

3.1 Metadaten Grundlagen

Daten managen – 3 Metadaten

Stellen wir uns vor, eine junge Versorgungsforscherin arbeitet an einem Projekt, das die Wirksamkeit neuer Therapien im ländlichen Raum untersuchen soll. Sie hat Zugriff auf eine große Datenbank voller anonymisierter Patient:innenakten, Studien und Befragungen. Doch als sie konkrete Informationen sucht, wie zum Beispiel etwa alle Datensätze über Diabetikerinnen über 65 Jahren in bestimmten Regionen, findet sie fast nichts. Die Daten sind zwar da, aber schlecht beschrieben.

Ein erfahrener Kollege erklärt ihr, dass jeder Datensatz wie ein Buch in einer Bibliothek ist: Ohne Titel, Autorenangabe oder Schlagwörter ist er praktisch unsichtbar. Erst durch Metadaten wie Alter, Geschlecht, Diagnose, Erhebungszeitraum oder Erhebungsmethode könne man erkennen, was in den Daten steckt und sie gezielt auffindbar machen. Unsere Forscherin beschließt, gemeinsam mit ihrem Kollegen, systematisch Metadaten zu ergänzen.

Im Gesundheitswesen existieren unendliche Mengen an Daten, die für die unterschiedlichsten Zwecke erhoben werden: Aus der Versorgung, Verwaltung und Forschung. Eine Zusammenführung dieser vielfältigen Daten könnte Lösungen für vielen noch unbeantwortete Fragen bedeuten, aber dafür müssen diese Daten zugänglich und (ebenso wichtig) auffindbar sein. Um Daten auffinden zu können, müssen sie gut beschrieben sein. Ansonsten bleiben wertvolle Erkenntnisse im Verborgenen – darum brauchen wir Metadaten.

Grundlagen

Metadaten sind „Daten über Daten“. Sie enthalten strukturierte Informationen über Merkmale anderer Ressourcen wie z.B. Datensätze, Dokumente oder Bücher. Die gegebenen Informationen beschreiben, erklären, kontextualisieren oder lokalisieren die Ressource und erleichtern so die Nutzung, das Auffinden, Verständnis, die Nachvollziehbarkeit, Analyse, Qualitätsbewertung oder die Verwaltung. Metadaten können neben wichtigen inhaltlichen und formalen Informationen auch Auskunft über rechtliche und ethische Rahmenbedingungen geben

Arten von Metadaten

- **Bibliografische Metadaten** sind dafür da, Daten zitieren zu können, sie auffindbar und thematisch einordbar zu machen. Zu ihnen gehören Informationen wie Titel, Autor:in, Stichwörter/Keywords, Kurzbeschreibung.
- **Administrative Metadaten** geben Informationen für die Verwaltung und Speicherung von Daten. Sie können z.B. Erstellungsdatum, Ort, Lizenzen und Zugriffsrechte enthalten
- **Kontextuelle Metadaten** beschreiben Schritte der Entstehung und Verarbeitung der Daten wie z.B. Informationen zum Projekt in dem die Daten entstehen, der Methodik oder der Arbeitsschritte
- **Strukturelle Metadaten** geben Informationen zur Datenstruktur. Sie beschreiben beispielsweise das Datenformat, die Tabellenstruktur, ggf. Relationen zu anderen Daten.
- **Technische Metadaten** speichern Informationen zu technischen Details wie Dateigröße, Dateiformat, Hard- oder Software. Viele von ihnen werden automatisch generiert.
- **Deskriptive Metadaten** bieten zusätzliche (meist umfangreichere) Informationen zum Inhalt und der Entstehung der Daten.

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt



Finanziert von der
Europäischen Union
NextGenerationEU

Metadaten über Datensätze sollen vor allem dafür sorgen, dass die Daten langfristig und personenunabhängig interpretierbar bleiben.

Metadaten können entweder in der Datei selber, in einer Readme-Datei, in einer Kalkulationstabelle (Excel-Tabelle), einer XML-Datei oder in einer Datenbank (wenn die Daten geteilt werden) gespeichert werden. Vor allem müssen sie von den Nachnutzenden verstanden werden, aber in der heutigen Zeit wird auch die Maschinenlesbarkeit immer relevanter, wenn die Daten beispielsweise in einem Repositorium veröffentlicht werden sollen. Durch gute Metadaten kann eine Schlagwortsuche oder Filterfunktionen ermöglicht werden, die das Auffinden der Daten erheblich erleichtern.

Persistente Identifikatoren (PID)

PIDs sind eine einzigartige Folge von alphanumerischen Zeichen, die dauerhaft auf eine Ressource verweisen. Die Identifikatoren werden nicht mehr verändert und zeigen auf einen Verzeichnisdienst (Resolver), der die Veränderungen von Zugriffsadressen erfasst. Sie dienen der Vermeidung von "toten" Links und ermöglichen eindeutigen Zuweisungen von Autor:innen auch bei Affiliations- oder Namenswechseln. Es kann auch sein, dass Datensätze unter mehreren URLs zu finden sind, was zu Verwirrung führen kann, weswegen eine eindeutige PID sinnvoll ist.

Beispiele:

DOI – Digital Object Identifier: sind Kennzeichen für digitale Objekte vor allem für wissenschaftliche Artikel und Publikationen. Sie bestehen aus der URL eines DOI Foundation Servers, einem Präfix für die Organisationskennung und einem Suffix für die Objekterkennung, sie sind dauerhaft zitierfähig und bei akademischen Einrichtungen kostenlos; <https://www.doi.org/>

ORCID – Open Researcher and Contributor ID: sind individuelle Identifikatoren für Forscher:innen. Sie werden vergeben um ihnen beispielsweise Publikationen, Forschungsförderungen und Peer Reviews eindeutig zuordnen zu können. Eine ORCID ist ein 16-stelliger alphanumerischer Code und wird von den Forschenden selbst gepflegt und verwaltet. Sie sind nicht an die Organisation gebunden bei der die Forschenden tätig sind (wie z.B. Mailadressen) und ermöglichen so eine dauerhafte und eindeutige Zuordnung. Eine kostenlose Registrierung ist möglich unter <http://orcid.org/>.

ROR ID – Research Organization Registry: sind persistente Identifikatoren für Organisationen wie z.B. Universitäten. Über das [ROR-Verzeichnis](#) lassen sich die Organisationen mit Metadaten wie den gängigen Bezeichnungen, dem Typ der Organisation, dem Standort und der Website finden.

URN – Uniform Resource Name: sind dauerhafte, ortsunabhängige Bezeichner für eine Ressource, meist eine Online-Publikation. Sie geben also einer Ressource einen eindeutigen und dauerhaft gültigen Namen zur eindeutigen Identifikation. URNs werden von der Deutschen Nationalbibliothek verwaltet.

Quellenverzeichnis

Blümm, M., Fritsch, K., Bock, S., Hackenbuchner, J., Arning, U., & Förstner, K. U. (2024).

06_LE_Metadaten. FDM@Studium.nrw Blended-Learning-Basiskurs

„Forschungsdatenmanagement“ (Version 1.0).

https://landesinitiativefdmnrw.github.io/FDMatStudium/thk/texte/06_LE_Metadaten.html

Böker, E. (2022). Metadaten und Metadatenstandards. Forschungsdaten.info. Abgerufen am 30.09.2025, von

<https://forschungsdaten.info/themen/beschreiben-und-dokumentieren/metadaten-und-metadatenstandards/>

Cramer, E., Krumreihn, A., Schaffert, C., & Stiene, M. (2024). Metadaten, Metadatenregister und Datenverknüpfungen im Gesundheitsbereich – eine Einordnung Forschung aktuell. <https://doi.org/10.53190/fa/202403>

DNB (2025). URN-Service. Abgerufen am 30.09.2025, von https://www.dnb.de/DE/Professionell/Services/URN-Service/urn-service_node.html

Forschungsdaten.info (2024). Persistente Identifikatoren. Abgerufen am 30.09.2025, von <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/persistente-identifikatoren/>

Ludwig, J. & Enke, H. (Hrsg.) (2013). Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Verlag Werner Hülsbusch, Glücksstadt.

Mertzen, D., Neuroth, H., Schneemann, C., Woywod, K., Haase, C., Jacob, B., Kroehling, M., Mittelbach, J., Straka, J., Szczukowski, A., & Weise, K. (2003) „Zertifikatskurs ‚Forschungsdatenmanagement für Studierende‘: Spring School 2023 der Landesinitiative für Forschungsdatenmanagement in Brandenburg“. FDM-BB. Zenodo. <https://doi.org/10.5281/zenodo.7936966>

MIT Data Management course (2016). Using Metadata to Find, Interpret & Share Your Data. Abgerufen am 30.09.2025, von <https://www.twillo.de/edu-sharing/components/render/01887843-1814-42e6-9ad5-6941a88d5c5a>

MIT Libraries Data Management Group (2016). Research Data Management: Strategies for Data Sharing and Storage. Using Metadata to Find, Interpret & Share Your Data. Abgerufen am 30.09.2025, von <https://www.twillo.de/edu-sharing/components/render/01887843-1814-42e6-9ad5-6941a88d5c5a>

Schmitz, D., Hausen, D. & Trautwein-Bruns, U. (2018). Forschungsdaten und ihre Metadaten. RWTH Aachen. https://doi.org/10.18154/RWTH-2018_231101

Schembera, B. (2024). Schlüsselqualifikation Forschungsdatenmanagement 4/8 (SQ): Dateninfrastrukturen. Abgerufen am 30.09.2025, von <https://www.oerbw.de/edu-sharing/components/render/7753ff28-917e-4f10-9d9b-34e39217cb93?closeOnBack=true>

Das Material ist im Rahmen des Projekts „DIM.RUHR: Datenkompetenzzentrum für die interprofessionelle Nutzung von Gesundheitsdaten in der Metropole Ruhr“ (Förderkennzeichen: 16DKZ2008[A-F]) entstanden. Dementsprechend spiegelt das Material in Version 1.0 den Stand von 2025 wider.

„**Daten managen - 3.1 Metadaten – Metadaten Grundlagen**“ von Anne Mainz und Sven Meister.

Dieses Werk und dessen Inhalt sind – sofern nicht anders angegeben – lizenziert unter CC BY 4.0. Ausgenommen aus der Lizenz sind die verwendeten Logos.

