



nationale
Forschungsdaten
Infrastruktur

for CULTURE

23. MusIS-Nutzertreffen – Datenaufbereitung und Datenqualität | 14.09.2023

Mechanismen des Datenqualitäts- Managements für Museen und Sammlungen

(nach Maßgabe von NFDI4Culture)

Dr. Celia Krause

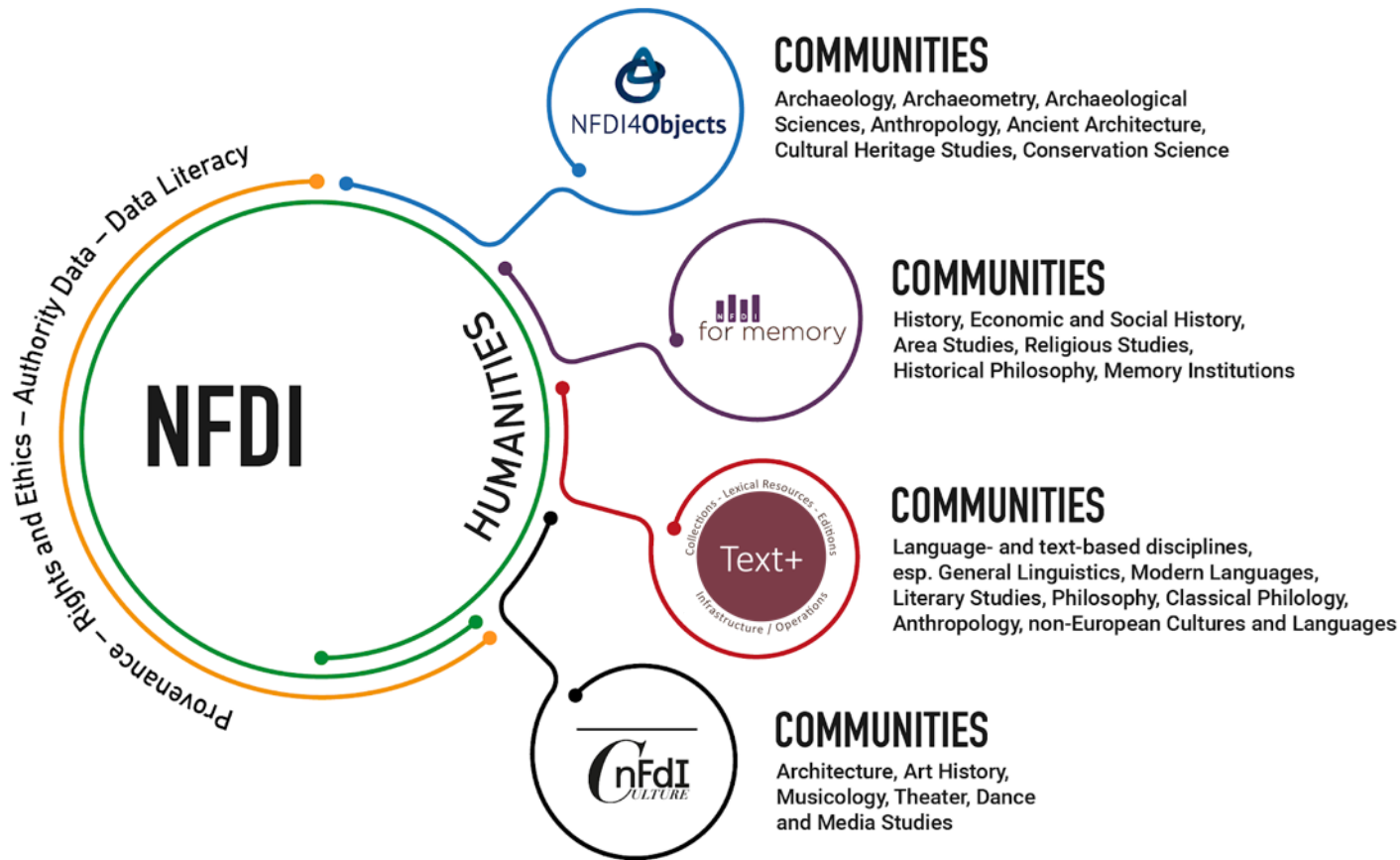
Gefördert durch die
Deutsche Forschungsgemeinschaft (DFG) - 441958017

DFG Deutsche
Forschungsgemeinschaft

1 ■

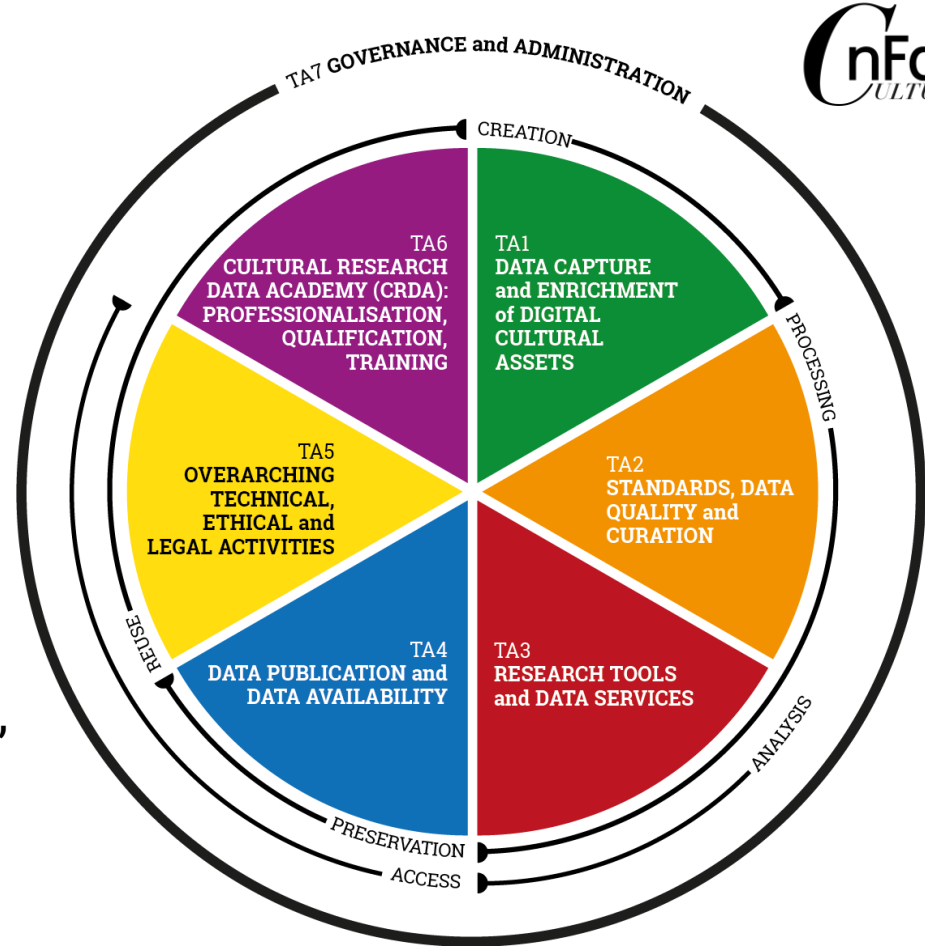
NFDI4Culture und Zielgruppen

Kooperation in den Geistes- und Kulturwissenschaften

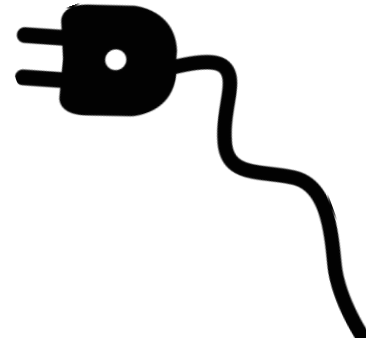
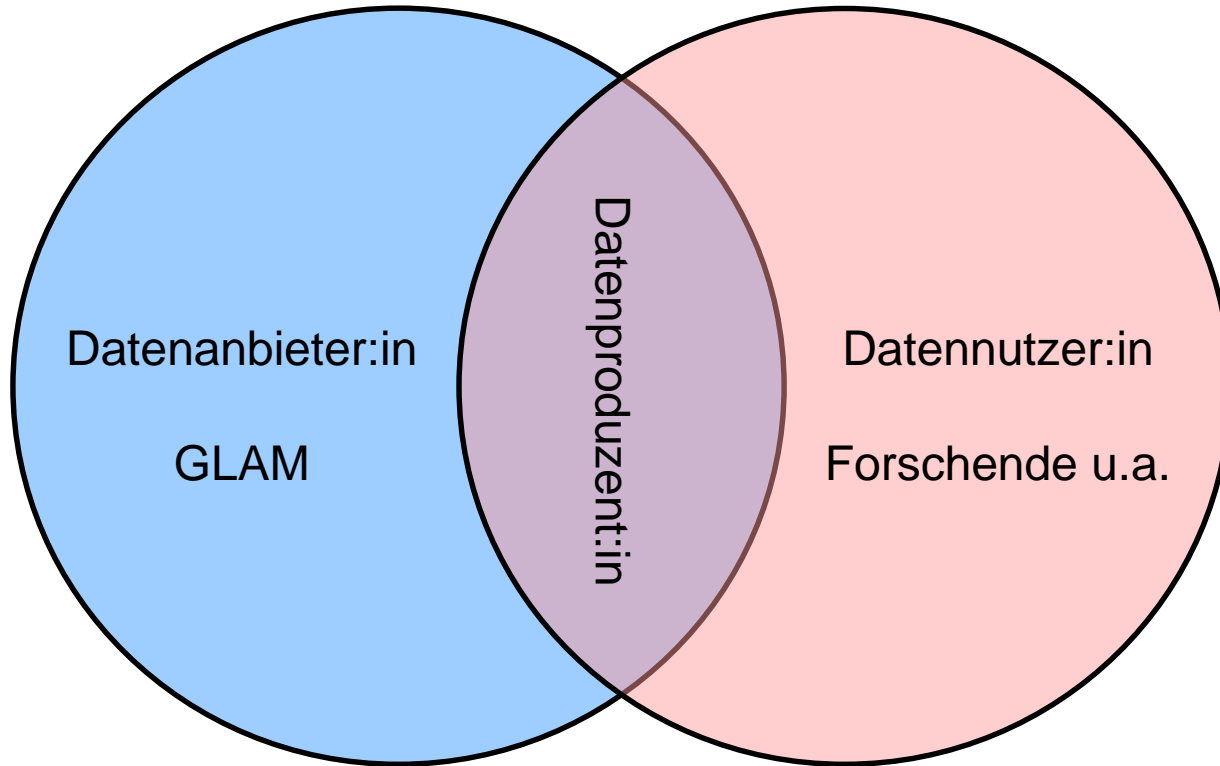


TASK AREAS in NFDI4Culture

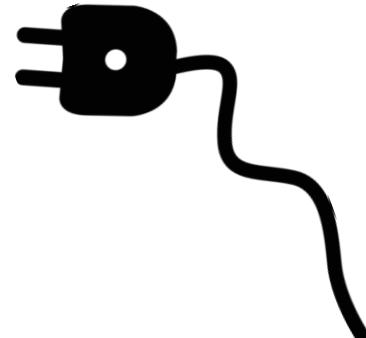
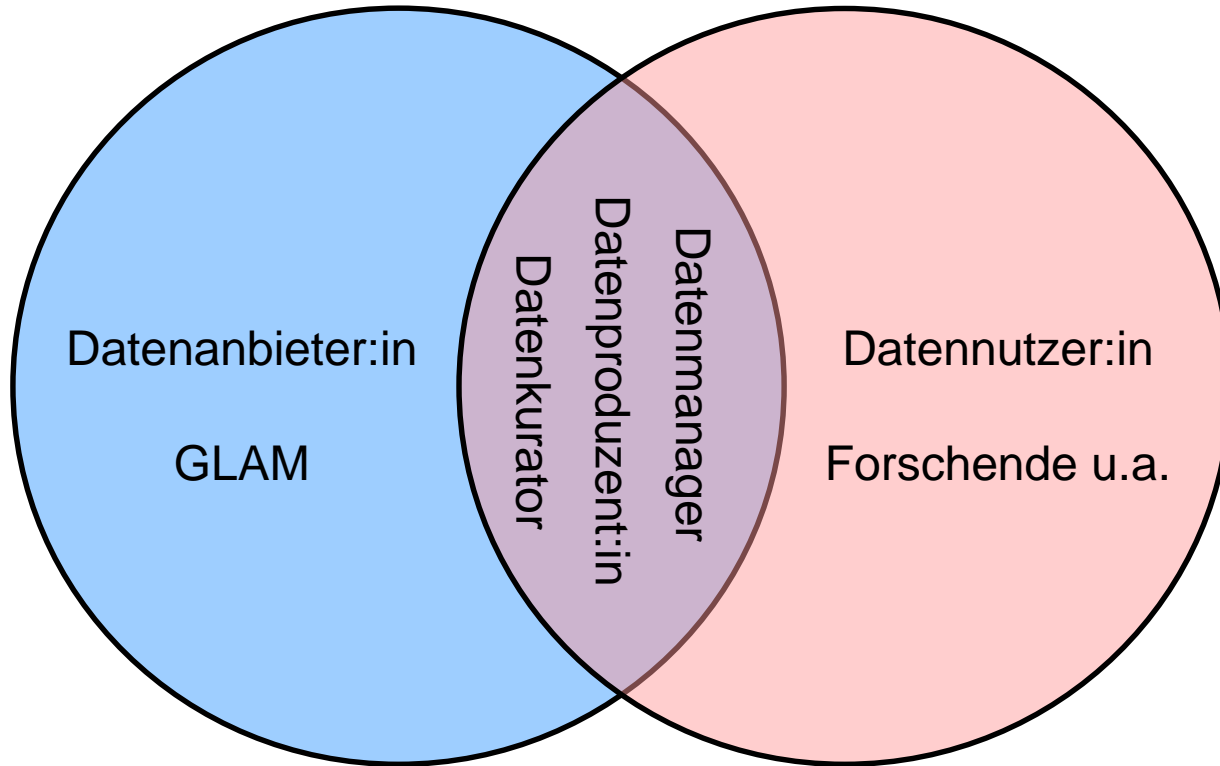
- Datenerfassung und -anreicherung (TA1)
- **Datenstandardisierung und Qualitätssicherung (TA2)**
- Forschungssoftware und Datendienste (TA3)
- Datenpublikation und -nachhaltigkeit (TA4)
- Ethische und rechtliche Fragen, übergreifende Infrastruktur (TA5)
- Qualifikations- und Trainingsangebote (TA6)



Hauptakteure in der Dateninfrastruktur






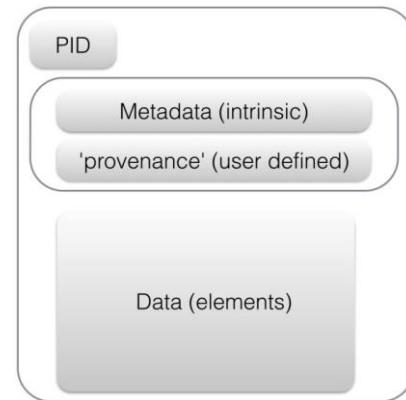
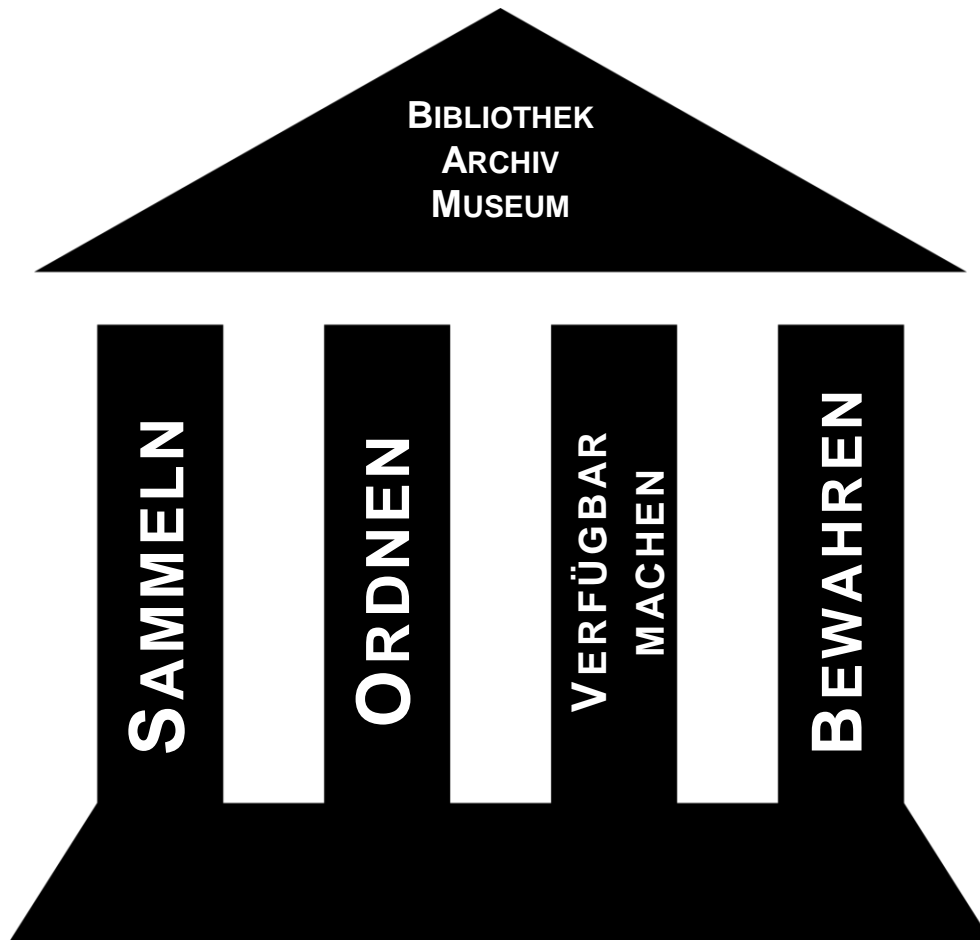
Hauptakteure in der Dateninfrastruktur





2. ■ Von Objekten und digitalen Sammlungen

- *Bibliothek* 
 - Erwerben, Erschließen, **Bereitstellen**, Bewahren
- *Archiv* 
 - Sammeln, **Verzeichnen**, **Bewahren**, Erforschen, Vermitteln
- *Museum* 
 - Sammeln, Dokumentieren, Erforschen, Bewahren, **Vermitteln und Präsentieren**



Canham, S. & Ohmann, C. (2016). Fig. 1.
(CC BY 4.0)

Collections as Data – Sammlungen als Daten



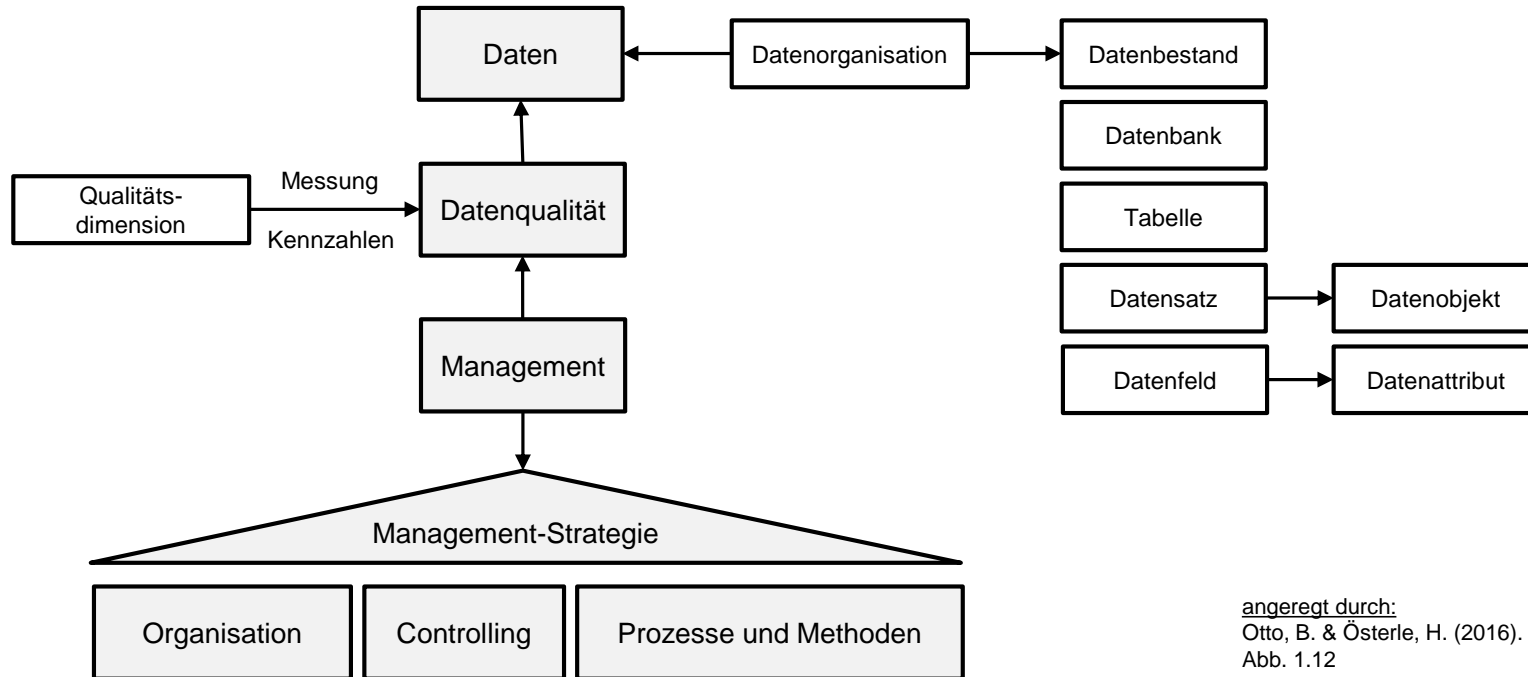
	physische Sammlung	digitale Sammlung	Sammlung als Daten
Objekt	möglichst originalgetreues Artefakt	wiedergabegetreues Surrogat (Ersatz für physisches Objekt)	Prozessierbares Objekt (Erschließung am Objekt)
<i>Verwendung</i>	Physische Ausstellung	Anschauungsmaterial, Online-Ausstellung (Websites, Portale)	Digitale Forschung, Big Data, computerbasierte Analyseverfahren
Dokumentation	Beschreibung (Karteikarten oder Inventarbücher)	Metadaten (Datenbank, Sammlungsmanagementsystem)	Metadaten (Datenbank (auch Graphdatenbank ...), Repository, Triple Store ...)
<i>Verwendung</i>	Sammlungserhaltung, „analoge“ Forschungszwecke (Provenienzfragen ...)	Arbeitsprozesse im Museum (Berichte, Leihverkehr, Ausstellungsplanung, Ö-Arbeit)	Weitergabe an Portale für die Recherche, Aggregatoren, Darstellung von Wissenskontexten (im Knowlegde Graph)
<i>Verwendung</i>		Präsentation auf der eigenen Website, ggfs. Herunterladen pro Datensatz	Basis für DH Projekte, Netzwerkanalyse, maschinelles Lernen, dump zum Download

3. ■ Digitale Strategie und Datenqualität

Datenqualität durch Management



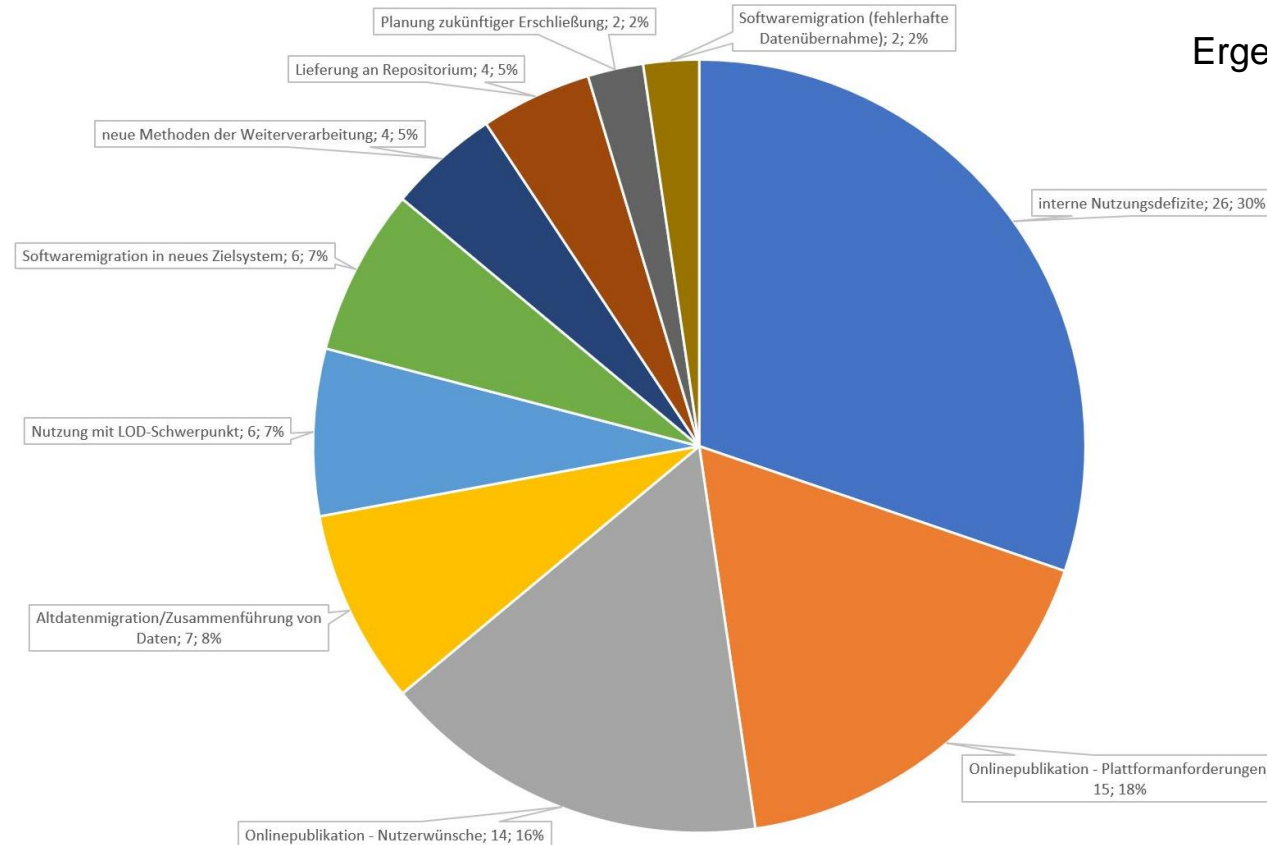
Institution	Projekt	Daten
Organisationsstruktur (Organigramm, Aufbauorganisation), Arbeitsbereiche, Zuständigkeiten	Planung, Durchführung	Organisation, Strukturierung Vorgehen im Datenlebenszyklus
<ul style="list-style-type: none">• <i>Bestandsaufbau</i> = Sammeln und Dokumentieren	Konzept	Einheiten und Zusammenhänge
<ul style="list-style-type: none">• <i>Bestandserhaltung</i> = Bewahren	Zeitplan	Datenmodelle
<ul style="list-style-type: none">• Erforschen	Typische Arbeitsschritte und Aufgabenverteilung	Datenfeldkatalog
<ul style="list-style-type: none">• <i>Bestandsvermittlung</i>	Prozesse	Policy, Richtlinie
Sammelprofil , Sammlungskonzept	Datenmanagementplan	Datenmanagementplan
Prozesse (Ablauforganisation)	Qualitätssicherung und -verbesserung	Qualitätssicherung und -verbesserung



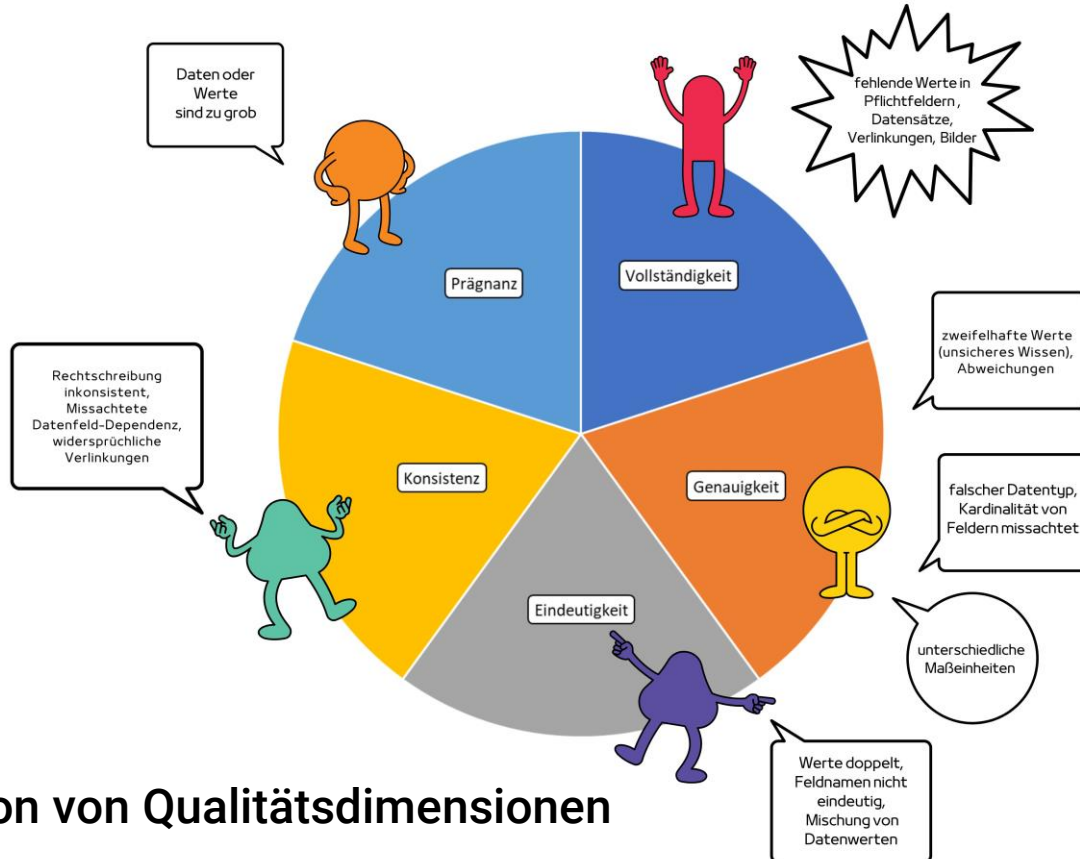
angeregt durch:
Otto, B. & Österle, H. (2016).
Abb. 1.12

Daten: Wo treten Qualitätsprobleme auf?

Ergebnisse einer Umfrage

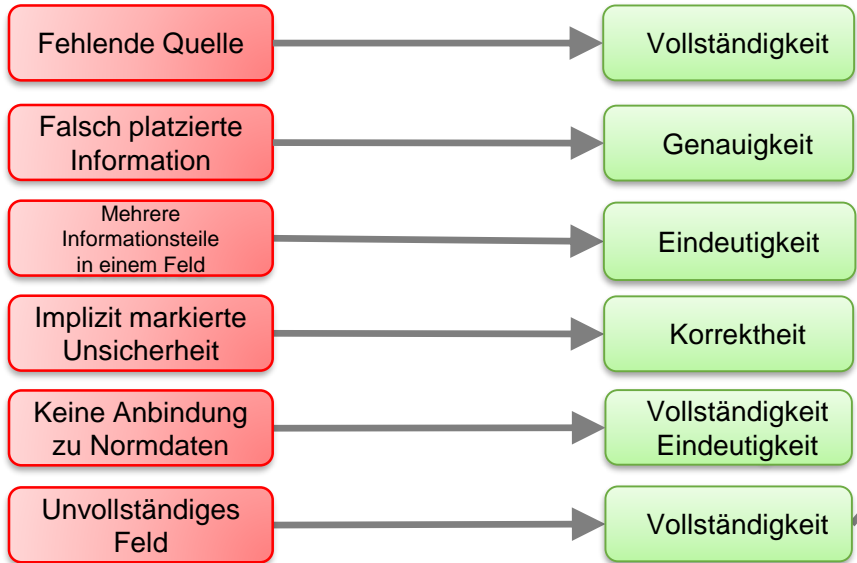


Daten: Vorgehen bei der Problembekämpfung



Definition von Qualitätsdimensionen

Daten: Vorgehen bei der Problembhebung



DATA01.1.2 Lack of data - incomplete fields

Description: Fields are only partially filled. Not all desired information is stated in the model. Fields can be incomplete due to a lack of knowledge (only partial information) or formal mistakes (lack of units). Hence there is only sometimes a possibility to complete those fields

Mainly affected quality feature: internal completeness (expected values are incomplete for attributes and/or elements)

Other affected quality features:

- semantic accuracy (incompleteness violates correctly represented information of the domain of interest)
- syntactic accuracy (if the corresponding specification for the data fields exists, the syntactic specification might be violated)
- precision (fields logically are, in most cases, not exactly as expected if the data field is partially filled)
- appropriateness (partly filled fields may affect the understandability of related elements)
- compliance (if the corresponding specification for the data fields exists, standards, conventions, or regulations will be violated)
- external completeness (if the information is intended to be given as a reference)

Impact on data quality: Information about the described entity is missing from the record.

Causes:

- Lack of research knowledge
- Implicit encoding of uncertainty (e.g., imprecision)
- Human error: Work has been interrupted and not been completed

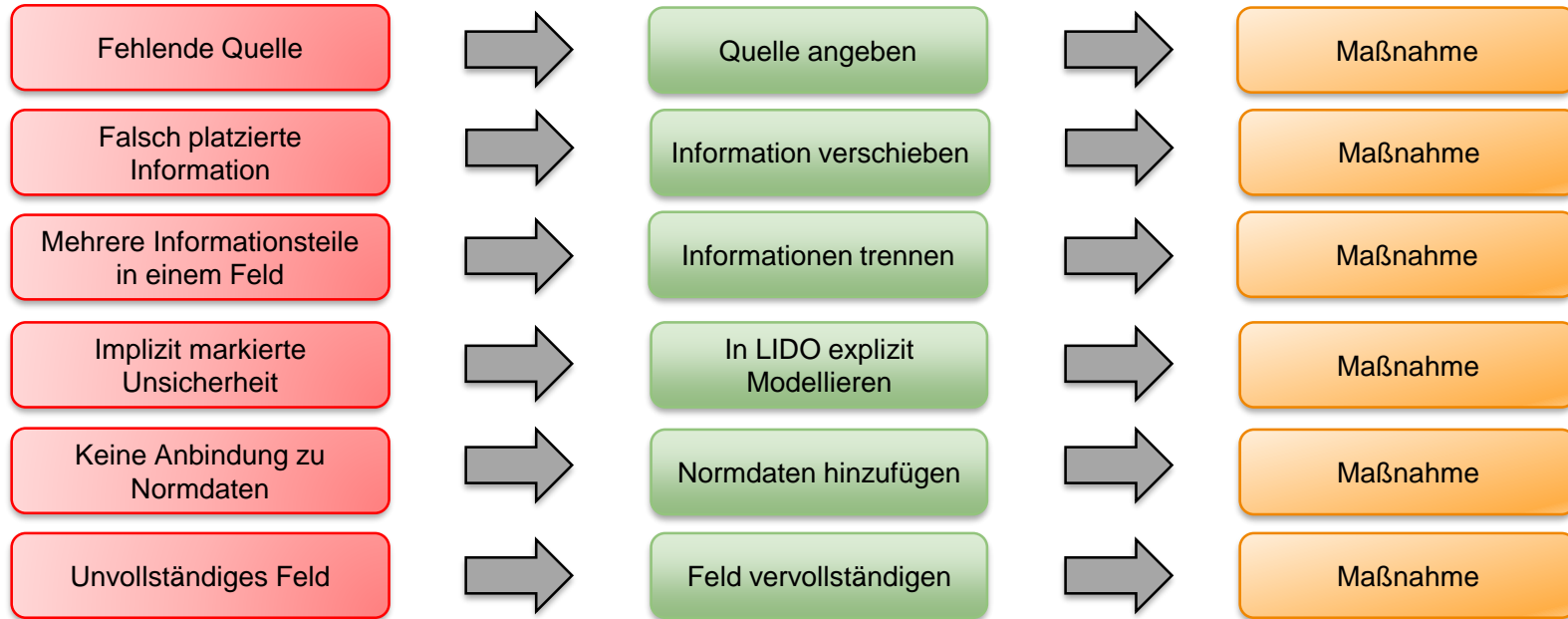
Root in the data life cycle: Collect

Identification: Fields do not contain all the desired information.

Target state: All fields meet their requirements on information density.

Benennung von Fehlern, Zuordnung zu Qualitätsdimensionen,
Beschreibung der Qualitätsprobleme

Daten: Vorgehen bei der Problembhebung



Formulierung von Anforderungen für gute Datenqualität,
Ableitung von Maßnahmen

4. ■

Einige Grundregeln für gute Datenqualität



Leitideen



Normierung und Standardisierung:

CIDOC CRM (ISO 21127:2014), ISO-Norm 800
„Datenqualität und Stammdatenqualität“, LIDO,
Normdaten, Vokabulare/Thesauri
(ISO 25964, DIN 1463)

Leitlinien/Policies: Forschungsdaten-Policies, Fair
Data Principles, Care Principles, Spectrum

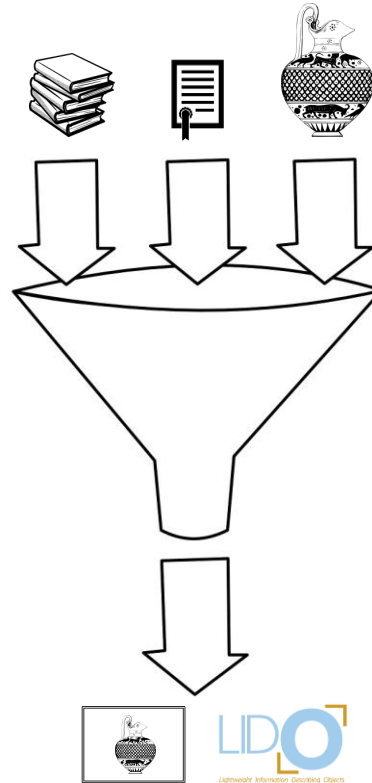
idealtypische Prozesse/Verfahren:

Datenlebenszyklus,
Workflows: OAIS (ISO 14721)

Validierung/Operationalisierung:

Datenmanagementpläne, Check-Verfahren
(Qualitätskontrolle)

Arbeitsgänge vom Publizieren her denken



1. Objekte in (Sinn-)Einheiten zerlegen,

z.B. *Textträger*, Text(block), Zeile, Wort, Schriftzeichen/Buchstabe

z.B. *Bildträger*, dargestellte Objekte/Personen, Teile davon ...

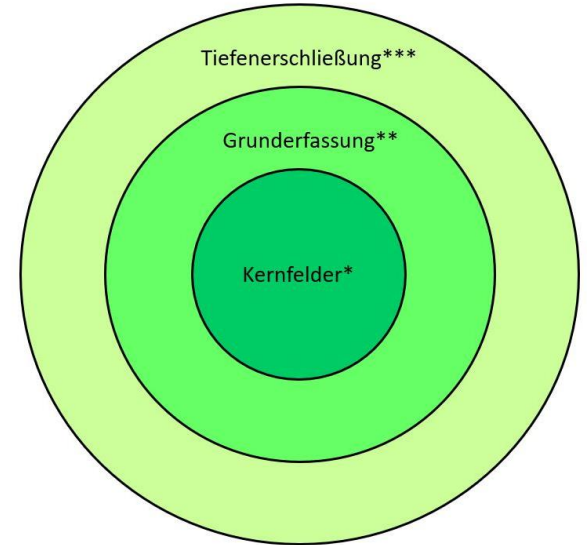
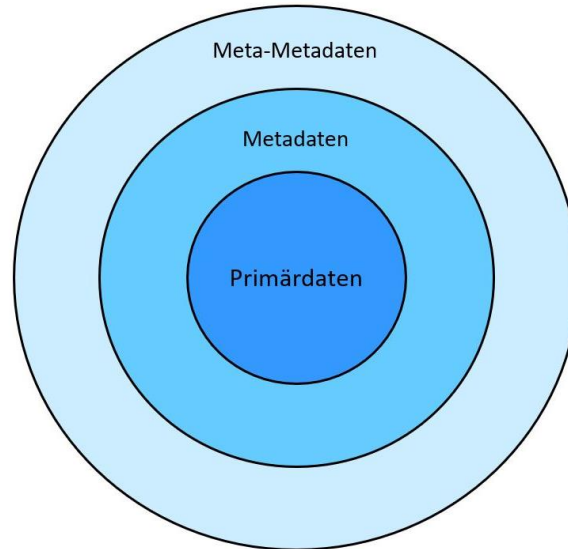
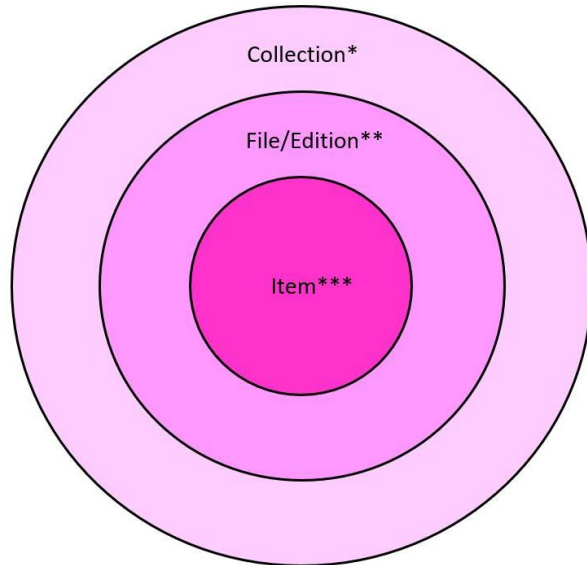
→ Trennen zwischen Träger und dargestelltem Gegenstand

2. Entitäten festlegen und als *eigene Informationseinheiten* behandeln,

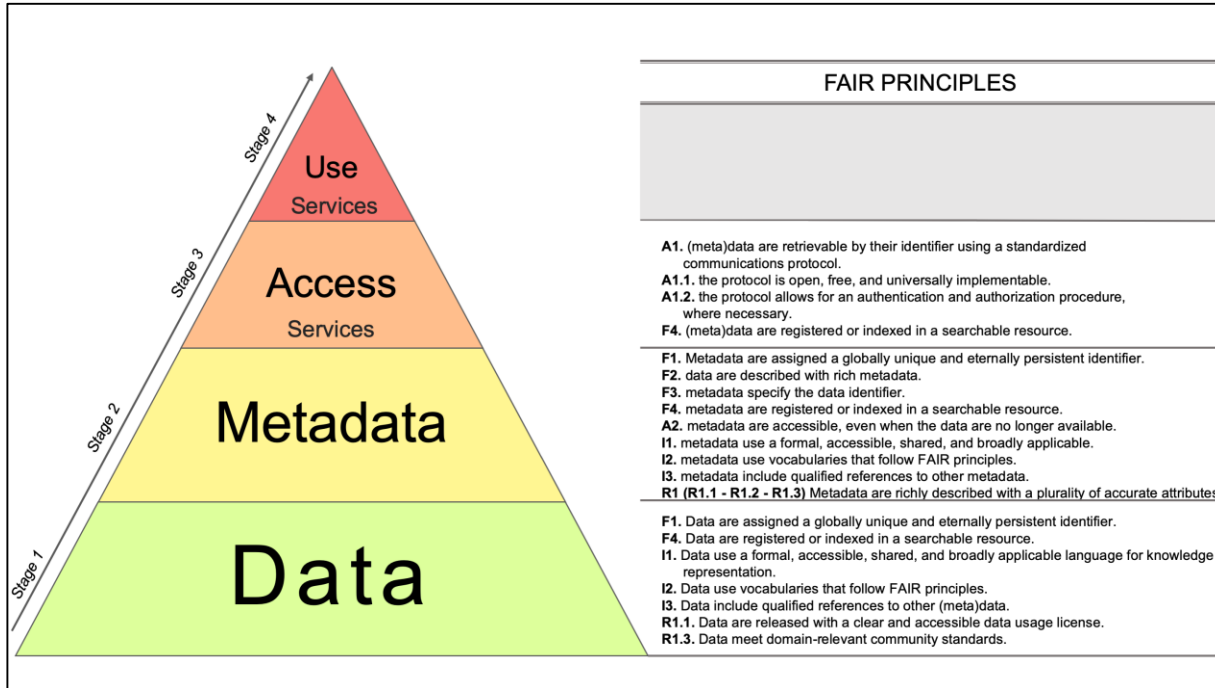
z.B. als Tabelle bei der Datenbankentwicklung

3. Für unterschiedliche Informationen *getrennte Datenfelder* anlegen

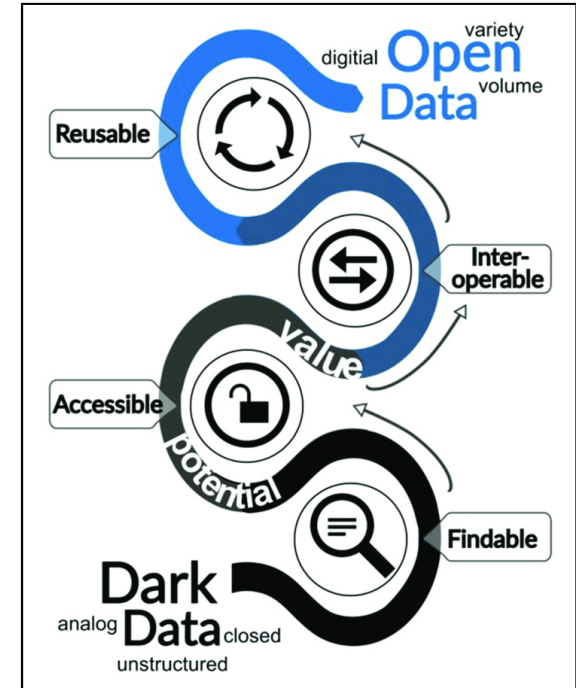
Denke in Schichten (Beispiele)



Qualitätsverbesserung ist ein gradueller Prozess



D. Bailo et al. (2019)



Easterday, K. et al. (2018) Fig. 1 (CC BY 4.0)

Let's talk #NFDIrocks !!

contact@nfdi4culture.de

Bailo, D. et al. (2019). Four-stages FAIR Roadmap - FAIR "Pyramid".
<https://doi.org/10.5281/zenodo.3299353>

Canham, S. & Ohmann, C. (2016). A metadata schema for data objects in clinical research. *Trials*. 17. <https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-016-1686-5>
Für die Grafik vgl. auch: [Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0](#)

Easterday, K. et al. (2018). From the Field to the Cloud: A Review of Three Approaches to Sharing Historical Data From Field Stations Using Principles From Data Science. *Frontiers in Environmental Science*. 6. <http://dx.doi.org/10.3389/fenvs.2018.00088>

Otto, B. & Österle, H. (2016). Datenqualität – eine Managementaufgabe. In: *Corporate Data Quality*. Berlin – Heidelberg. https://doi.org/10.1007/978-3-662-46806-7_1