

Forschungsdatenrepositorien, -datenbanken & -langzeitarchive

Eine Abgrenzung

Juliane Watson

Begriff,
Hauptfunktion,
Granularität,
Permanenz

Forschungsdatenbank

(heute)

- Ist für **Nutzung** und Suche der Forschenden ausgelegt
- Granularität ist auf Ebene einer **Entität** und
- Darstellung und Präsentation eines Datensatzes
- Es handelt sich um ein **dynamisches** System
- Datensätze sind selten **versioniert**
- Daten innerhalb eines Systems häufig eher **homogen** (z.B. Bild mit Text und Verknüpfungen)



Repository

(digitales)

- Ausgelegt auf **Publikation** und **Referenz**
- eines “datasets” (Gruppe) - Granularität
- Es handelt sich um ein **wachsendes** System
- datasets teilweise **versioniert**
- datasets Innerhalb eines Systems häufig eher **heterogen** (Publikation einer Tabelle v.s. kompletter Kataloge, Forschung von mehreren Jahren etc. Granularität sehr unterschiedlich)

Langzeit- archiv

(nobody knows)

- Ausgerichtet auf **Erhaltung**
- einer **Datei**, wobei diese unterschiedlich viele DS z.B. enthalten kann bzw. auch viele Dateien einen DS ausmachen
- Langzeitarchiv ist **statisch**, es gibt im Normalfall maximal mehrere Versionen von **Metadaten**
- Daten eher heterogen aber häufig Sammlungsstrategie (fachlich, technisch...)
- Unterschiedliche Preservation level: Erhaltung des Bitstream, logische & semantische E.

Datenmodelle &- repräsentation

Unterschiede:
Repräsentation

- **FDB:**
 - Entitäten mit stabilen IDs
 - Dynamische Präsentation in Interface (Web)
 - Url oft logisch strukturiert, selten versioniert
- **Repositorium**
 - Dataset mit Landing page
 - Einheitliche ID / Versionierung
 - Weniger granular aber zitierfähig
- **Langzeitarchiv**
 - Webpräsentation nicht notwendig (aber häufig, was zu Verwirrung mit Repo führt)

Datenmodelle &- repräsentation

Unterschiede:
Datenmodell

- **FDB:**
 - Datenmodell einheitlich
 - Häufig gewachsene Systeme
 - sehr spezifisch und selten wiederverwendbar
- **Repository**
 - Datenmodell nicht einheitlich
- **Langzeitarchiv**
 - Datenmodell nicht einheitlich, aber dokumentiert (präzise), inklusive Vokabular (abh. von Preservation level)

Metadaten

Unterschiede

- **FDB:**
 - Fachspezifische Beschreibungen (heterogen)
 - Einzelne akteure häufig via PID oder relationalen Tabellen, manchmal kein Change-tracking, manchmal keine Autor:in
 - Präsentation des Datensatzes kann sich ändern
 - Mapping auf Minimal-Standards (DC) häufig möglich

Metadaten

Unterschiede

- **Repository**
 - Häufig standardisierte Metadaten (DataCite)
Maschinenlesbar
 - Pflichtfelder (Autor, Titel, Jahr...)
 - Minimum wird ausgefüllt
 - Leere Felder
 - da zur Präsentation von Ergebnissen
verwendet viel Prosa ++

Metadaten

Unterschiede

- **Langzeitarchiv**
 - Viele Standard-Metadaten notwendig, aber ausgerichtet auf Erhaltung (PREMIS)
 - Fokus auf technischer Provenienz & Risikoabwägung
 - Metadaten durch Kontrolle etwas vollständiger als bei Repo (abh. vom preservation level)
 - Zusätzlich: AIP

Änderungen

Was ändert sich?

- **FDB:**
 - Datensatz (Versionierung meist implizit)
 - manchmal Change-History, insgesamt selten
- **Repository**
 - Versionierung DOI (Concept + Version)
 - Auf Ebene des gesamten Datasets, nicht auf Basis von einzelner Datei geplant (Change-History Best practice)
 - Zitation neue Version
- **Langzeitarchiv**
 - Technisch: neue Version mögl, aber nicht geplant (Checksum: keine Änderung)



Zitation

- **FDB:**
 - Selten formal festgelegt
 - ID oder url/PID
- **Repository**
 - Häufig: How to cite /DOI
- **Langzeitarchiv**
 - DOI Auflösung bei DIP aber häufig Zitation
kein Bestandteil des “Auftrags”
 - Wichtig ist Referenz (Daten auf welchem
Band)



spezifisch

allgemein

Forschungsdatenbank

Repositorium

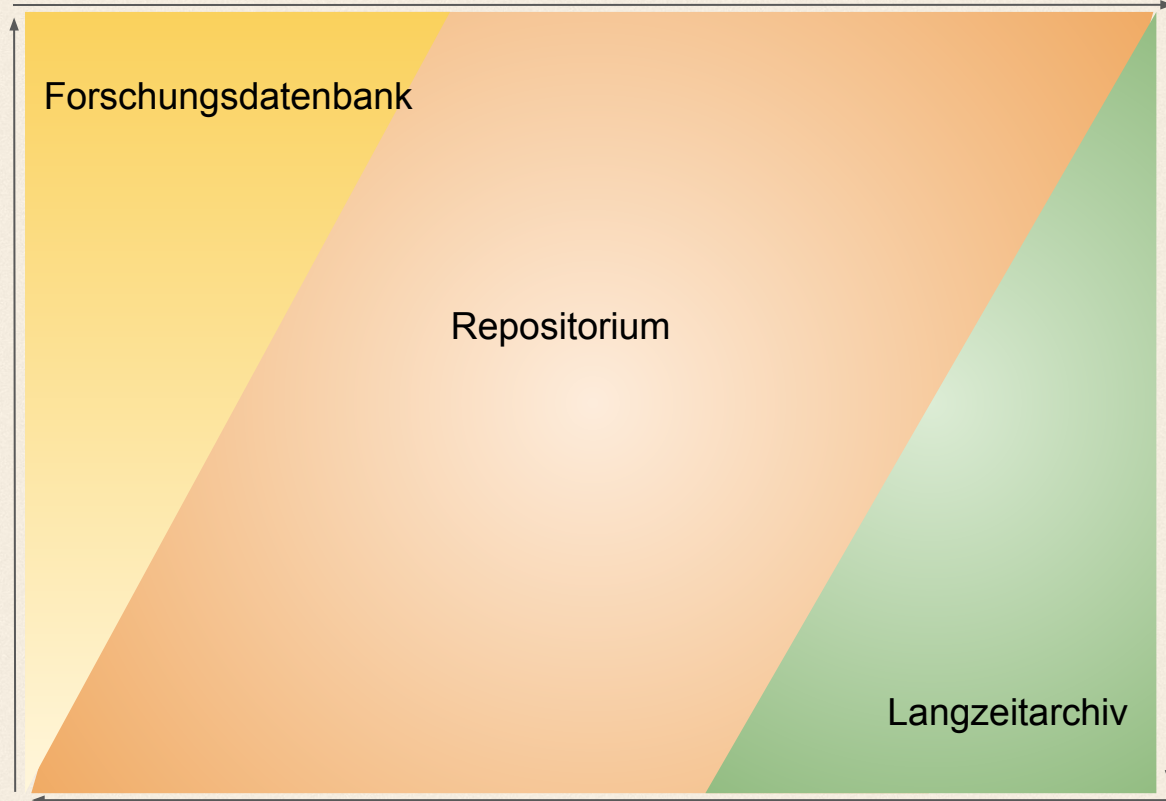
Langzeitarchiv

Einfach:
Zugriff & Retrieval

Notwendigkeit
Abgrenzung
Änderung und
einzelner
Versionen

dynamisch

statisch





Fragen?
Anregungen? Kritik?