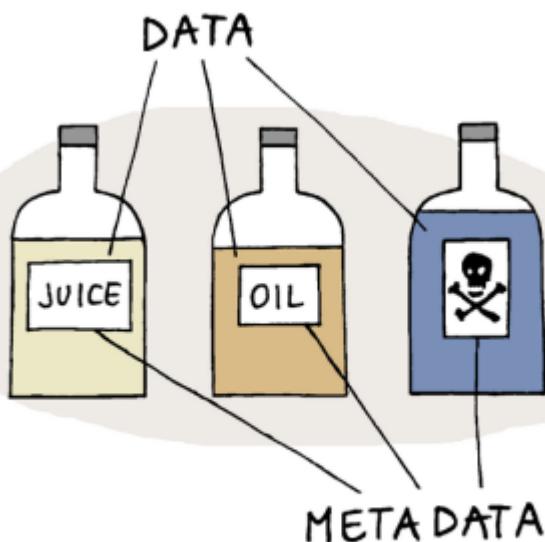


Was sind Metadaten?



 Dataedo /cartoon

Piotr Datnado

Quelle: <https://dataedo.com/cartoon/data-vs-metadata-4/>



Bei Metadaten handelt sich um „strukturierte Informationen über andere Daten bzw. Ressourcen und deren Merkmale“ (siehe <https://forschungsdaten.info/praxis-kompakt/glossar/>).

Als Ressourcen sind hier v.a. Forschungsdaten gemeint, aber auch für Publikationen, Poster, Abstracts und weitere wissenschaftliche Veröffentlichungen spielen Metadaten eine bedeutende Rolle. Metadaten erleichtern das Auffinden in Suchmaschinen und Repositoryn und das Instituts- und disziplinübergreifende Verständnis vorliegender Forschungsdaten und machen ihre Interpretation und eine spätere Nachnutzung durch andere Personen überhaupt erst möglich. Vor allem in großen mehrjährigen Forschungsprojekten mit verschiedenen Mitarbeitenden erleichtert das einheitliche Verwenden von Metadaten das Verständnis über die erstellten und gespeicherten Forschungsdaten enorm. Sie bieten Kontext und Zusatzinformationen zu vorhandenen Daten und sind damit im weitesten Sinne „**Daten über Daten**“. Die Vergabe von Metadaten ist ein Baustein zur Umsetzung der FAIR-Prinzipien [Link zum Wiki-Artikel] bei der Veröffentlichung von Forschungsdaten. Die FAIR-Prinzipien sollen die Auffindbarkeit, die Zugänglichkeit, die Interoperabilität und die Wiederverwendbarkeit von Forschungsdaten sicherstellen. Metadaten sollen zumindest bei den folgenden Fragen (analog den **5W1H-Fragen**) helfen, Antworten zu liefern:



- Wer hat die Daten erstellt?
- Was ist der Inhalt der Daten?
- Wann wurden die Daten erstellt?
- Wo wurden die Daten erstellt (Koordinaten)?
- Wie wurden die Daten erstellt?



Warum wurden die Daten erstellt?

===== Arten von Metadaten ===== Es gibt verschiedene Arten von Metadaten. Dabei wird zwischen bibliographischen und administrativen für alle Formen von Forschungsdaten unterschieden, hinzu kommen je nach Fachdisziplin noch Prozessmetadaten sowie inhaltsbeschreibende bzw. deskriptive Metadaten (vgl. siehe

<https://forschungsdaten.info/themen/beschreiben-und-dokumentieren/metadata-und-metadatenstandards/>) ^Arten von Metadaten ^Beispiele ^Ziele ^ | Bibliografisch | Titel, Autor_innen/Mitwirkende, Abstract, Keywords, Sprache, Identifier [Link zum Wiki-Artikel] | Ermöglichen die Zitation der Daten, Dienen der Auffindbarkeit und thematischen Eingrenzung | | Administrativ | Datum/Zeitpunkt, Ort/Koordinaten, Nutzungsrechte/Lizenzen, Zugriffsrecht, Dateiformat, Dateigröße | Helfen bei der Verwaltung der Daten und deren langfristiger Erhaltung | | Prozessmetadaten | Projekt, Methodik, Arbeitsschritte, Hilfsmittel, Geräte/Instrumente, Anwendungsprogramme | Aufzeigen der verwendeten Methoden und Hilfsmittel bei der Entstehung und Verarbeitung der Daten | | Inhaltsbeschreibende bzw. deskriptive | Fachbegriffe, domänen spezifisches Wissen, Kontextinformationen | Disziplinspezifische Auskunft über zusätzliche Informationen zu Inhalt und Entstehung der Daten |



Tipp: Vergeben Sie Metadaten am besten direkt beim Entstehen der Forschungsdaten. Je größer das Forschungsvorhaben ist und je komplexer und variantenreicher die anfallenden Daten sind, desto schwieriger wird es, dies im Nachhinein zu rekonstruieren.

===== Hilfsmittel bei der Vergabe von Metadaten ===== Es gibt verschiedene Möglichkeiten Metadaten zu erfassen und im Zusammenhang mit den Daten zu speichern: * Metadaten in den Daten (z.B. im Header von Dateien) * Metadaten an den Daten (z.B. im Dateinamen) * Metadaten bei den Daten (z.B. als eigene ReadMe-Datei) * Metadaten in einer Datenbank oder einem Wiki * Verknüpfung zwischen Metadaten mittels Identifier (z.B. DOI) und den Daten Da die meisten der genannten Möglichkeiten nur begrenzte Informationen liefern können, ist ein sinnvolles Zusammenspiel zu empfehlen. Redundante Angaben können dabei nicht vermieden werden, dienen aber dem Abgleich und dürfen sich nicht widersprechen. Werden Metadaten nicht nur menschen-, sondern auch **maschinenlesbar** erstellt, können diese Informationen mittels Schnittstellen über verschiedene Systeme hinweg verknüpft und ausgetauscht werden und erfahren dadurch eine größere Verbreitung. Dabei können Metadaten unabhängig oder zusammen mit den eigentlichen Forschungsdaten gespeichert werden. Zum Beispiel ist das Erstellen einer **ReadMe-Datei** [Link zum Wiki-Artikel] sinnvoll. Durch die Nutzung von **Tools** zur (automatischen) Vergabe von Metadaten kann die Arbeit erheblich erleichtert und die Qualität der Dateneinreichung verbessert werden. Für generische Metadaten eignen sich folgende Services: * [DataCite Metadatengenerator](#): Hier werden Metadaten aus dem Metadatenstandard [Link nach unten] von [DataCite](#) erstellt. Das Ergebnis ist eine XML-Datei mit interoperablen Metadaten. Diese Datei kann zusammen mit den Forschungsdaten und evtl. weiteren Dokumenten gespeichert und veröffentlicht werden. Dieses Videotutorium von FDM Bayern (2020) zeigt die Beschreibung von Forschungsdaten durch die Nutzung des DataCite-Metadatengenerators und des [DataCite-Best-Practice-Guides](#): <https://youtu.be/y7Xullpa6gk> * [CodeMeta generator](#) der [CodeMeta Initiative](#) * weitere Tools für die Metadatenerstellung listet die Research Data Alliance auf: <https://rdamsc.bath.ac.uk/tool-index>

Wozu dienen Metadaten?

- Verwaltung und Organisation von Forschungsdaten sowie Identifizierung, Gruppierung und Kategorisierung von Daten
- Auffindbarkeit: effiziente Recherche nach Forschungsdaten sowie innerhalb größerer Datenmengen
- Nachvollziehbarkeit/Reproduzierbarkeit: aufgrund der Kenntnisse über Prozess der Datenerhebung, -verarbeitung und -analyse
- Nachnutzbarkeit: längerfristige Weiter- und Wiederverwendung von Forschungsdaten aufgrund von Kontext- und Zugriffsmöglichkeiten (inkl. rechtlicher Bedingungen)
- Langzeitarchivierung
- Referenzierung der eigenen Forschungsdaten
- Einheitliche Metadaten erleichtern die (interdisziplinäre und internationale) Kooperation
- Vergleichbar- und Verknüpfbarkeit von Forschungsdaten aus unterschiedlichen Quellen
- Metadatenvergabe ist Teil guter wissenschaftlicher Praxis



===== Was sind Metadatenstandards? =====



Quelle: <https://xkcd.com/927> Für die verschiedenen Anwendungsbereiche werden Metadaten zu generischen oder fachspezifischen Schemas zusammengeführt. Einige solcher Schemas sind zu Standards in fachspezifischen Communitys geworden, andere werden z.B. durch die NFDI (Link zum Wiki-Artikel) (National Forschungsdateninfrastruktur für Deutschland) entwickelt bzw. etabliert. Sie gewährleisten eine konsistente und inhaltlich/strukturell gleichförmige Beschreibung von ähnlichen Datensätzen und dienen damit der Vereinheitlichung, dem Verständnis und damit der besseren Nachnutzbarkeit von Forschungsdaten. Metadatenstandards sind i.d.R. menschen- und maschinenlesbar. **Beispiel:** Die Sprache eines Datensatzes kann in Form von Metadaten erfasst werden, dafür wären die Angaben „deutsch“, „German“, „deu“ oder „ger“ möglich. In einem Metadatenstandard ist die Angabe für die Sprache festgelegt, hier evtl. „deu“ nach der Norm ISO 639-3. Auch Datumsangaben können sehr vielfältig sein und werden durch die Nutzung eines Standards vereinheitlicht. Die verschiedenen Anwendungsbereiche ergeben sich aus den verschiedenen Datenformaten der Forschungsdaten und den verschiedenen Fachdisziplinen, in denen sie entstehen (siehe Abb.). Hier gilt es, den passenden Standard für die eigene Forschung festzulegen



und zu nutzen.

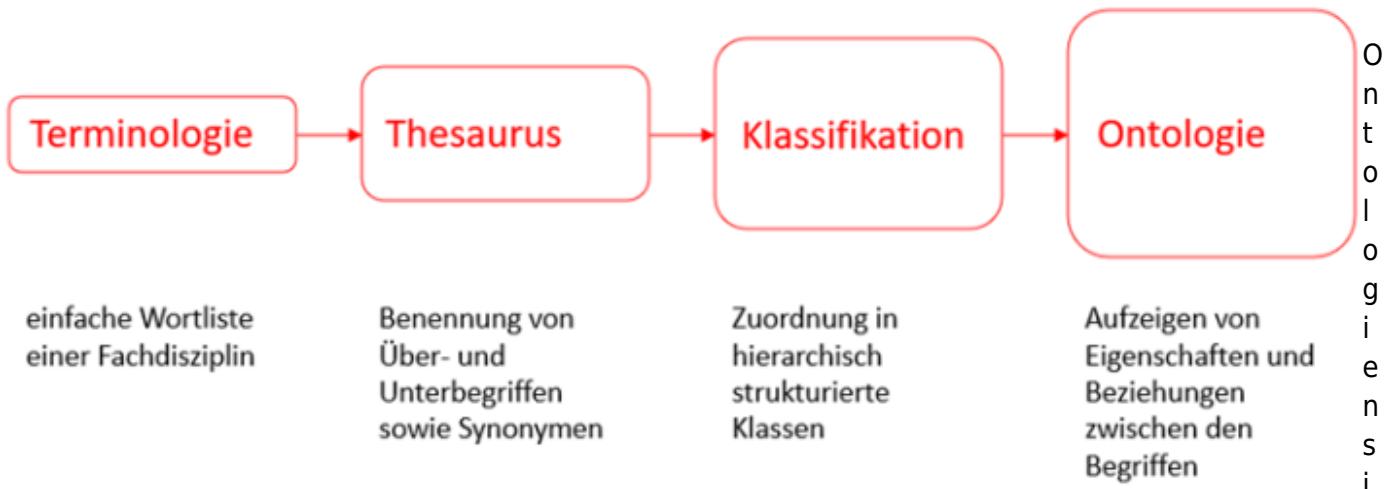
Quelle: Jenn Riley: Seeing Standards: A Visualization of the Metadata Universe.

<https://jennriley.com/metadata-map/> Viele Systeme, in denen Publikationen und/oder Forschungsdaten veröffentlicht werden (z.B. Repositorien), nutzen ein vorhandenes Metadatenschema. Während also alle Felder beim Upload von Veröffentlichungen ausgefüllt werden, werden bereits Metadaten vergeben. ===== Generische Metadatenstandards ===== Fächerübergreifende Metadatenstandards sind für die Beschreibung aller Forschungsdaten nutzbar, können aber nur schlecht disziplinspezifische Inhalte abbilden. Auch die genutzten Metadatenstandards in vielen generischen Repositorien (z.B. [Zenodo](#)) bieten keine fachspezifischen Beschreibungen von Forschungsdaten an und reichen für eine umfassende Beschreibung kaum aus. ^Name des Standards ^Erläuterung ^ | [DataCite](#) | weit verbreiteter Standard, für die bibliographische Beschreibung von Forschungsdaten zur Registrierung von DOIs | | [Dublin Core](#) | Internationaler Standard, für die Beschreibung von elektronischen Ressourcen | | [DCAT](#) | Sehr breit aufgestelltes Metadatenschema, entwickelt von [W3C](#) | | [Schema.org](#) | für die Beschreibung von Webseiten-Text zur weiteren Verwendung durch Internet-Suchmaschinen (wird von den führenden Suchmaschinen-Anbietern genutzt) | | [PREMIS](#) | International anerkannter Standard, für digitale Langzeitarchivierung, basiert auf OAIS (Open Archival Information System) | | [METS](#) | Metadata Encoding & Transmission Standard; ein Container-Standard, der in sieben Abschnitte (Kopfteil, Erschließungsangaben, Verwaltungsangaben, Dateiabschnitt, Strukturbeschreibung, Strukturverknüpfung und Verhalten) geteilt ist, in denen jeweils unterschiedliche Metadatenstandards ausgewählt werden können | ===== Fachspezifische Metadatenstandards ===== Es gibt eine Vielzahl fachspezifischer Metadatenstandards aufgrund der sehr unterschiedlichen Bedürfnisse je Fachcommunity. Die folgende Tabelle zeigt eine kleine Auswahl fachspezifischer Metadatenstandards, orientiert an den Fachgebieten der Jade Hochschule:

^Fachdisziplinen ^Name des Standards ^ | Architektur | [Brick](#) (A Uniform Metadata Schema For Buildings) | | Bildungswissenschaften | [Metadatenset Forschungsdaten Bildung](#) | | Geisteswissenschaften | [EAD](#), [TEI P5](#), [TEI Lex0](#) | | Geowissenschaften | [AgMES](#), [CSDGM](#), ISO 19115 | | Gesundheitswissenschaften/Medizin | [Medical Subject Headings](#) | | Klimawissenschaften | [CF Conventions](#) | | Kunst- und Kulturwissenschaften | [CDWA](#), [MIDAS-Heritage](#) | | Sozial- und Wirtschaftswissenschaften | Data Documentation Initiative (DDI) Standard der [DDI Alliance](#), [TEI P5](#) (Guidelines for Electronic Text Encoding and Interchange) | | Naturwissenschaften | [ICAT Schema](#) im ICAT Project, [Darwin Core](#) (Biodiversity Information Standard), [CIF](#), [CSMD](#), [EML](#) | | Ingenieurwissenschaften | Projekt [AIMS](#) („Applying Interoperable Metadata Standards“) mit dem Ziel zur Schaffung einer Plattform zum Gestalten und Teilen von Metadatenschemata zur Anwendung in den Ingenieurwissenschaften | Quelle: Krähwinkel, E., Langner, P., Lipp, R., & Pietsch, A. M. (2022, März 22). HeFDI Data OER: Forschungsdatenmanagement – eine Online-Einführung. Zenodo.

<https://doi.org/10.5281/zenodo.6373596> Kapitel 4.4 Viele weitere fachspezifische Metadatenstandards sind auf folgenden Seiten aufgelistet und erläutert: * [Fairsharing.org](#) vom e-Research Centre der University of Oxford * [Disciplinary Metadata Catalogue](#) des Digital Curation Centre (DCC) – inkl. Anwendungsprofile, Anwendungsfälle und Tools * [Metadata Directory](#) von der [Metadata Standards Directory Working Group](#) der [Research Data Alliance \(RDA\)](#) – Community-basiert ===== Was sind Terminologien / Ontologien? ===== Metadatenschemata geben eine Struktur für Metadaten vor. Sie legen fest, welche Angaben in welchem Format erfasst werden sollen. Terminologien standardisieren den Inhalt bestimmter Metadatenfelder, d.h. sie bilden Wortlisten, aus denen man wählen kann. So können zum Beispiel unterschiedliche (oder auch fehlerhafte) Schreibweisen vereinheitlicht werden. Sie eignen sich gut, um innerhalb eines Forschungsprojektes die Verwendung von Begriffen

festzulegen und damit einheitlich zu verwenden. Dabei kann die inhaltliche Breite und Tiefe solcher kontrollierten Vokabulare sehr unterschiedlich sein:



nd v.a. relevant bei der maschinellen Verarbeitung großer Datenmengen mittels KI. Einen Überblick über bestehende Terminologien gibt das Basic Register of Thesauri, Ontologies and Classifications BARTOC. Weitere Beispiele für kontrollierte Vokabulare: * [Linked Open Vocabularies \(LOV\)](#) vor allem für den ingenieurwissenschaftlichen Bereich * [Umweltklassifikation](#) und der [Umweltthesaurus](#) vom Umweltbundesamt * Thesaurus für die Sozialwissenschaften von GESIS:

<https://lod.gesis.org/thesoz/de/> * Physics Subject Headings (PhySH): <https://physh.aps.org/> * Mathematics Subject Classification (MSC): <https://mathscinet.ams.org/msc/msc2010.html> * Art and Architecture Thesaurus (AAT): <http://www.aat-deutsch.de/> * Medical Subject Headings (MeSH): <https://www.nlm.nih.gov/mesh/> * Standard Thesaurus Wirtschaft (STW): <https://zbw.eu/stw/version/latest/about.de.html> * Getty Thesaurus of Geographic Names Online * Biodiversity Thesaurus Darüber hinaus ist auch die Verwendung von **Normdaten** zur Beschreibung von Forschungsdaten sinnvoll. Im Unterschied zu den oben genannten Vokabularen, haben Normdaten jeweils eigene IDs, die diese eindeutig referenzierbar machen. Vor allem in der maschinellen Verarbeitung und dem Austausch von Daten zwischen verschiedenen Systemen können Normdaten eindeutig interpretiert werden. Beispiel für Normdaten: * GND – Gemeinsame Normdatei der Deutschen Nationalbibliothek <https://explore.gnd.network/> * Geonames für geographische Begriffe * ORCID für Personendaten * [Open Funder Registry](#) für Forschungsförderer und deren geförderte Projekte Auch die **NFDI** [Link zum Wiki-Artikel] (Nationale Forschungsdateninfrastruktur für Deutschland) hat den Wert von Ontologien und Terminologien erkannt und viele Konsortien bieten eigene Lösungen an bzw. arbeiten derzeit an eigenen Lösungen: NFDI4Ing – NFDI für Ingenieurwissenschaften: Ontologie-Sammlung <https://terminology.nfdi4ing.de/ts/> NFDI4Mat – NFDI für Materialwissenschaft und Werkstoffkunde: Ontologie-Service in Planung NFDI4Cat – NFDI für Wissenschaften rund um Katalyse: Ontologie Sammlung <https://www.nfdi4cat.org/services/ontologie-sammlung/> + (Meta-)Datenstandards basierend auf Use Cases in Planung NFDI4Chem – Chemistry Consortium in the NFDI: Terminology Service <https://terminology.nfdi4chem.de/ts/> NFDI4Biodiversity – Konsortium für Biodiversität und Umweltdaten: Terminologie-Service (teilweise noch Work in Progress) <https://terminologies.gfbio.org/> NFDI4Microbiota – Konsortium für Microbiota-Forschung: (Meta-)Datenstandards und Terminologie-Service in Planung FAIRagro – FAIRe Dateninfrastruktur für die Agrosystemforschung: Terminologie-Service in Planung NFDI4DataScience – NFDI for Data Science and Artificial Intelligence: Nutzung von ORKG <https://orkg.org/> NFDI4Energy – NFDI für die interdisziplinäre Energiesystemforschung: Metadatenregister für Digitale Objekte in Planung

NFDI4Culture – Konsortium für Forschungsdaten zu materiellem und immateriellem kulturellem Erbe:
Antilope zur Terminologie-Suche <https://service.tib.eu/annotation/>
BERD@NFDI – NFDI für Betriebswirtschaftslehre, Volkswirtschaftslehre und verwandte Daten:
Ontologie-Dienst für Daten deutscher Unternehmen (Work in Progress)
Text+ - Konsortium für text- und sprachbasierte Forschungsdaten: föderierte Metadaten- und
Dateninfrastruktur in Planung
Quelle: NFDI-N(HAW)igator, Stand: 03/2025 ===== Empfehlungen =====

From:
<https://forschung-wiki.jade-hs.de/> - **Forschung-Wiki**



Permanent link:
<https://forschung-wiki.jade-hs.de/de/fdm/metadata/start?rev=1751450474>

Last update: **2025/07/02 10:01**