ fachspezifisch sind. Mit den zwei allgemeinen Elementen können Literaturreferenzen²¹⁹ und Kontaktdaten von Personen²²⁰ hinterlegt werden.

Ein neu entwickeltes Metadatenschema für die Simulationswissenschaften, *EngMeta*²²¹, repräsentiert die Möglichkeit, Metadatenschemata zu kombinieren, um den Datenaustausch zwischen verschiedenen Datenbanken sowie Repositorien und die Metadaten-Interoperabilität zu unterstützen.²²² Konzipiert ist es aus Komponenten von *DataCite* (Zitation und generelle Metadaten), *PREMIS* (technische Angaben wie Datei-Größe) und *CodeMeta* (Software-Angaben) sowie weiteren Schemata, die Angaben zum wissenschaftlichem Arbeitsablauf und Versuchshardware standardisieren²²³ (Anhang, Abb. 4). Das Schema ist inzwischen in ein institutionelles Daten-Repository integriert.²²⁴ Eine weitere Möglichkeit für mehr Interoperabilität zwischen den verschiedenen Standards ist die Entwicklung eines flexiblen Metadatenmodells, das gemäß des Entity-Attribute-Value-Modells für jedes Metadatenelement einen eigenen Datensatz erstellt.²²⁵ Damit lassen sich Elemente flexibel hinterlegen und nach Bedarf in Mustervorlagen zusammenfassen.²²⁶ Nichtsdestotrotz sollten diese Felder Elementen etablierter Metadatenschemata entsprechen, um Interoperabilität zu sichern.

²¹³ International Union of Crystallography - CIF; Data name categories.

²¹⁴ International Union of Crystallography - CIF; Data name categories und International Union of Crystallography - CIF; Appendix; CIF Dictionary (Core Version 1991).

²¹⁵ Könnecke u. a. (2015).

²¹⁶ *NeXus* - 3.3.1. Base class definitions.

²¹⁷ *NeXus* - 3.3.2. Application definitions.

²¹⁸ *NeXus* - 3.3.1. Base class definitions.

²¹⁹ *NeXus* - 3.3.1.7. *NXcite*.

²²⁰ *NeXus* - 3.3.1.55. *NXuser*.

²²¹ Schembera & Iglezakis (2019).

²²² Schembera & Iglezakis (2019, S. 130).

²²³ Schembera & Iglezakis (2019, S. 130).

²²⁴ Hermann, Iglezakis & Seeland (2019, S. 3).

²²⁵ Scott u. a. (2014, S. 39).

²²⁶ Scott u. a. (2014, S. 43).

Standardisierungsbestrebungen in den Ingenieurwissenschaften scheinen international nur von der RDA-Interessensgruppe Research Data Management in Engineering²²⁷ auszugehen, die sich aus Wissenschaftlern und Bibliothekaren zusammensetzt,²²⁸ und von dem weltweiten Berufsverband Institute of Electrical and Electronics Engineers (IEEE)²²⁹. In Deutschland strebt die nationale Forschungsinfrastruktur NFDI4ing ebenfalls Standardisierung von Metadaten schemata in den IW an.²³⁰ Im Gegensatz zur Biologie führt der Mangel an etablierten Standards in einigen Bereichen der IW dazu, dass Bibliotheken dort eine führende Rolle bei der Etablierung neuer, interoperabler und den FAIR-Prinzipien entsprechender Metadatenstandards einnehmen können.²³¹ Dabei können Bibliotheken sich für die Integration wichtiger, fachunabhängiger Elemente, wie persistente Identifikatoren, Rechte und Lizenzen, zugehöriges Projekt, Aspekte der Langzeitarchivierung und weitere,²³² einsetzen. Eine Zusammenarbeit mit Wissenschaftlern, um die tatsächlichen Ansprüche der Nutzer zu eruieren, wie beispielsweise bei der Erstellung des neuen Metadaten schemas EngMeta²³³, ist trotzdem nötig. Bei interdisziplinärer ingenieurwissenschaftlicher Forschung oder der Nutzung von Methoden aus anderen Disziplinen sollte überprüft werden, ob fachfremde Metadaten schemata vorhanden sind, die sich für diese Forschungsdaten eignen. Bibliotheken könnten in diesem Fall Unterstützung und Beratung anbieten.²³⁴ Vereinzelt werden in IW, in Abwesenheit von etablierten Standards und Infrastrukturen, innovative Lösungen für das ingenieurwissenschaftliche FDM eingeführt, wie das flexible Metadatenmodell von Scott und seinen Mitarbeitern.²³⁵ Angesichts der steigen-

²²⁷ Research Data Alliance - Interessensgruppe Research Data Management in Engineering, Hermann u. a. (2019, S. 4).

²²⁸ Research Data Alliance - Interessensgruppe Research Data Management in Engineering; Members.

²²⁹ Rat für Informationsinfrastrukturen (2019, S. A-10).

²³⁰ Hermann u. a. (2019, S. 4).

²³¹ Iglezakis & Schembera (2018, S. 58–59).

²³² Iglezakis & Schembera (2018, S. 54–55), Schembera & Iglezakis (2019, S. 130), Lindlar u. a. (2020) Langzeitarchivierungsaspekte werden besonders relevant beim Transfer von FD aus FD-Repositories an Archive. Eine entsprechende Übergabe von archivwürdigen FD wird z. B. nach dem DINI-Zertifikat für Repositorien behandelt (Lindlar u. a. (2020, S. 2), Müller u. a. (2017, S. 29–30)). Wird Langzeitarchivierung nicht früh mitbedacht, entsteht das Risiko, dass FD nur mit großem Aufwand oder, bei entsprechend strikten Vorgaben eines Archivs, gar nicht archiviert werden können (Lindlar u. a. (2020, S. 2)).

²³³ Schembera & Iglezakis (2019, S. 128), Iglezakis & Schembera (2018).

²³⁴ Iglezakis & Schembera (2018, S. 58–59).

²³⁵ Scott u. a. (2014).

den Anzahl von Metadatschemata ist ein flexibles Modell, allerdings mit standardisierten, zentral hinterlegten Metadatenelementen, die als Ontologie oder als mehrere Ontologien im RDF-Format modelliert sind, eventuell eine Möglichkeit um Flexibilität und Interoperabilität zu gewähren. Aus den vorhandenen Elementen ließen sich Anwendungsprofile für einzelne Disziplinen erstellen. Ein ähnliches Konzept verfolgt die FDM-Plattform Dendro, in dessen System sich allerdings Anwendungsprofile für einzelne Nutzer statt für Disziplinen erstellt lassen.²³⁶ Das grundlegende Konzept ließe sich trotzdem möglicherweise auf Metadaten eines Fachbereichs oder sogar disziplinübergreifend übertragen.

Generell sind auch die Anzahl der Metadatenfelder insgesamt und der Anteil an obligatorischen Feldern kritisch zu betrachten. Ist der Aufwand für Wissenschaftler zu groß, Datensätze mit entsprechenden Metadaten zu veröffentlichen, stellt dies eine Hürde dar.²³⁷ Verschiedene Anwendungen und Software können für Wissenschaftler sowohl die Dokumentation von Metadaten im Forschungsprozess als auch die Veröffentlichung von Daten erleichtern.

4. Anwendungen für die Metadatendokumentation

Die Dokumentation von Metadaten im Zusammenhang mit Forschungsergebnissen lässt sich auch als Forschungsdokumentation beschreiben und wird von Wissenschaftlern während des Forschungsalltags durchgeführt.²³⁸ Im Optimalfall würde sie generell zu hochqualitativen Metadaten führen, die eine Veröffentlichung von FD ohne Mehraufwand ermöglichen.²³⁹ Die Art der Dokumentation ist stark von der Art der Forschung in der jeweiligen Disziplin abhängig. Beispielsweise wird die Dokumentation der experimentellen Forschung in den Naturwissenschaften und der Medizin während des Forschungsprozesses (z. B. die Auswahl und Behandlung der Versuchsobjekte²⁴⁰) in Laborbüchern festgehalten.²⁴¹ Diese zum Teil immer noch analog geführten Bücher enthalten dementsprechend die für eine Datenpublikation notwendigen Metadaten.

²³⁶ Castro u. a. (2017, S. 189–190).

²³⁷ Rat für Informationsinfrastrukturen (2019, S. 38), Iglezakakis & Schembera (2018, S. 59).

²³⁸ Krause (2016, S. 78–79), Alexander Ball u. a. (2012).

²³⁹ Iglezakakis & Schembera (2018, S. 59), Hermann u. a. (2018, S. 34).

²⁴⁰ Krause (2016, S. 79).

²⁴¹ Krause (2016, S. 78).

Mit der Einführung von softwarebasierten elektronischen Laborbüchern, ELNs (electronic laboratory notebooks), liegen Metadaten nun elektronisch vor und die Verknüpfung einzelner ELN-Einträge und der digitalen Forschungsdaten ist möglich.²⁴² Eine große Anzahl an kommerziellen ELNs wird bereits angeboten²⁴³, allerdings liegt der Fokus eher auf der Vereinfachung des Forscheralltags als der strukturierten Aufbereitung von Metadaten.²⁴⁴ Eine Ausnahme ist das ELN-Projekt LabBook²⁴⁵ der IBM, in dem Metadatenelemente in einem flexiblen Graph-Schema mit Beziehungen untereinander, ähnlich dem Semantic Web, hinterlegt sind.²⁴⁶ Diese Elemente basieren allerdings nicht notwendigerweise²⁴⁷ auf bestehenden Standards oder integrierten Ontologien²⁴⁸, sondern können auch von Nutzern oder durch automatische Extraktion aus Datensätzen dem Graphen hinzugefügt werden.²⁴⁹ Einen ähnlichen Ansatz verfolgen Kanza und ihre Mitautoren mit dem Vorschlag, Technologien des Semantic Webs wie RDF und Ontologien, allerdings in Übereinstimmung mit Metadatenstandards, im ELN breit zu implementieren.²⁵⁰ Sie sprechen sich auch für eine Etablierung der automatischen Daten- und Metadaten-Übertragung von Laborgeräten in ELNs aus.²⁵¹ Die Anbindung von ELNs an Repositorien über Schnittstellen würde zusätzlich die FD-Veröffentlichung aus dem ELN heraus einfach gestalten.²⁵² Verschiedene ELN-Open-Source-Projekte (z.B. Chemotion²⁵³ und openBIS²⁵⁴) ermöglichen eine Beteiligung von Bibliotheken an der Entwicklung dieser Anwendungen, im Zuge derer sie Expertisen bezüglich Metadatenstandards, besonders im disziplinübergreifenden Bereich und Semantic-Web-Technologien einbringen könnten.

²⁴² Krause (2016, S. 79–80).

²⁴³ Kanza u. a. (2017, S. 2–3), Harvard University - Harvard Biomedical Data Management; Electronic lab notebooks.

²⁴⁴ Kanza u. a. (2017, S. 10).

²⁴⁵ Kandogan u. a. (2015).

²⁴⁶ Kandogan u. a. (2015, S. 433–434).

²⁴⁷ Kandogan u. a. (2015, S. 432).

²⁴⁸ Kandogan u. a. (2015, S. 437).

²⁴⁹ Kandogan u. a. (2015, S. 434).

²⁵⁰ Kanza u. a. (2017, S. 2, 9–10, 11). Entsprechend plädieren auch Nussbeck und ihre Mitautoren für die Implementierung von Metadatschemata in ELNs (Nussbeck u. a. (2014, S. 633)), allerdings ohne Bezug auf Technologien des Semantic Webs.

²⁵¹ Kanza u. a. (2017, S. 10).

²⁵² Hermann u. a. (2018, S. 38) Hermann und ihre Mitautoren erwähnen die Anbindung des ELNs Chemotion (Chemotion - <https://chemotion.net/>) an ein Repository (Hermann u. a. (2018, S. 38)).

²⁵³ Tremouilhac u. a. (2017).

²⁵⁴ Bauch u. a. (2011).

Neben ELNs existieren weitere Dokumentationsmethoden, wie die Virtuelle Forschungsumgebung²⁵⁵, das Workflow-Management²⁵⁶ und das Workflow-Tracking.²⁵⁷ Da die Forschung der IW, besonders die Simulationswissenschaften, intensiv mit Software-Code und selbstgeschriebener Software arbeitet²⁵⁸, wird in diesem Artikel das Workflow-Tracking näher betrachtet, das aus der Software-Entwicklung stammt.²⁵⁹ Dabei wird innerhalb dieser Methode der Fokus auf Git²⁶⁰ gelegt, dem populärsten System zur Versionskontrolle (version control system, VCS).²⁶¹ Git ermöglicht, mit jeder Veränderung einer Computermodellierung, Softwareentwicklung oder informatischen Datenauswertung, die zu einer neuen Version von Dateien und Code führt, eine Momentaufnahme der Dateien zu speichern und zu dokumentieren.²⁶² Ein Beispiel für FDM mit Git-Instanzen ist die Anwendung MetaGit²⁶³ oder der Client des Projekts RePlay-DH, der für die Digital Humanities entwickelt wurde²⁶⁴ und mitsamt Metadatenchema an IW angepasst werden könnte.²⁶⁵ Der RePlay-Client ermöglicht sowohl die Metadaten-Dokumentation nach Schemata, die auf dem Dublin Core-Metadatenchema basieren²⁶⁶, als auch eine zukünftige Anbindung an ein Repositorium²⁶⁷, sodass Forschungsergebnisse mit Metadaten aus dem Client heraus veröffentlicht werden können. Das Projekt wurde als Kooperation zwischen Wissenschaftlern und Bibliothekaren konzipiert.²⁶⁸

²⁵⁵ Hermann u. a. (2018, S. 35–37).

²⁵⁶ Hermann u. a. (2018, S. 38–39).

²⁵⁷ Hermann u. a. (2018, S. 39–41).

²⁵⁸ Hermann u. a. (2019, S. 1).

²⁵⁹ Hermann u. a. (2018, S. 39).

²⁶⁰ Blischak, Davenport & Wilson (2016).

²⁶¹ Hermann u. a. (2018, S. 40).

²⁶² Blischak u. a. (2016, S. 3–4).

²⁶³ Scott, Johnston & Cox (2017, S. 89–91).

²⁶⁴ Hermann u. a. (2018, S. 41–42), Gärtner, Hahn & Hermann (2018).

²⁶⁵ Hermann u. a. (2018, S. 42).

²⁶⁶ Gärtner u. a. (2018, S. 289).

²⁶⁷ Gärtner u. a. (2018, S. 289–290), Hermann u. a. (2018, S. 42).

²⁶⁸ Die Autoren des Artikels über das Projekt gehören dem Institut für maschinelle Sprachverarbeitung (Institute for Natural Language Processing) der Universität Stuttgart, dem Kommunikations- und Informationszentrum (Communication and Information Centre) der Universität Ulm und der Bibliothek (Library) der Universität Stuttgart an (Gärtner u. a. (2018, S. 284)). Eine Projektbeschreibung ist auch auf der Seite der Bibliothek der Universität Stuttgart vorhanden (Bibliothek der Universität Stuttgart - RePlay-DH; Projektbeschreibung).

5. Fazit

Um vor allem die Auffindbarkeit und Nachnutzung der Daten entsprechend der FAIR-Prinzipien zu ermöglichen, müssen die Daten ausreichend mit Metadaten annotiert werden, die wiederum über Metadatenschemata bis zu einem gewissen Maße standardisiert werden. Standardisierungsbestrebungen inklusive Metadatenschemata können sich für verschiedene Disziplinen stark unterscheiden. Die molekular ausgerichtete Biologie pflegt eine lange Tradition des Datenaustausches, besitzt eine große Anzahl an etablierten Metadatenstandards, Datenbanken, Repositorien sowie Organisationen, die Standards entwickeln und Daten kuratieren. Generell geht Datenaustausch oft auf eine Selbstverpflichtung der Disziplin, wie bei der molekularen Biologie, oder auf eine gesetzliche Anforderung zurück.²⁶⁹ In den Ingenieurwissenschaften besteht dagegen eine geringer ausgeprägte Kultur des Datenaustausches, da in dieser Disziplin stärker anwendungsbezogene und industrienahere Forschung mit großen Datenmengen betrieben wird, wodurch Hürden für den Datenaustausch primär ingenieurwissenschaftlicher Daten entstehen. Durch die Interdisziplinarität der IW existieren in dieser Disziplin einige Metadatenschemata, die sich eher anderen Fachbereichen zuordnen lassen, während primär ingenieurwissenschaftliche Schemata rar sind.

Die fachspezifischen Standards beider Disziplinen zeigen in manchen Fällen Interoperabilität über Disziplingrenzen hinweg, aber oft besitzen die Standards nur wenige allgemeine disziplinenübergreifende Metadaten. Das lässt sich bei vielen Schemata, die schon seit längerer Zeit etabliert sind, wie das *GenBank Sequence Format*, *SBML*, *CIF* und *NeXus* (Kapitel 3) beobachten. In diesen Fällen kann man vermuten, dass die Standards spezifisch auf die Bedürfnisse der Fachcommunity ausgerichtet sind und Datenaustausch über Disziplingrenzen hinweg bei ihrer Erstellung nicht in Betracht gezogen wurde. Die Signifikanz des fachübergreifenden Datenaustausches nimmt allerdings zurzeit im Zusammenhang mit den FAIR-Prinzipien zu.²⁷⁰ Von Vorteil wäre eine Erweiterung der bestehenden Standards mit etablierten fachunabhängigen Standards, wie *Dublin Core* und vor allem dem FD-spezifischen *DataCite*. Eine solche Kombination mit bestehenden generischen Standards ist beispielsweise bei den Schemata *Darwin Core* und *EngMeta* ersichtlich. Wichtige allgemeine Metadaten können

²⁶⁹ Austin u. a. (2017, S. 78).

²⁷⁰ Wilkinson u. a. (2016, S. 4).

aber über die Möglichkeiten von *DataCite* hinausgehen, wie Elemente, die das zugehörige Forschungsprojekt oder Aspekte der Langzeitarchivierung beschreiben.²⁷¹ Impulse zur Einbeziehung solcher generischen Standards können von Bibliotheken in gegenwärtige Standardisierungsbestrebungen, beispielsweise die Entwicklung und Bearbeitung von Standards eingebracht werden. Darüber hinaus könnte von ihnen eine Entwicklung in Richtung standardisierter Metadatenelementen in Ontologien mit flexibler Zusammenstellung zu Schemata oder Anwendungsprofilen verfolgt werden, um sowohl Fachspezifität als auch Interoperabilität zu ermöglichen. Weitere Unterstützung können Bibliotheken bei der Entwicklung von Anwendungen bieten, die eine standardisierte Eingabe von Forschungsdaten während des Forschungsalltags erleichtern, sodass ein vermehrter Arbeitsaufwand für Wissenschaftler bei der letztlichen Veröffentlichung der Daten vermieden wird. Um Projekte und Entwicklungen auf Nutzerbedürfnisse aufzubauen, sollten existierende wissenschaftliche Institutionen und Austauschplattformen wie RDA einbezogen werden. Generell ergeben sich viele Gelegenheiten der Forschungsunterstützung durch Bibliotheken im Zusammenhang mit der starken Proliferation von Metadatschemata und -standards sowie bei der Herausforderung für Forscher, qualitativ hochwertige Forschungsdokumentation zu generieren.

²⁷¹ Iglezakis & Schembera (2018, S. 55), Schembera & Iglezakis (2019, S. 130), Lindlar u. a. (2020).

Anhang

Tab. 1: Übersicht über die in diesem Bericht erwähnten Metadatenschemata und -standards

Disziplin	Metadatenschema oder -standard	Behandeln- des Kapitel	Details	Referenz
Generisch				
	Dublin Core	2.1	Anhang, Tab. 2	Caverlee u. a. (2009, S. 947–949)
	Schema.org	2.1	-	Guha, Guha u. a. (2016)
	PREMIS	2.1	Anhang, Abb. 1	PREMIS Editorial Committee (2015)
	DataCite	2.1	Anhang, Tab. 3	DataCite Metadata Working Group (2019)
Biologie				
	GenBank Sequence Format	3.1	Anhang, Tab. 4	National Center for Biotechnology Information - Sample GenBank record. GenBank flat file format
	Systems Biology Markup Language (SMBL)	3.1	Anhang, Tab. 5	Michael Hucka u. a. (2019)
	Darwin Core	3.1	Anhang, Tab. 6, Abb. 2	Wieczorek u. a. (2012)
	Bioschemas	3.1	-	Gray u. a. (2017)
	Gene Ontology	2, 3.1	-	Ashburner u. a. (2000)
	Access to Biological Collections Data (ABCD) Schema	3.1	-	Holetschek, Dröge, Güntsch & Berendsohn (2012)
	Minimum Information for Biological and Biomedical Investigations (MIBBI)	2, 3.1	-	Taylor u. a. (2008)

Disziplin	Metadatenchema oder -standard	Behandeln- des Kapitel	Details	Referenz
Ingenieurwissenschaften				
	Core Scientific Metadata model (CSMD)	3.2	Anhang, Abb. 3	Matthews & Fisher (2013)
	Crystallographic Information Framework (CIF)	3.2	Anhang, Tab. 7	International Union of Crystallography – CIF; Data name categories
	NeXus	3.2	Anhang, Tab. 8	Könnecke u. a. (2015)
	EngMeta	3.2	Anhang, Abb. 4	Schembera & Iglezakis (2019)

Tab. 2: 15 Elemente von Simple Dublin Core²⁷² mit Definitionen aus derselben Quelle.²⁷³

	DC Element Name	Definition
1.	Title	“A name given to the resource.”
2.	Creator	“An entity primarily responsible for making the resource.”
3.	Subject	“The topic of the resource.”
4.	Description	“An account of the resource.”
5.	Publisher	“An entity responsible for making the resource available.”
6.	Contributor	“An entity responsible for making contributions to the resource.”
7.	Date	“A point or period of time associated with an event in the lifecycle of the resource.”
8.	Type	“The nature or genre of the resource.”
9.	Format	“The file format, physical medium, or dimensions of the resource.”
10.	Identifier	“An unambiguous reference to the resource within a given context.”
11.	Source	“A related resource from which the described resource is derived.”
12.	Language	“A language of the resource.”
13.	Relation	„A related resource.“
14.	Coverage	“The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.”
15.	Rights	“Information about rights held in and over the resource.”

²⁷² Caverlee u. a. (2009, S. 948).

²⁷³ Caverlee u. a. (2009, S. 948).

Tab. 3: DataCite 4.2 Pflicht-Elemente. Die Definitionen sind entsprechend der DataCite Metadata Working Group²⁷⁴ zitiert (vgl. angegebene Seitenzahlen).

Property	Definition
Identifier	“The Identifier is a unique string that identifies a resource. For software, determine whether the identifier is for a specific version of a piece of software, (per the Force11 Software Citation Principles ¹¹), or for all versions.” (S. 12)
Creator	“The main researchers involved in producing the data, or the authors of the publication, in priority order. To supply multiple creators, repeat this property.” (S. 12)
Title	“A name or title by which a resource is known. May be the title of a dataset or the name of a piece of software.” (S. 13)
Publisher	“The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource. This property will be used to formulate the citation, so consider the prominence of the role. For software, use Publisher for the code repository. If there is an entity other than a code repository, that "holds, archives, publishes, prints, distributes, releases, issues, or produces" the code, use the property Contributor/contributorType/hostingInstitution for the code repository.” (S. 13–14)
PublicationYear	“The year when the data was or will be made publicly available. In the case of resources such as software or dynamic data where there may be multiple releases in one year, include the Date/dateType/ dateInformation property and sub-properties to provide more information about the publication or release date details.” (S. 14)
ResourceType	“A description of the resource.” (S. 14)

²⁷⁴ DataCite Metadata Working Group (2019).

Tab. 4 GenBank Sequence Format. Die Kommentare sind zitiert aus dem Sample GenBank record.²⁷⁵

Field	Comments
LOCUS	“The LOCUS field contains a number of different data elements, including locus name, sequence length, molecule type, GenBank division, and modification date. Each element is described below. “
Locus Name	“The RefSeq database of reference sequences assigns formal locus names to each record, based on gene symbol. RefSeq is separate from the GenBank database, but contains cross-references to corresponding GenBank records.”
Sequence Length	“Number of nucleotide base pairs (or amino acid residues) in the sequence record.”
Molecule Type	“The type of molecule that was sequenced.”
GenBank Division	“The GenBank division to which a record belongs is indicated with a three letter abbreviation.”
Modification Date	“The date in the LOCUS field is the date of last modification. “
DEFINITION	“Brief description of sequence; includes information such as source organism, gene name/protein name, or some description of the sequence's function (if the sequence is non-coding).”
ACCESSION	“The unique identifier for a sequence record.”
VERSION	“A nucleotide sequence identification number that represents a single, specific sequence in the GenBank database.”
GI	“”GenInfo Identifier” sequence identification number, in this case, for the nucleotide sequence. If a sequence changes in any way, a new GI number will be assigned.”
KEYWORDS	“Word or phrase describing the sequence.”
SOURCE	“Free-format information including an abbreviated form of the organism name, sometimes followed by a molecule type.”
Organism	“The formal scientific name for the source organism (genus and species, where appropriate) and its lineage, based on the phylogenetic classification scheme used in the NCBI Taxonomy Database.”

²⁷⁵ National Center for Biotechnology Information - Sample GenBank record. GenBank flat file format.

Field	Comments
REFE- RENCE	“Publications by the authors of the sequence that discuss the data reported in the record.”
AU- THORS	“List of authors in the order in which they appear in the cited article.”
TITLE	“Title of the published work or tentative title of an unpublished work.”
JOURNAL	“MEDLINE abbreviation of the journal name.”
PUBMED	“PubMed Identifier (PMID).“
Direct Sub- mission	“Contact information of the submitter, such as institute/department and postal address.”
FEATURES	“Information about genes and gene products, as well as regions of biological significance reported in the sequence.”
source	“Mandatory feature in each record that summarizes the length of the sequence, scientific name of the source organism, and Taxon ID number.”
Taxon	“A stable unique identification number for the taxon of the source organism.”
CDS	“Coding sequence; region of nucleotides that corresponds with the sequence of amino acids in a protein (location includes start and stop codons).”
<1..206	“Base span of the biological feature indicated to the left, in this case, a CDS feature.”
protein_id	“A protein sequence identification number, similar to the Version number of a nucleotide sequence.”
GI	“"GenInfo Identifier" sequence identification number, in this case, for the protein translation.”
translation	“The amino acid translation corresponding to the nucleotide coding sequence (CDS).“
gene	“A region of biological interest identified as a gene and for which a name has been assigned.”
comple- ment	“Indicates that the feature is located on the complementary strand.”
Other Fea- tures	
ORIGIN	“The ORIGIN may be left blank, may appear as "Unreported," or may give a local pointer to the sequence start, usually involving an experimentally determined restriction cleavage site or the genetic locus (if available).”

Tab. 5 SBML (Level 3 Core, Ver. 2)-Komponenten. Die Bedeutungen sind entsprechend der angegebenen Seitenzahlen zitiert aus der Language Specification von SBML.²⁷⁶

Compo- nent	Meaning
Function definition	“A named mathematical function that may be used throughout the rest of a model.” (S.9)
Unit defini- tion	“A named denition of a new unit of measurement.” (S. 9)
Compart- ment	“A well-stirred container of finite size where species may be located. Compartments may or may not represent actual physical structures.” (S. 9)
Species	“A pool of entities of the same kind located in a compartment and participating in reactions (processes).” (S. 9)
Parameter	“A quantity with a symbolic name. In SBML, the termparameteris used in a generic sense to refer to named quantities regardless of whether they are constants or variables in a model.” (S. 9)
Initial As- signment	“A mathematical expression used to determine the initial conditions of a model.” (S. 9)
Rule	“A mathematical expression added to the set of equations constructed based on the reactions defined in a model.” (S. 10)
Constraint	“A means of detecting out-of-bounds conditions during a dynamical simulation and optionally issuing diagnostic messages.” (S. 10)
Reaction	“A statement describing some transformation, transport or binding process that can change the amount of one or more species.” (S. 10)
Event	“A statement describing an instantaneous, discontinuous change in one or more symbols of any type (species, compartment, parameter, etc.) when a triggering condition is satisfied.” (S. 10)

²⁷⁶ Michael Hucka u. a. (2019).

Tab. 6 Darwin Core Elemente in der Record-level-Klasse (Ver. 2014-12-23). Die Definitionen sind zitiert aus dem Darwin Core Quick Reference Guide.²⁷⁷ Weitere Klassen wie Occurrence und Organism sind für Darwin Core spezifiziert, wurden aber nicht in diese Tabelle mitaufgenommen.²⁷⁸

Term	Definition
type	“The nature or genre of the resource.”
modified	“The most recent date-time on which the resource was changed.”
language	“A language of the resource.”
license	“A legal document giving official permission to do something with the resource.”
rightsHolder	“A person or organization owning or managing rights over the resource.”
accessRights	“Information about who can access the resource or an indication of its security status. Access Rights may include information regarding access or restrictions based on privacy, security, or other policies.”
bibliographicCitation	“A bibliographic reference for the resource as a statement indicating how this record should be cited (attributed) when used.”
references	“A related resource that is referenced, cited, or otherwise pointed to by the described resource.”
institutionID	“An identifier for the institution having custody of the object(s) or information referred to in the record.”
collectionID	“An identifier for the collection or dataset from which the record was derived.”
datasetID	“An identifier for the set of data. May be a global unique identifier or an identifier specific to a collection or institution.”
institutionCode	“The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record.”
collectionCode	“The name, acronym, coden, or initialism identifying the collection or data set from which the record was derived.”
datasetName	“The name identifying the data set from which the record was derived.”
ownerInstitutionCode	“The name (or acronym) in use by the institution having ownership of the object(s) or information referred to in the record.”
basisOfRecord	“The specific nature of the data record.”
informationWithheld	“Additional information that exists, but that has not been shared in the given record.”

²⁷⁷ Biodiversity Information Standards - Darwin Core quick reference guide.

²⁷⁸ Biodiversity Information Standards - Darwin Core quick reference guide.

Term	Definition
dataGeneralizations	“Actions taken to make the shared data less specific or complete than in its original form. Suggests that alternative data of higher quality may be available on request.”
dynamicProperties	“A list of additional measurements, facts, characteristics, or assertions about the record. Meant to provide a mechanism for structured content.”

Tab. 7 CIF Data Name Categories. Die Beschreibungen sind zitiert aus International Union of Crystallography – CIF; Data Name Categories.²⁷⁹

Data name	Description
category	
audit	“data provide a record of the CIF creation and subsequent updating”
atom	“data are in two separate categories: those that describe atom sites in a crystal structure (i.e. <code>_atom_site_</code> data names) and those that describe the properties of the atom types that occupy these sites (i.e. <code>_atom_type_</code> data names).”
cell	“data record the cell parameters, method of measurement, conditions etc.”
chemical	“data specify the composition and chemical properties of the compound. The <code>_chemical_formula_</code> items must agree with those that specify the density, unit-cell and Z values.”
chemical_conn	“data specify the 2D chemical structure for molecular species.”
computing	“data record the computer programs used in the crystal structure analysis.”
database	“data are only specified by database managers and should only appear in a CIF if they originate from this source.”
diffrn	“data are the diffraction measurements.”
exptl	“data record the crystal measurements, such as density, shape, size etc.”
geom	“data describe the standard molecular and crystal geometry, as calculated from the contents of the <code>_atom_</code> , <code>_cell_</code> and <code>_symmetry_</code> data.”
journal	“data are entries used by the journals' staff in processing a CIF.”
publ	“data are used when submitting CIF data to a journal for publication.”
refine	“data describe the structure refinement parameters.”
refln	“data specify the reflection items used to determine the <code>_atom_</code> data items. The <code>_refln_</code> data refer to individual reflections and must be included in looped lists.”
reflns	“data specify the reflection items used to determine the <code>_atom_</code> data items. The <code>_reflns_</code> data specify the parameters that apply to all reflections.”
symmetry	“data specify the space-group symmetry.”

²⁷⁹ International Union of Crystallography - CIF; Data name categories.

Tab. 8 NeXus Base Class Definitions (Ver. 2020.1). Die Beschreibungen sind zitiert aus dem NeXus-Manual.²⁸⁰ Innerhalb der Klassen werden Applikationen verwendet, die in dieser Tabelle nicht aufgeführt sind.²⁸¹

Base Class Definitions	Description
NXaperture	“A beamline aperture.“
NXattenuator	“A device that reduces the intensity of a beam by attenuation.“
NXbeam	“Properties of the neutron or X-ray beam at a given location.“
NXbeam_stop	“A device that blocks the beam completely, usually to protect a detector.“
NXbending_magnet	“A bending magnet.“
NXcapillary	“A capillary lens to focus the X-ray beam.“
NXcite	“A literature reference.“
NXcollection	“An unvalidated set of terms, such as the description of a beam line.“
NXcollimator	“A beamline collimator.“
NXcrystal	“A crystal monochromator or analyzer.“
NXcylindrical_geometry	“Geometry description for cylindrical shapes.“
NXdata	“NXdata describes the plottable data and related dimension scales.“
NXdetector	“A detector, detector bank, or multidetector.“
NXdetector_group	“Logical grouping of detector elements.“
NXdetector_module	“Geometry and logical description of a detector module.“
NXdisk_chopper	“A device blocking the beam in a temporal periodic pattern.“
NXentry	“(required) NXentry describes the measurement.“
NXenvironment	“Parameters for controlling external conditions.“
NXevent_data	“NXevent_data is a special group for storing data from neutron.“
NXfermi_chopper	“A Fermi chopper, possibly with curved slits.“
NXfilter	“For band pass beam filters.“
NXflipper	“A spin flipper.“
NXfresnel_zone_plate	“A fresnel zone plate.“
NXgeometry	“legacy class - recommend to use NXtransformations now.“

²⁸⁰ NeXus - 3.3.1. Base class definitions.

²⁸¹ NeXus - 3.3.2. Application definitions.

Base Class	Definition- Description
NXgrating	“A diffraction grating, as could be used in a soft X-ray monochromator”
NXguide	“A neutron optical element to direct the path of the beam.”
NXinsertion_device	“An insertion device, as used in a synchrotron light source.”
NXinstrument	“Collection of the components of the instrument or beamline.”
NXlog	“Information recorded as a function of time.”
NXmirror	“A beamline mirror or supermirror.”
NXmoderator	“A neutron moderator.”
NXmonitor	“A monitor of incident beam data.”
NXmonochromator	“A wavelength defining device.”
NXnote	“Any additional freeform information not covered by the other base classes.”
NXobject	“This is the base object of NeXus”
NXoff_geometry	“Geometry (shape) description.”
NXorientation	“legacy class - recommend to use NXtransformations now.”
NXparameters	“Container for parameters, usually used in processing or analysis.”
NXpdb	“A NeXus transliteration of a PDB file, to be validated only as a PDB.”
NXpinhole	“A simple pinhole.”
NXpolarizer	“A spin polarizer.”
NXpositioner	“A generic positioner such as a motor or piezo-electric transducer.”
NXprocess	“Document an event of data processing, reconstruction, or analysis for this data.”
NXreflections	“Reflection data from diffraction experiments.”
NXroot	“Definition of the root NeXus group.”
NXsample	“Any information on the sample.”
NXsample_component	“One group like this per component can be recorded For a sample consisting of multiple components.”
NXsensor	“A sensor used to monitor an external condition.”
NXshape	“legacy class - (used by NXgeometry) - the shape and size of a component.”
NXslit	“A simple slit.”
NXsource	“The neutron or x-ray storage ring/facility.”
NXsubentry	“Group of multiple application definitions for “multi-modal” (e.g. SAXS/WAXS) measurements.”

Base Class Definiti- Description

ons

NXtransformations “Collection of axis-based translations and rotations to describe a geometry.”

NXtranslation “legacy class - (used by NXgeometry) - general spatial location of a component.”

NXuser “Contact information for a user.”

NXvelocity_selector “A neutron velocity selector”

NXxraylens “An X-ray lens, typically at a synchrotron X-ray beam line.”

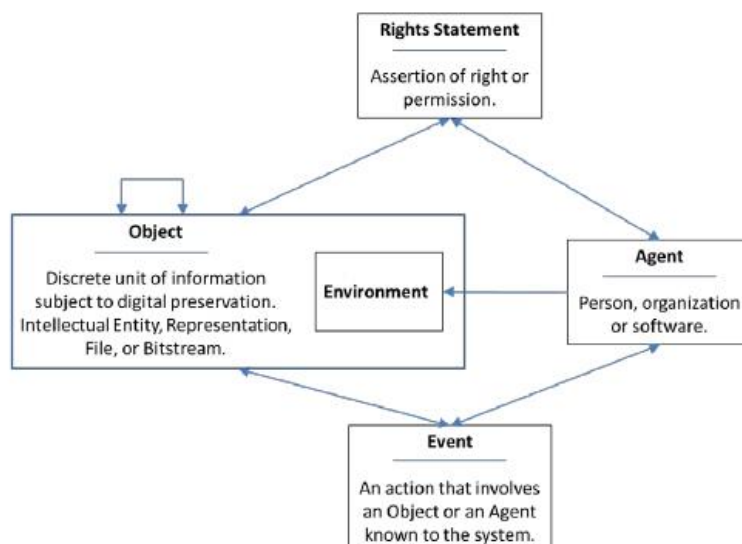


Abb. 1 PREMIS Datenmodell, Version 3:²⁸² Die Boxen repräsentieren Entitäten, während Pfeile die Beziehung zwischen den Entitäten darstellen.²⁸³ Abbildung entnommen aus dem PREMIS Data Dictionary for Preservation Metadata, Seite 6.²⁸⁴

²⁸² PREMIS Editorial Committee (2015, S. 6).

²⁸³ PREMIS Editorial Committee (2015, S. 6).

²⁸⁴ PREMIS Editorial Committee (2015, S. 6)

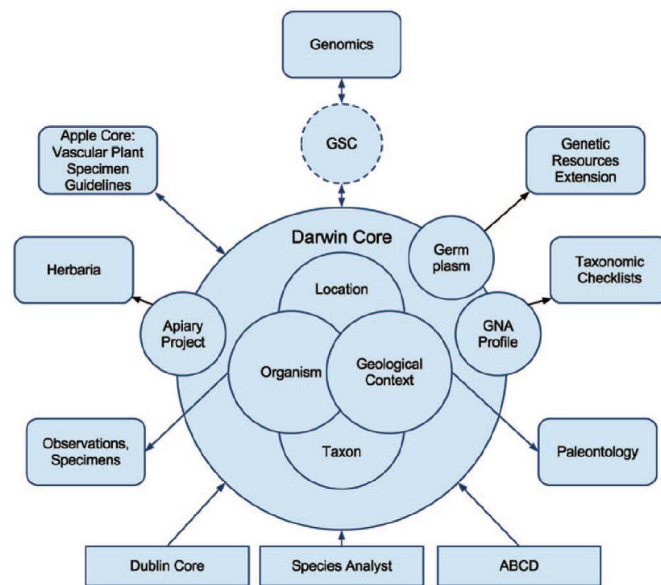


Abb. 2 Darwin Core, inklusive der Zusammenhänge mit anderen Standards und Schemata wie Dublin Core.²⁸⁵ Abbildung entnommen aus Wiczorek et al. (2012), Seite 2.²⁸⁶

²⁸⁵ Wiczorek u. a. (2012, S. 2).

²⁸⁶ Wiczorek u. a. (2012, S. 2)

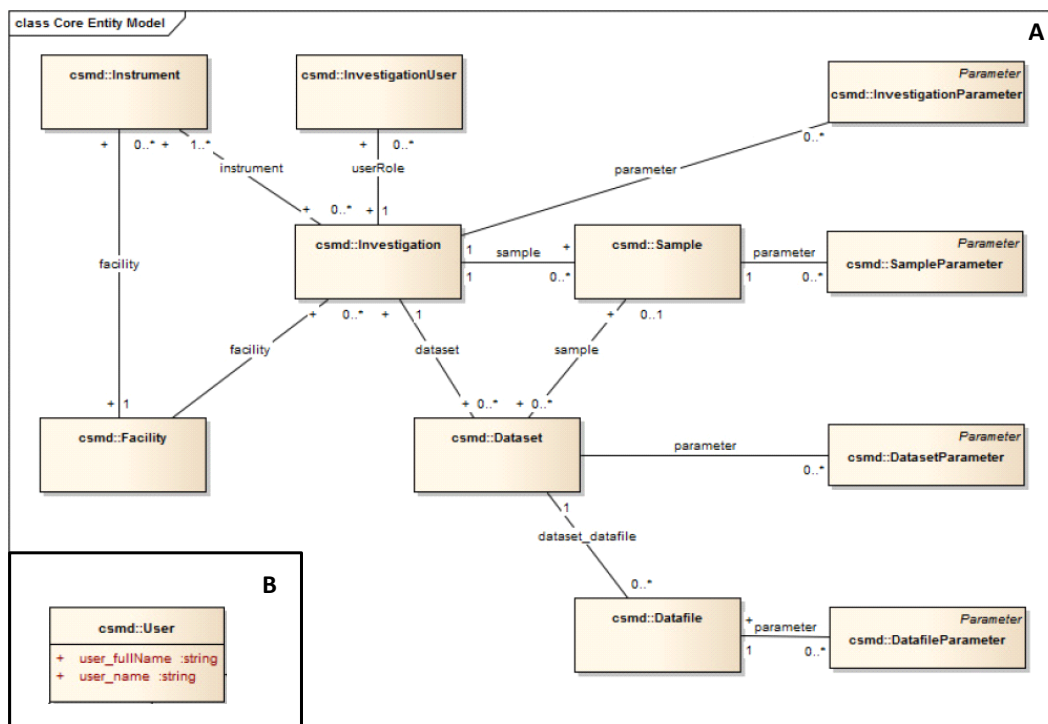


Abb. 3 CSMD-4 Hauptelemente im class Core Entity Model (A)²⁸⁷ mit dem Element „User“ aus dem „UML Class Diagram for the full CSMD“²⁸⁸ als Einsatz (B). Abbildungskomponenten entnommen aus Matthews & Fisher (2013), Seite 2 und 3.²⁸⁹

²⁸⁷ Matthews & Fisher (2013, S. 2).

²⁸⁸ Matthews & Fisher (2013, S. 3–4).

²⁸⁹ Matthews & Fisher (2013, S. 2–3)

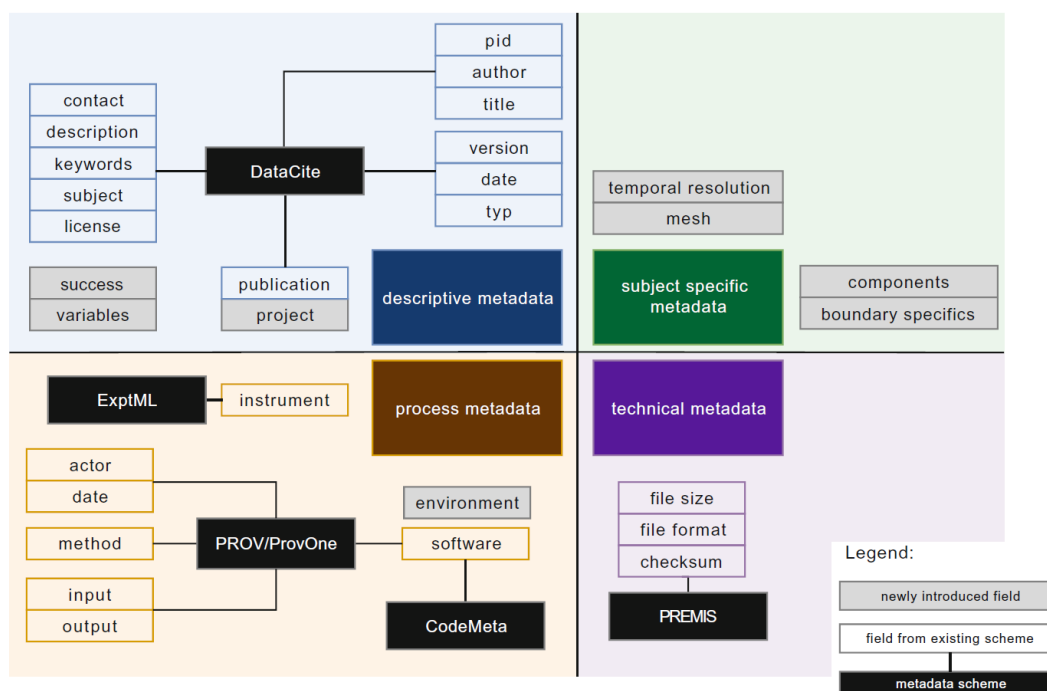


Abb. 4 Komponenten und Felder des EngMeta Metadatenmodells.²⁹⁰ Abbildung entnommen aus Schembera & Iglezakis (2019), Seite 130.²⁹¹

²⁹⁰ Schembera & Iglezakis (2019, S. 130).

²⁹¹ Schembera & Iglezakis (2019, S. 130)

Literatur

- Akers, K. G. & Green, J. A. (2014). Towards a symbiotic relationship between academic libraries and disciplinary data repositories: A Dryad and University of Michigan case study. *International Journal of Digital Curation*, 9(1), 119–131. doi:10.2218/ijdc.v9i1.306
- Antoniou, G. & Plexousakis, D. (2009). Semantic Web. In L. Liu & M. T. Özsu (Hrsg.), *Encyclopedia of Database Systems* (S. 2579–2583). Boston: Springer US. doi:10.1007/978-0-387-39940-9_1320
- Arroyo, S. & Siorpaes, K. (2014). Ontologies and ontology languages. In M.-A. Sicilia (Hrsg.), *Handbook of Metadata, Semantics and Ontologies* (S. 141–156). New Jersey; London; Singapore; Beijing; Shanghai; Hong Kong; Taipei; Chennai: World Scientific Publishing Co.
- Ashburner, M. u. a. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), 25–29. doi:10.1038/75556
- Austin, C. C. u. a. (2017). Key components of data publishing: using current best practices to develop a reference model for data publishing. *International Journal on Digital Libraries*, 18(2), 77–92. doi:10.1007/s00799-016-0178-2
- Ball, Alex u. a. (2016). *RDA Metadata Standards Directory working group: Final report*. <https://www.rd-alliance.org/system/files/MSDWG-Final-Report.pdf> (abgerufen am 03. 02. 2020).
- Ball, Alex & Neilson, C. (2010). *Curation of research data in the disciplines of engineering. SCARP case study 7. Version 1.2*. http://www.dcc.ac.uk/sites/default/files/documents/publications/case-studies/SCARP_B4812_EngCase_v1_2.pdf (abgerufen am 06. 01. 2020).
- Ball, Alexander u. a. (2012). Visualizing research data records for their better management. *Journal of Digital Information*, 13(1). <https://journals.tdl.org/jodi/index.php/jodi/article/view/5917/5892> (abgerufen am 06. 01. 2020).
- Ball, Alexander u. a. (2014). Building a disciplinary Metadata Standards Directory. *International Journal of Digital Curation*, 9(1), 142–151. doi:10.2218/ijdc.v9i1.308
- Bauch, A. u. a. (2011). OpenBIS: A flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics*, 12(1), 1–19. doi:10.1186/1471-2105-12-468
- Benson, D. A. u. a. (2011). GenBank. *Nucleic Acids Research*, 39(SUPPL. 1), D32–D37. doi:10.1093/nar/gkq1079
- Berman, F. (2019). *The Research Data Alliance – The first five years*. <https://www.rd-alliance.org/research-data-alliance—first-five-years> (abgerufen am 11. 01. 2020).
- Bibliothek der Universität Stuttgart - RePlay-DH; Projektbeschreibung. <https://www.ub.uni-stuttgart.de/forschen-publizieren/forschungsdatenmanagement/projekte/replay/> (abgerufen am 21. 02. 2020).
- Biodiversity Information Standards - Darwin Core quick reference guide. <http://rs.tdwg.org/dwc/terms/> (abgerufen am 12. 02. 2020).
- Biodiversity Information Standards - <https://www.tdwg.org/>. (angerufen am 21. 02. 2020)
- Bioschemas - What Is Bioschemas? <https://bioschemas.org/> (abgerufen am 12. 02. 2020).
- Blischak, J. D., Davenport, E. R. & Wilson, G. (2016). A quick introduction to version control with Git and GitHub. *PLoS Computational Biology*, 12(1), e1004668. doi:10.1371/journal.pcbi.1004668

- Brooksbank, C. & Quackenbush, J. (2006). Data standards: A call to action. *OMICS: A Journal of Integrative Biology*, 10(2), 94–99. doi:10.1089/omi.2006.10.94
- Candela, L. u. a. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology*, 66(9), 1747–1762. doi:10.1002/asi.23358
- Caplan, P. (2017). *Understanding PREMIS*. Washington (DC).
<https://www.loc.gov/standards/premis/understanding-premis-rev2017.pdf> (abgerufen am 24. 10. 2020).
- Caracciolo, C. u. a. (2019). Semantics for data in agriculture: A community-based wish list. In E. Garoufallou u. a. (Hrsg.), *Metadata and Semantic Research. MTSR 2018. Communications in Computer and Information Science* (Bd. 846, S. 340–345). Cham: Springer. doi:10.1007/978-3-030-14401-2_32
- Castro, J. A. u. a. (2017). Involving data creators in an ontology-based design process for metadata models. In M. C. Malta, A. A. Baptista & P. Walk (Hrsg.), *Developing Metadata Application Profiles* (S. 181–214). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-2221-8.ch008
- Caverlee, J., Mitra, P. & Laarsgard, M. (2009). Dublin Core. In L. Liu & M. T. Özsu (Hrsg.), *Encyclopedia of Database Systems* (S. 947–949). Boston: Springer US. doi:10.1007/978-0-387-39940-9_894
- Chamongsri, N. (2019). Metadata standards for palm leaf manuscripts in asia. In E. Garoufallou u. a. (Hrsg.), *Metadata and Semantic Research. MTSR 2018. Communications in Computer and Information Science* (Bd. 846, S. 242–254). Cham: Springer. doi:10.1007/978-3-030-14401-2_23
- Chemotion - <https://chemotion.net/>. (abgerufen am 21. 02. 2020)
- Coyle, K. (2017). Application profiles: An overview. In M. C. Malta, A. A. Baptista & P. Walk (Hrsg.), *Developing Metadata Application Profiles* (S. 1–15). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-2221-8.ch001
- Data Readiness Group - People. <https://sansonegroup.eng.ox.ac.uk/> (abgerufen am 09. 02. 2020).
- DataCite - Assign DOIs. <https://datacite.org/does.html> (abgerufen am 05. 03. 2020).
- DataCite - Getting Started. <https://support.datacite.org/docs/getting-started> (abgerufen am 01. 11. 2020).
- DataCite Metadata Working Group. (2019). *DataCite metadata schema documentation for the publication and citation of research data. Version 4.2*. <https://doi.org/10.5438/BMJT-BX77> (abgerufen am 09. 02. 2020).
- Deutsche Forschungsgemeinschaft (DFG). (2019). *Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex*.
https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf (abgerufen am 15. 02. 2020).
- Deutsche Forschungsgemeinschaft (DFG) - Fachsystematik.
https://www.dfg.de/dfg_profil/gremien/fachkollegien/faecher/index.jsp (abgerufen am 13. 12. 2020).
- Deutsches Institut für Normung e.V. (DIN) - ISO 23081.1.
<https://www.din.de/de/mitwirken/normenausschuesse/nid/normen/wdc-beuth:din21:282205888> (abgerufen am 13. 02. 2020).

- Gärtner, M., Hahn, U. & Hermann, S. (2018). Supporting sustainable process documentation. In G. Rehm & T. Declerck (Hrsg.), *Language Technologies for the Challenges of the Digital Age. GSCL 2017. Lecture Notes in Computer Science* (Bd. 10713, S. 284–291). Cham: Springer Verlag. doi:10.1007/978-3-319-73706-5_24
- Global Biodiversity Information Facility - Data Standards. <https://www.gbif.org/standards> (abgerufen am 12. 02. 2020).
- Global Biodiversity Information Facility - What is GBIF? (abgerufen am 12. 02. 2020)
- GO FAIR - GO FAIR Initiative. <https://www.go-fair.org/go-fair-initiative/> (abgerufen am 05. 03. 2020).
- Graham, C. u. a. (2011). Species distribution modeling and the challenge of predicting future distributions. In S. Herzog u. a. (Hrsg.), *Climate change and biodiversity in the tropical Andes* (S. 295–310). Montevideo ; Paris: Inter-American Institute for Global Change Research (IAI) and Scientific Committee on Problems of the Environment (SCOPE). doi:10.13140/2.1.3718.4969
- Gray, A. J. G., Goble, C. & Jimenez, R. C. (2017). Bioschemas: From potato salad to protein annotation. In N. Nikitina u. a. (Hrsg.), *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks. CEUR Workshop Proceedings* (S. 1963:1–4). Aachen: CEUR-WS, RWTH. <http://ceur-ws.org/Vol-1963/#paper579> (abgerufen am 16. 02. 2020).
- Greenberg, J., Swauger, S. & Feinstein, E. (2013). Metadata capital in a data repository. In *International Conference on Dublin Core and Metadata Applications* (S. 140–150). <https://dcpapers.dublincore.org/pubs/article/view/3678> (abgerufen am 05. 01. 2020).
- Group On Earth Observation - GEO Community. https://www.earthobservations.org/geo_community.php (abgerufen am 12. 02. 2020).
- Gruber, T. (2009). Ontology. In L. Liu & M. T. Özsu (Hrsg.), *Encyclopedia of Database Systems* (S. 1963–1965). Boston: Springer US. doi:10.1007/978-0-387-39940-9_1318
- Guha, Guha, R. V., Brickley, D. & Macbeth, S. (2016). Schema.org: Evolution of structured data on the web. *Communications of the ACM*, 59(2), 44–51. doi:10.1145/2844544
- Harvard University - Harvard Biomedical Data Management; Electronic lab notebooks. <https://datamanagement.hms.harvard.edu/electronic-lab-notebooks> (abgerufen am 14. 02. 2020).
- Heery, R. & Patel, M. (2000). Application profiles: Mixing and matching metadata schemas. *Ariadne*, (25). <http://www.ariadne.ac.uk/issue/25/app-profiles/> (abgerufen am 13. 08. 2020).
- Hermann, S. u. a. (2018). Nachträglich ist nicht gleich nachnutzbar: Ansätze für integrierte Prozessdokumentation im Forschungsalltag. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB*, 5(3), 32–45. doi:10.5282/O-BIB/2018H3S32-45
- Hermann, S., Iglezakis, D. & Seeland, A. (2019). Requirements for finding research data and software. *Proceedings in Applied Mathematics & Mechanics*, 19(1), e201900480. doi:10.1002/pamm.201900480
- Holetschek, J. u. a. (2012). The ABCD of primary biodiversity data access. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology*, 146(4), 771–779. doi:10.1080/11263504.2012.740085

- Howard, T. u. a. (2010a). *Understanding and characterizing engineering research data for its better management. Version 1.0*. Bath: University of Bath. <https://researchportal.bath.ac.uk/en/publications/understanding-and-characterizing-engineering-research-data-for-it> (abgerufen am 06. 01. 2020).
- Howard, T. u. a. (2010b). *Opportunities for and barriers to engineering research data re-use. Version 1.0*. Bath: University of Bath. <https://researchportal.bath.ac.uk/en/publications/opportunities-for-and-barriers-to-engineering-research-data-re-us> (abgerufen am 06. 01. 2020).
- Howe, D. u. a. (2008). The future of biocuration. *Nature*, 455, 47–50. doi:10.1038/455047a
- Hucka, M. u. a. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4), 524–531. doi:10.1093/bioinformatics/btg015
- Hucka, Michael u. a. (2019). The Systems Biology Markup Language (SBML): Language specification for level 3 version 2 core. <http://sbml.org/Special/specifications/sbml-level-3/version-2/core/release-2/sbml-level-3-version-2-release-2-core.pdf> (abgerufen am 11. 02. 2020).
- Hugo, W. u. a. (2017). Global infrastructures for biodiversity data and services. In M. Walters & R. J. Scholes (Hrsg.), *The GEO Handbook on Biodiversity Observation Networks* (S. 259–291). Cham: Springer International Publishing. doi:10.1007/978-3-319-27288-7_11
- Iglezakis, D. & Schembera, B. (2018). Anforderungen der Ingenieurwissenschaften an das Forschungsdatenmanagement der Universität Stuttgart - Ergebnisse der Bedarfsanalyse des Projektes DIPL-ING. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB*, 5(3), 46–60. doi:10.5282/O-BIB/2018H3S46-60
- International Organization for Standardization - ISO 15836-1:2017. <https://www.iso.org/standard/71339.html> (abgerufen am 09. 02. 2020).
- International Organization for Standardization - ISO 15836-2:2019. <https://www.iso.org/standard/71341.html> (abgerufen am 09. 02. 2020).
- International Standard Organisation (ISO) - ISO 19115-1:2014. <https://www.iso.org/standard/53798.html> (abgerufen am 13. 08. 2020).
- International Union of Crystallography - CIF; Appendix; CIF Dictionary (Core Version 1991). <http://ww1.iucr.org/iucr-top/cif/standard/cifstd15.html> (abgerufen am 13. 02. 2020).
- International Union of Crystallography - CIF; Data name categories. <http://ww1.iucr.org/iucr-top/cif/standard/cifstd7.html> (abgerufen am 13. 02. 2020).
- Johannes, P. C. u. a. (2013). *Beweissicheres elektronisches Laborbuch: Anforderungen, Konzepte und Umsetzung zur langfristigen, beweiswerterhaltenden Archivierung elektronischer Forschungsdaten und -dokumentation*. Baden-Baden: Nomos Verlagsgesellschaft GmbH & Co. KG.
- Kandogan, E. u. a. (2015). LabBook: Metadata-driven social collaborative data analysis. In H. Ho u. a. (Hrsg.), *2015 IEEE International Conference on Big Data (Big Data)* (S. 431–440). Institute of Electrical and Electronics Engineers Inc. doi:10.1109/BigData.2015.7363784
- Kanza, S. u. a. (2017). Electronic lab notebooks: can they replace paper? *Journal of Cheminformatics*, 9(1), 1–15. doi:10.1186/s13321-017-0221-3
- Kindling, M. u. a. (2017). The landscape of research data repositories in 2015: A re3data analysis. *D-Lib Magazine*, 23(3–4). doi:10.1045/march2017-kindling

- Könnecke, M. u. a. (2015). The NeXus data format. *Journal of Applied Crystallography*, 48(1), 301–305. doi:10.1107/S1600576714027575
- Krause, E. (2016). Elektronische Laborbücher im Forschungsdatenmanagement – Eine neue Aufgabe für Bibliotheken? *ABI Technik*, 36(2), 78–87. doi:10.1515/abitech-2016-0013
- Lautenbacher, C. C. (2006). The Global Earth Observation System of Systems: Science serving society. *Space Policy*, 22(1), 8–11. doi:10.1016/j.spacepol.2005.12.004
- Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek (TIB) - Die FAIR Data Prinzipien für Forschungsdaten. <https://blogs.tib.eu/wp/tib/2017/09/12/die-fair-data-prinzipien-fuer-forschungsdaten/> (abgerufen am 13. 08. 2020).
- Leonelli, S. (2016). *Data-centric biology: a philosophical study*. Chicago ; London: The University of Chicago Press.
- Li, K., Lin, X. & Greenberg, J. (2016). Software citation, reuse and metadata considerations: An exploratory study examining LAMMPS. *Proceedings of the Association for Information Science and Technology*, 53(1), 1–10. doi:10.1002/pr2.2016.14505301072
- Library of Congress - PREMIS Implementation Registry. <https://www.loc.gov/standards/premis/registry/premis-fulllist.php> (abgerufen am 24. 10. 2020).
- Lin, D. u. a. (2020). The TRUST Principles for digital repositories. *Scientific Data*. Nature Research. doi:10.1038/s41597-020-0486-7
- Lindlar, M. u. a. (2020). “You say potato, I say potato“ Mapping digital preservation and research data management concepts towards collective curation and preservation strategies. *International Journal of Digital Curation*, 15(1), 26. doi:10.2218/ijdc.v15i1.728
- Matthews, B. & Fisher, S. (2013). CSMD: the Core Scientific Metadata Model. <http://icatproject-contrib.github.io/CSMD/CSMD-4.0.pdf> (abgerufen am 13. 02. 2020).
- Maxson Jones, K., Ankeny, R. A. & Cook-Deegan, R. (2018). The Bermuda Triangle: The Pragmatics, Policies, and Principles for Data Sharing in the History of the Human Genome Project. *Journal of the History of Biology*, 51(4), 693–805. doi:10.1007/s10739-018-9538-7
- McQuilton, P. u. a. (2016). BioSharing: curated and crowd-sourced metadata standards, databases and data policies in the life sciences. *Database*, 2016, baw075. doi:10.1093/database/baw075
- McQuilton, P. u. a. (2019). Helping the consumers and producers of standards, repositories and policies to enable FAIR data. *Data Intelligence*, 2(1–2), 151–157. doi:10.1162/dint_a_00037
- Metadata Standards Catalog - Biological diversity. <https://rdamsc.bath.ac.uk/subject/Biological%2Bdiversity> (abgerufen am 12. 02. 2020).
- Metadata Standards Catalog - Biology. <https://rdamsc.bath.ac.uk/subject/Biology> (abgerufen am 11. 02. 2020).
- Metadata Standards Catalog - Engineering. <https://rdamsc.bath.ac.uk/subject/Engineering> (abgerufen am 17. 02. 2020).
- Metadata Standards Catalog - Index of subjects. <https://rdamsc.bath.ac.uk/subject-index> (abgerufen am 13. 02. 2020).

- Metadata Standards Catalog - Natural sciences.
<https://rdamsc.bath.ac.uk/subject/Natural%2Bsciences> (abgerufen am 11. 02. 2020).
- Metadata Standards Catalog - Terms of use. <https://rdamsc.bath.ac.uk/terms-of-use> (abgerufen am 09. 02. 2020).
- Müller, U. u. a. (2017). *DINI Certificate for Open Access Repositories and Publication Services 2016*. (Deutsche Initiative für Netzwerkinformation (DINI), Hrsg.). Humboldt-Universität zu Berlin.
<https://doi.org/10.18452/18178> (abgerufen am 25. 10. 2020).
- National Center for Biotechnology Information - Sample GenBank record. GenBank flat file format.
<https://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html> (abgerufen am 11. 02. 2020).
- National Information Standards Organization. (o. J.). *ISO/TC 46/SC11N800R1 Building a metadata schema –Where to start*. Baltimor: National Information Standards Organization (NISO).
[https://committee.iso.org/files/live/sites/tc46sc11/files/documents/N800R1 Where to start-advice on creating a metadata schema.pdf](https://committee.iso.org/files/live/sites/tc46sc11/files/documents/N800R1%20Where%20to%20start-advice%20on%20creating%20a%20metadata%20schema.pdf) (abgerufen am 22. 01. 2020).
- Nature - Supplementary information. <https://www.nature.com/nature/for-authors/supp-info> (abgerufen am 10. 02. 2020).
- Neumann, J. & Brase, J. (2014). DataCite and DOI names for research data. *Journal of Computer-Aided Molecular Design*, 28(10), 1035–1041. doi:10.1007/s10822-014-9776-5
- NeXus - 3.3.1. Base class definitions.
https://manual.nexusformat.org/classes/base_classes/index.html#base-class-definitions (abgerufen am 13. 02. 2020).
- NeXus - 3.3.1.55. NXuser.
https://manual.nexusformat.org/classes/base_classes/NXuser.html#nxuser (abgerufen am 19. 02. 2020).
- NeXus - 3.3.1.7. NXcite. https://manual.nexusformat.org/classes/base_classes/NXcite.html#nxcite (abgerufen am 19. 02. 2020).
- NeXus - 3.3.2. Application definitions.
<https://manual.nexusformat.org/classes/applications/index.html#application-definitions> (abgerufen am 13. 02. 2020).
- Nussbeck, S. Y. u. a. (2014). The laboratory notebook in the 21st century. *EMBO reports*, 15(6), 631–634. doi:10.15252/embr.201338358
- Open Biological and Biomedical Ontology Foundry - <http://www.obofoundry.org/>. (abgerufen am 05. 03. 2020)
- Pampel, H. u. a. (2013). Making research data repositories visible: The re3data.org registry. *PLoS ONE*, 8(11), e78080. doi:10.1371/journal.pone.0078080
- Parsons, M. A. (2013). Building global partnerships-second plenary meeting of the research data alliance. *D-Lib Magazine*, 19(11–12). doi:10.1045/november2013-parsons
- Plant Physiology - Instructions for authors. https://pphys.msubmit.net/cgi-bin/main.plex?form_type=display_auth_instructions (abgerufen am 10. 02. 2020).
- PLOS - Data availability. <https://journals.plos.org/plosone/s/data-availability> (abgerufen am 10. 02. 2020).

- PREMIS Editorial Committee. (2015). *PREMIS Data Dictionary for Preservation Metadata, Version 3.0*. <https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf> (abgerufen am 24. 10. 2020).
- Rat für Informationsinfrastrukturen. (2019). *Herausforderung Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel*. Göttingen. urn:nbn:de:101:1-2019112011541657732737 (abgerufen am 05. 01. 2020).
- RDA Metadata Standards Directory - <http://rd-alliance.github.io/metadata-directory/>. (abgerufen am 18. 02. 2020)
- Reilly, S. u. a. (2011). *Report on integration of data and publications*. Bremerhaven: Alfred-Wegener-Institut. <hdl.handle.net/10013/epic.40198> (abgerufen am 16. 02. 2020).
- Research Data Alliance - About RDA. <https://www.rd-alliance.org/about-rda> (abgerufen am 18. 02. 2020).
- Research Data Alliance - Arbeitsgruppe FAIRSharing Registry: connecting data policies, standards & databases. <https://www.rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html> (abgerufen am 09. 02. 2020).
- Research Data Alliance - Arbeitsgruppe Research Metadata Schemas: Schema crosswalks visualizations. <https://rd-alliance.github.io/Research-Metadata-Schemas-WG/> (abgerufen am 18. 10. 2020).
- Research Data Alliance - <https://www.rd-alliance.org/>. (abgerufen 18. 02. 2020)
- Research Data Alliance - Interessensgruppe Biodiversity Data Integration. <https://www.rd-alliance.org/groups/biodiversity-data-integration-ig.html> (abgerufen am 12. 02. 2020).
- Research Data Alliance - Interessensgruppe Research Data Management in Engineering; Members. <https://www.rd-alliance.org/node/61116/members> (abgerufen am 21. 02. 2020).
- Research Data Alliance - Interessensgruppe Research Data Management in Engineering. <https://www.rd-alliance.org/node/61116/> (abgerufen am 14. 02. 2020).
- Riley, J. (2017). *Understanding metadata : what is metadata, and what is it for?* Baltimore: National Information Standards Organization (NISO). <https://www.niso.org/publications/understanding-metadata-2017> (abgerufen am 19. 01. 2020).
- Salimi, N. & Vita, R. (2006). The biocurator: Connecting and enhancing scientific data. *PLoS Computational Biology*, 2(10), e125. doi:10.1371/journal.pcbi.0020125
- Sansone, S.-A. u. a. (2019). FAIRsharing as a community approach to standards, repositories and policies. *Nature Biotechnology*, 37(4), 358–367. doi:10.1038/s41587-019-0080-8
- Sansone, S.-A. & Rocca-Serra, P. (2016). *Interoperability standards - Digital objects in their own right*. London: Wellcome Trust. <https://doi.org/10.6084/M9.FIGSHARE.4055496.V1> (abgerufen am 02. 02. 2020).
- Schembera, B. & Iglezakis, D. (2019). The genesis of EngMeta - A metadata model for research data in computational engineering. In E. Garoufallou u. a. (Hrsg.), *Metadata and Semantic Research. MTSR 2018. Communications in Computer and Information Science* (Bd. 846, S. 127–132). Cham: Springer. doi:10.1007/978-3-030-14401-2_12
- Scott, M. u. a. (2014). A framework for user driven data management. *Information Systems*, 42, 36–58. doi:10.1016/j.is.2013.11.004

- Scott, M., Johnston, S. J. & Cox, S. J. (2017). Metagit: Decentralised metadata management with Git. *Information Systems*, 65, 78–92. doi:10.1016/j.is.2016.09.002
- Simons, N. & Richardson, J. (2013). *New content in digital repositories: The changing research landscape. Chandos Information Professional Series*. Oxford; Cambridge; New Delhi: Chandos Publishing. doi:10.1533/9781780634098
- Specka, X. u. a. (2019). The BonaRes metadata schema for geospatial soil-agricultural research data – Merging INSPIRE and DataCite metadata schemes. *Computers and Geosciences*, 132, 33–41. doi:10.1016/j.cageo.2019.07.005
- Sugimoto, S. u. a. (2016). Permanence and temporal interoperability of metadata in the linked open data environment. In *International Conference on Dublin Core and Metadata Applications* (S. 45–54). <https://dcpapers.dublincore.org/pubs/article/view/3822> (abgerufen am 26. 01. 2020).
- Taylor, C. F. u. a. (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. *Nature Biotechnology*, 26(8), 889–896. doi:10.1038/nbt.1411
- Thalhath, N. u. a. (2019). Yet Another Metadata Application Profile (YAMA): Authoring, versioning and publishing of application profiles. *International Conference on Dublin Core and Metadata Applications*, 114–125. <https://dcpapers.dublincore.org/pubs/article/view/4055> (abgerufen am 05. 02. 2020).
- Tremouilhac, P. u. a. (2017). Chemotion ELN: an Open Source electronic lab notebook for chemists in academia. *Journal of Cheminformatics*, 9(1), 54. doi:10.1186/s13321-017-0240-0
- Tristram, F. u. a. (2015). *Öffentlicher Abschlussbericht von bwFDM-Communities - Wissenschaftliches Datenmanagement an den Universitäten Baden-Württembergs*. Karlsruhe: Karlsruher Institut für Technologie (KIT). <https://doi.org/10.5445/IR/1000083272> (abgerufen am 16. 02. 2020).
- Whyte, A. & Pryor, G. (2011). Open science in practice: Researcher perspectives and participation. *International Journal of Digital Curation*, 6(1), 199–213. doi:10.2218/ijdc.v6i1.182
- Wieczorek, J. u. a. (2012). Darwin core: An evolving community-developed biodiversity data standard. *PLoS ONE*, 7(1), e29715. doi:10.1371/journal.pone.0029715
- Wilkinson, M. D. u. a. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 1–9. doi:10.1038/sdata.2016.18
- Wissenschaftliche Kommission Niedersachsen. (2001a). *Forschungsevaluation an niedersächsischen Hochschulen. Biologie. Bericht der Gutachter*. Hannover: Niedersächsisches Ministerium für Wissenschaft und Kultur. https://www.wk.niedersachsen.de/publikationen/evaluationsberichte/faecherbezogene_forschungsevaluation_an_universitaeten/evaluationsberichte-137324.html (abgerufen am 28. 01. 2020).
- Wissenschaftliche Kommission Niedersachsen. (2001b). *Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen. Ingenieurwissenschaften. Berichte und Empfehlungen*. Hannover: Niedersächsisches Ministerium für Wissenschaft und Kultur. https://www.wk.niedersachsen.de/taetigkeitsbereiche/wissenschaftspolitische_fragen/empfehlungen-zu-wissenschaftspolitischen-themen-164545.html (abgerufen am 28. 01. 2020).