

Metadaten für Forschungsdaten: Welche Standards gibt es?

Uwe Jensen, GESIS – Leibniz-Institut für Sozialwissenschaften

24.04 - Metadaten für Forschungsdaten - eine Aufgabe für Information Professionals!?

101. Deutscher Bibliothekartag 2012, Hamburg

Metadaten sind ...

- **Informationen**, die in (möglichst) strukturierter Form analoge oder digitale Forschungsdaten (Objekte) beschreiben.
 z. B. Felderhebungen, Großinstrumente (Teleskope etc.), Simulationen, Laborversuche, Museumsobjekte, Textdokumente ...
 Was sind Forschungsdaten? Nur im Kontext der Disziplin zu beantworten ...
 - **Werkzeuge**, um Forschungsdaten nachhaltig zu managen, zu erschließen, zu verstehen und zu benutzen (NISO 2004)
 - heterogen & vielfältig > **Vielzahl** ist kaum überschaubar ...
- **Wie also mit der Frage in 20 Minuten umgehen?**

Metadaten im Forschungsdatenzyklus

Ansatz: Praktische Frage aus dem Datenmanagement aufgreifen!

- **Welche Metadaten sind** bei der Erzeugung von Forschungsdaten im Forschungsprozess und für die Nachnutzung von Bedeutung?
- Welche **Metadatenstandards** können genutzt werden?

Vorgehen: Beispiele! aus 4 (groben!) Phasen im Datenzyklus aufzeigen:

1. Studiendesign und Erhebung von Forschungsdaten
2. Dokumentation von Forschungsdaten
3. Publikation, Zitation und Recherche von Forschungsdaten
4. Langzeitarchivierung von Forschungsdaten

1. Studiendesign und Datenerhebung

Metadaten zum **Entstehungskontext** der Daten

Metadaten zum Studiendesign:

Forschungsfrage, Theorierahmen, Forschungsstrategie / Design der Daten + der Erhebung (Messinstrument / Methodik), geplante Datenanalyse > Berichte; Forschungsstandards

- **DDI** (Data Documentation Initiative) **Standard zur Dokumentation von Umfragedaten (>Lifecycle):**
- MD zur **Studie als Ganzes** (Thema, Instrument, Sample, Methode) > **Datensatz** > **Variablen**
- MD zu **Messinstrument & Entwicklung**: Fragetext & Antwortkategorien > Versionen > Pre-test
 - **Inter-Kultureller Kontext**: Mehrsprachigkeit > Übersetzungsprozesse (Validität der Frage in) ...

Metadaten über die Datenerhebung und die Bedingungen im ‚Feld‘:

- **ISO 20252:2006**: formuliert **Qualitätskriterien** für die Markt-, Meinungs- & Sozialforschung
- **MD zur Datenerzeugung: Stichprobe, Interviewerschulung, Protokollierung Feldarbeit**:
 - prozessproduzierte **Paradaten** liefern **Metadaten zur Feldarbeit & Datenqualität**
z.B. durch Kontaktdaten, Interview-Ausfälle; schwierige Fragen
 wichtig für Datenanalyse (Gründe Non-Response?), Qualitätssicherung Messinstrument ...

2. Dokumentation von Forschungsdaten

Metadaten zur **Bedeutung** von Daten

Design und Bedeutung der Daten muss methodisch definiert sein

- MD beim Messen: Einheiten numerischer Daten > Indikatoren
- Welche ‚Einheiten‘ entstehen bei Analysen von Bilddaten, Texten ...

Fachspezifische Anforderungen an Metadaten auf verschiedenen Ebenen durch **Metadaten-Regelwerke (Frameworks) spezifiziert**

- Sind ein System sicher ergänzender Standards, Normen, kontrollierter Vokabulare, technischer Verfahren ...

2. Dokumentation von Forschungsdaten

(1) Beispiel: Daten- & Metadatenmodell CERA-2

- **Beschreibt Metadaten von Simulationsdaten der Klimaentwicklung**
 - Regeln vorgegeben von World Data Center for Climate (WDC-C Hamburg)
 - Ziel: Daten verschiedener Institute einheitlich beschreiben & verknüpfen > archivieren > erschließen (CERA = Climate & Environmental Retrieval and Archive)

- CERA-2 MD: **Beschreibung von Experiment & Datensatz:**
 - **Projekt:** ForscherInnen, Zeitraum / **Variablen**, Codierung / **Datenstruktur**, -format
 - **Standardnamen für Variablen:** Datenvergleich (Luftdruck) (+ Raum/Zeit Angaben) > **CF Konventionen (Climate and Forecast Metadata Convention)**
 - **Datensatzbeschreibung:** Skalen; Codelisten; Datenerzeugung, Datenqualität
 - **Langzeitarchivierung:** Persistent Identifier, Errata oder neue Datenversionen

2. Dokumentation von Forschungsdaten

(2) Disziplinspezifische bzw. datenspezifische Metadaten-Standards - Beispiele

Geowissenschaften / Klimaforschung

- **Basis: ISO 19115:2003 Geographic Information - Metadata** (400 Items)
 - Grundlage beim Aufbau Geodateninfrastrukturen > GDI-DE > Europa > ...
(INSPIRE Initiative der EC: Informationsdienste & kompatible Geodaten – Basis: einheitliche Metadaten)
 - Relevanz auch für andere Disziplinen: Geo-Referenzierung von Daten
- Disziplinspezifische Metadaten: z. B. QuakeML (Daten Seismologie)

Medizin > Forschung

- **CDISC (Clinical Data Interchange Standards Consortium)** entwickelt Life Cycle orientiertes Metadatenformat für Daten klinischer Studien

Amtliche Statistik (statischen Ämtern):

- **SDMX (Statistical Data & Metadata Exchange) ISO/TS 17369:2005**

Geisteswissenschaften u.a.

- **Text Encoding Initiative (TEI:P5) - Kodierung & Austausch Texten**

2. Dokumentation von Forschungsdaten

(3) Klassifikationssysteme & Vokabulare

Weitere Bausteine zur standardisierten Dokumentation von (Daten) Objekten und ihren Eigenarten

Klassifikationssysteme und kontrollierte Vokabulare - Beispiele:

- ISO Normen für geografische Einheiten (ISO 3166), Sprachen (ISO 639-1:2002) oder Codes für Berufe (ISCO)
- Medizin: ICD Klassifikation der Krankheiten und verwandter Gesundheitsprobleme
SNOMED CT Systematisierte Nomenklatur der Medizin
- DDI Standard: Umfragemethodik (Art, Häufigkeit, Stichprobe der Befragung ...)

Beispiele: technische Repräsentation kontrollierter Vokabulare

- Genericode: Standardmodell & Präsentation in XML für kontrollierte Vokabulare
- SKOS (Simple Knowledge Organization System): kontr. Vokabulare Semantik Web

3. Publikation, Zitation & Recherche FD

Publikation Analyse-Ergebnissen: üblich Aufsätze, Monographien, ...

- **Metadaten für Texte:** Dublin Core, MARC (Machine-Readable Cataloguing) ...

> Veröffentlichung & Zitation Daten als eigenständige Publikationsform

- Nutzung v. **Persistent Identifier** (Speicherort unabhängig > dauerhafte Infrastruktur)
- Gewinn: langfristige Zitation (Impact) / Recherche / Sichtbarkeit: Forscher & Produkt /
- Verknüpfung Literatur & Datenpublikation:
z.B. Geowiss.-, Umweltdaten: Journal ESSD (Earth System Science Data) & PANGAEA
- PID Systeme: **Handle** (Geistesw., Psycholinguistik), **URN** (Bibliotheken), PURL, ...

DataCite: **DOI (Digital Object Identifier)** z. B: **doi:10.4232/1.10834**

- Klima, Medizin, SozWiss., Pädagogik, Biodiversität (Vergabe: TIB, ZBMED, ZBW, GESIS)
- DataCite **Metadatenstandard:** (mandatory fields für alle Datenpublikationen)
 - > optionale Felder zur Spezifikation domainspezifischer Metadaten
 - > z.B. **da|ra Metadaten** für Daten Sozialwissenschaften & Wirtschaftswissenschaften


ZA4600: ALLBUS/GGSS 2008 (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2008)


[Bibliographic Citation](#) [Content](#) [Methodology](#) [Data & Documents](#) [Errata & Versions](#) [Further Remarks](#) [Groups](#)


Dataset Number of Units: 3469

Number of Variables: 800

Analysis System(s): SPSS, Stata

Current Version 2.0.0, 2011-8-3, [doi:10.4232/1.10834](https://doi.org/10.4232/1.10834) 

[Availability](#)  A - Data and documents are released for academic research and teaching.


Download of Data and [Documents](#)  Download possible for registered users, please login or register.


You don't have to pay fees for downloads.


[Datasets](#) [Questionnaires](#) [Codebooks](#) [Other Documents](#) [DDI Documents](#)

• [ZA4600_cdb.pdf](#) (Codebook) 3 MBytes

• [ZA4600_cdb_Supplement.pdf](#) (Codebook) 446 KBytes

 [Data Access](#) to ALLBUS Studies

 You can order this study via shopping cart
[add to shopping cart](#)

 [\(General access to studies and data sets at the GESIS Data Archive for the Social Sciences\)](#)

Bsp.: Erschließung & Präsentation von DDI basierten Metadaten einer Datenpublikation „doi:10.4232/1.10834“

Oben: Kontext

Mitte: Datensatz: Version & doi

Unten: Download Daten & Dokumentation

4. Langzeitarchivierung von F-daten

(1) Modelle & Kriterienkataloge zur Etablierung & Betrieb digitaler Langzeitarchive

Hintergrund > Ziele der LZA: Speichern & Erhalt | langfristige Nutzbarkeit & Interpretierbarkeit > Verfügbarkeit

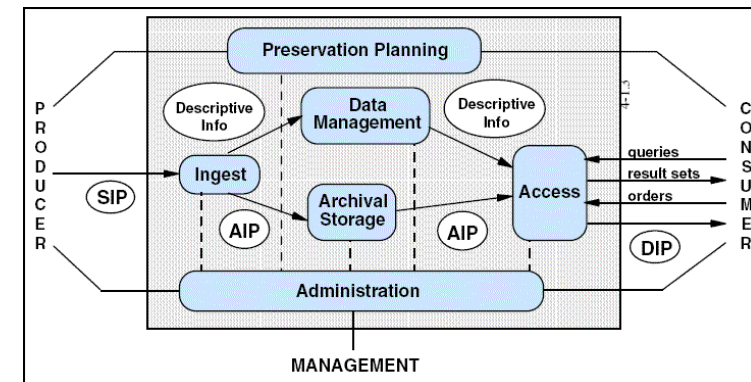
Ist laufender Entwicklungsprozess (Infrastrukturen) > Anforderungen & Qualitätssicherung > Normen, Methoden, Werkzeuge

- Leitmodell für Langzeitarchive

Open Archival Information System (OAIS)

- ISO14721:2003
- Begriffs-, Struktur- und Organisationskonzept (Technik <Organisation> Akteure);
- Keine Vorgaben für die technische & organisatorische Implementation > Anwendungskontext

Akteure – Funktionseinheiten - Objekte



- Kriterienkataloge zu Organisation & Vertrauenswürdigkeit Langzeitarchive

- TRAC (Trustworthy Repositories Audit & Certification: Criteria and Checklist)
- DRAMBORA (Digital Repository Audit Method Based on Risk Assessment)
- Data Seal of Approval
- NESTOR Kriterienkatalog / Enzyklopädie zur LZA

4. Langzeitarchivierung von F-daten

(2) LZA Metadaten (domainunspezifisch) > Interoperabilität

Beispiele für Metadatenstandards in der Langzeitarchivierung

- **PREMIS** (Preservation Metadata: Implementation Strategies)
 - **MD**: notwendige Semantische Informationen über Archivsystem & Management Bibliotheken: LMER (Long-term preservation Metadata for Electronic Resources)
- **METS** (Metadata Encoding and Transmission Standard)
 - **MD zur Dokumentation digitaler Sammlungen** (FD, Dokument, Bilder...)
 - **7 Abschnitte**: u. a. **Dateien** (file section), **Struktur** (structural map) & **administrative** Metadaten (technische, Rechte ...) einer Kollektion
 - **METS Dokumente können andere MD verknüpfen bzw. integrieren** wie z. B. Metadaten aus PREMIS, Dublin CORE, MARC ...
 - METS stellt **OAIS kompatibles Informationspaket** für den LZA Workflow bereit: Einlieferung SIP > Archivierung AIP > Bereitstellung DIP

Zusammenfassung: Welche Standards gibt es?

(1) Weitere Antwort aus Sicht einer funktionalen Einteilung

Unterscheidung nach Metadatentypen / -arten

Decken unterschiedlichen Beschreibungs-/Informationsbedarf je nach Anwendungskontext ab

- **Deskriptive:** Inhalt des Objektes (Studientitel ... Variable ... Messeinheit ...)
- **Strukturelle:** innerer Aufbau der Objektkomponenten / Beziehung zu anderer Objekten
- **Administrative:** Management des Objektes (ID, Version ... Rechte, Fristen)
 - **Rechtmanagement** Verwertungs- und Urheberachweise, Zugangsregeln ...
 - **Technische:** Herstellung des Objektes (Dateiformat ... Hardware / Software ...)
 - **Provenance:** Geschichte des Objektes: Erhaltungsmaßnahmen, Authentizität, Integrität

LZA benutzt im Prinzip alle Typen als „**Preservation Metadata**“ in ihrem Anwendungskontext

Zusammenfassung: Welche Standards gibt es? (2) Beantwortung durch Strukturierung & Visualisierung des Metadata Universe



Content: Jenn Riley – Design: Devin Becker

Quelle: <http://www.dlib.indiana.edu/~jenrile/metadatamap/>

Zum Schluss > noch mehr Antworten ...

Die Frage „Welche Standards gibt es?“ möchte ich zum Schluss an die Autoren aus 11 Disziplinen zur Beantwortung weiterreichen.

Sie antworten umfassend auf die Frage nach **Metadaten bzw. fachspezifischen Metadaten-Schemata in ihrer Disziplin:**

Sozialwissenschaften

Psycholinguistik

Pädagogik und Erziehungswissenschaft

Geisteswissenschaften

Altertumswissenschaften

Geowissenschaften

Klimaforschung

Funktionelle Biodiversität

Medizin

Teilchenphysik

Astronomie und Astrophysik

Beschrieben in Neuroth et. al. (2012) „Langzeitarchivierung von Forschungsdaten - Eine Bestandsaufnahme“

Vielen Dank für Ihr Interesse

Uwe.Jensen@gesis.org

Literatur zu Metadaten > Standards, Akronyme, Verweise

- [Glossary of Metadata Standards](#). (2010). Riley und Becker
- [Konzeptstudie Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie](#) (2010). TIB (Technische Informationsbibliothek) Hannover, FIZ Chemie Berlin & Universität Paderborn
- [Langzeitarchivierung von Forschungsdaten](#) - Eine Bestandsaufnahme. (2012). Neuroth et.al. (Hrsg)
- [Metadata Resources Guide](#) (2004). Government of Alberta
- [Metadaten und Standards](#). (2011). Jensen, Katsanidou, Zenk-Möltgen
- [nestor-Handbuch](#): Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (2010). Neuroth et.al. (Hrsg)
- [Understanding Metadata](#). National Information Standards Organization. NISO Press (2004)