Forschungsdatenmanagement - Langzeitverfügbarkeit - NFDI : Was macht der Verbund?

21. BVB-Verbundkonferenz | 8. November 2022

Heidi Weig (UB Regensburg)
Robert Günther (UB Bayreuth)

Projekt „Digitale Langzeitverfügbarkeit im BVB“

1. Forschungsdatenmanagement und Langzeitverfügbarkeit

Was ist Langzeitverfügbarkeit?

- Archivierung und Verfügbarhaltung von Daten und Metadaten für mind. 10 Jahre, bis zu unbegrenzter Zeit
- Bitstream Preservation – Logical Preservation – Semantic Preservation

FDM und Digitale LZV – Wozu?

- **Effizienz**: Auffindbarkeit und Zugänglichkeit im Sinne der *FAIR*-Prinzipien und *Open Science*
- **Replikation, Qualitätssicherung, Validierung**: breite Nachnutzbarkeit von Daten und Forschungsergebnissen, Sicherung gegen Manipulation von Information und Wissen
- **Erhalt** des wissenschaftlichen und kulturellen Erbes, und der Überprüfbarkeit von Narrativen: gegen ein „digital dark age“

2. Forschungsdaten und das Langzeitarchiv der ZDI

Projekt „Digitale Langzeitverfügbarkeit im BVB“

Erkenntnisse des Berichts *Perspektiven der digitalen Langzeitarchivierung in Bayern*, 2017:

Langzeitverfügbarkeit …

- geht über die kurz- und mittelfristige Bereitstellung von Publikationen und Daten in Repositorien hinaus.
- kann nicht auf objektklassen- und bedarfsübergreifende Patentlösungen zurückgreifen.
- ist ein Prozess der des kontinuierlichen Einsatzes technischer und personeller Kapazitäten, sowie technischer, organisatorischer, rechtlicher, und wissenschaftlicher Expertise bedarf.
- kann nur durch Kooperation mehrerer Institutionen gewährleistet werden.

2. Forschungsdaten und das Langzeitarchiv der ZDI

Projekt „Digitale Langzeitverfügbarkeit im BVB“

Konzeption einer LZV-Infrastruktur—

- gemäß dem **Open Archival Information System** (OAIS, ISO 14721):

  An OAIS is an Archive […] consisting of an organization […] of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community.

- basierend auf einer **zentral-dezentralen Verteilung** von Verantwortlichkeit, Leistungen und Kompetenzen.

- mit Softwarelösung **Rosetta** (Ex Libris): Lizenznehmer Verbundzentrale, Betrieb durch die VZ und das Leibniz Rechenzentrum.

2. Forschungsdaten und das Langzeitarchiv der ZDI

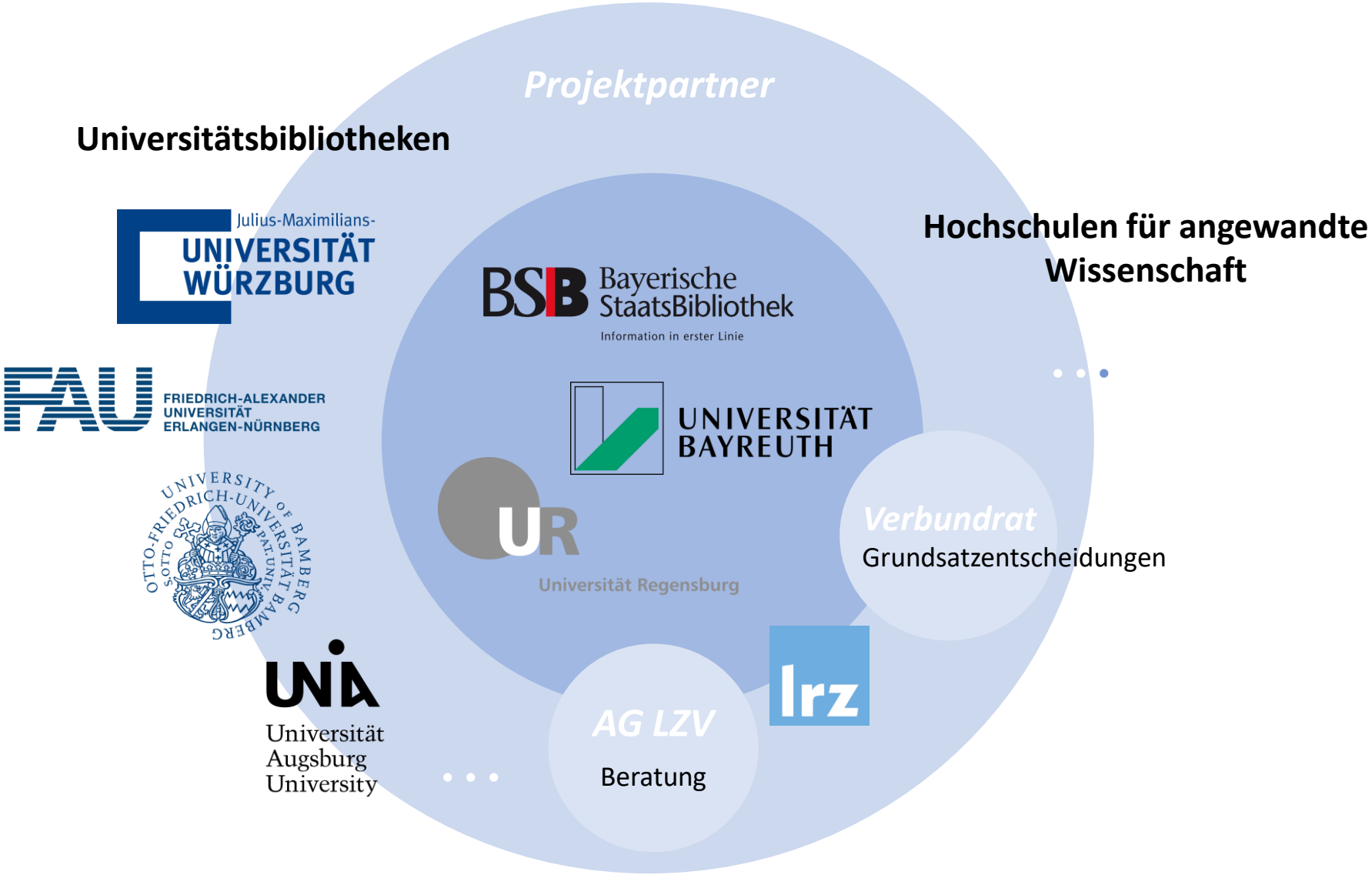
Projekt „Digitale Langzeitverfügbarkeit im BVB“

Projektpartner: UB Regensburg (Projektleitung), UB Bayreuth, BSB

Phase 1: 2019-2021 — Aufbau, Modelllösungen, Policies

- Aufbau der zentralen Infrastruktur in der Verbundzentrale
- Modell-Workflows für Digitalisate, Open Access-Publikationen, Forschungsdaten
- Entwicklung von Preservation Policies, Archivierungsvereinbarung, und Kostenmodell
- Vernetzung mit Nachnutzungs-Partner und komplementären Projekten

Projekt „Digitale Langzeitverfügbarkeit im BVB“ – Organisation



2. Forschungsdaten und das Langzeitarchiv der ZDI

Projekt „Digitale Langzeitverfügbarkeit im BVB“

Phase 2: 2022-2025 — Konsolidierung und Nachnutzung

- Finalisierung von Workflows und Dokumenten
- Öffentlichkeitsarbeit
- Zertifizierung durch Core Trust Seal
- Weiterführung und Ausweitung von Kooperationen

2. Forschungsdaten und das Langzeitarchiv der ZDI

Forschungsdatenmanagement: Was machen Universitätsbibliotheken?

- Mitwirkung bei institutionellen **Forschungsdatenpolicies**, z.B. auch im Bezug auf verwaiste Daten
- Beratung bei rechtlichen Aspekten: **Datenschutz, Urheber- und Verwertungsrechte**
- **Forschungsunterstützung** (Beratung, Datenmanagementpläne, Bereitstellung von Softwarelösungen etc.)
- Aufbau **personeller Strukturen** an Institutionen: Forschungsdatenreferent /inn/en, Data Stewards / Curators
- Einrichtung von **Forschungsdatenzentren**

2. Forschungsdaten und das Langzeitarchiv der ZDI

Langzeitverfügbarkeit von Forschungsdaten

Probleme:

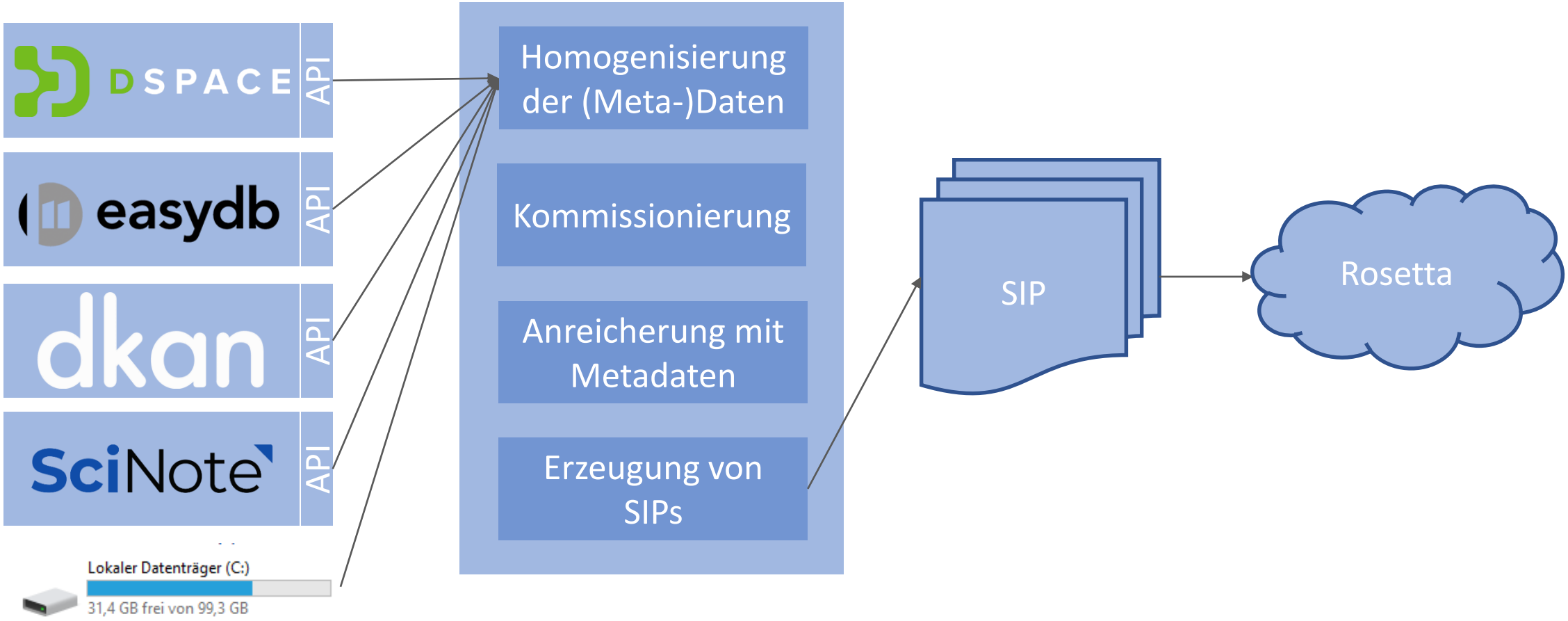
- Heterogenität der Datenformate
- Heterogenität der Speicherorte
- Heterogenität der vorhandenen Metadaten

➢ **Idee:**
Homogenisierung mittels eines niederschwelligen webbasierten Pre-Ingest Tools.

➢ **Pre-Ingest Tool „FDOrganizer“**

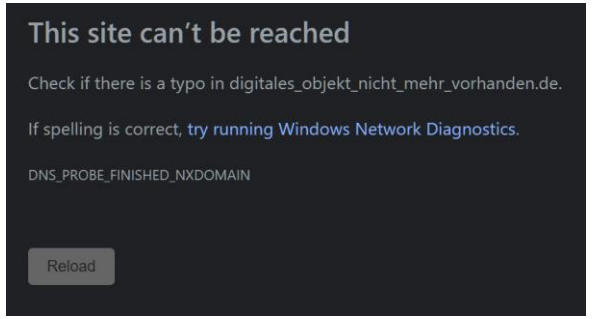
2. Forschungsdaten und das Langzeitarchiv der ZDI

FDOrganizer



2. LaVe - Aufbau eines lokalen PID-Resolver

Digitale Objekte können nicht mehr zugänglich sein aus diversen Gründen,
z.B. Link Rot, Content Drift, aktive oder unbewusste Löschung



→ **Lösung: PID-Resolver**

Zentrale PID-Resolver (z.B. DOIs, URNs) unterstützen nicht:

- größere Mengen an digitalen Objekten
- Referenzierung dynamischer digitaler Objekte
- digitale Objekte, die nur einem beschränkten Personenkreis zugänglich gemacht werden sollen
- digitale Objekte, die zur Weiterverarbeitung gedacht sind, aber nicht langfristig gespeichert werden sollen

- Des Weiteren stellen sie einen *Single Point of Failure* dar.

→ **Lösung: Lokale PID-Resolver**

IDENTIFIER

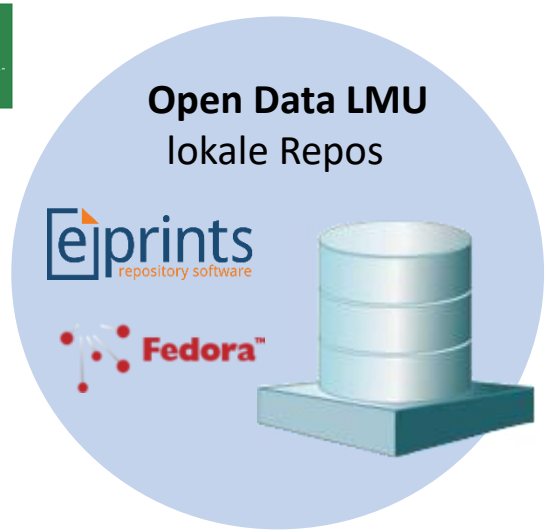
DOI: https://doi.org/10.48543/bbc1c88e-3af3-4af4-8468-c684e01e8a87

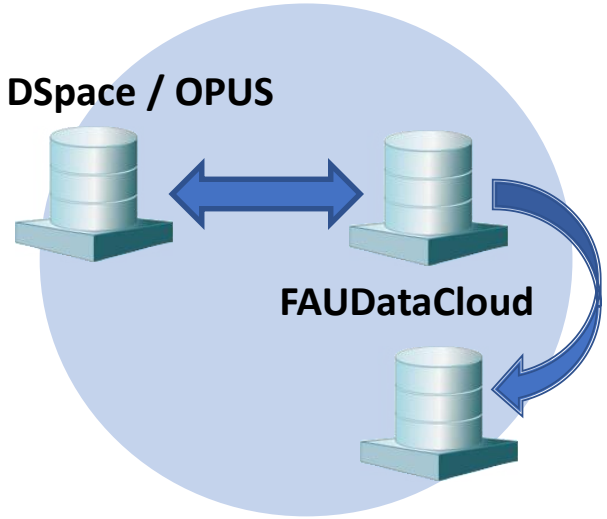
ImUB: bbc1c88e-3af3-4af4-8468-c684e01e8a87

mit Dank an Dr. Jürgen Rohrwild, FAU / LaVe

3. LaVe – Vorbereitung von Daten für externe Ingests am Beispiel Rosetta

Forschungsdaten und damit verwandte Informationen lagern an Hochschulen in unterschiedlichen Systemen





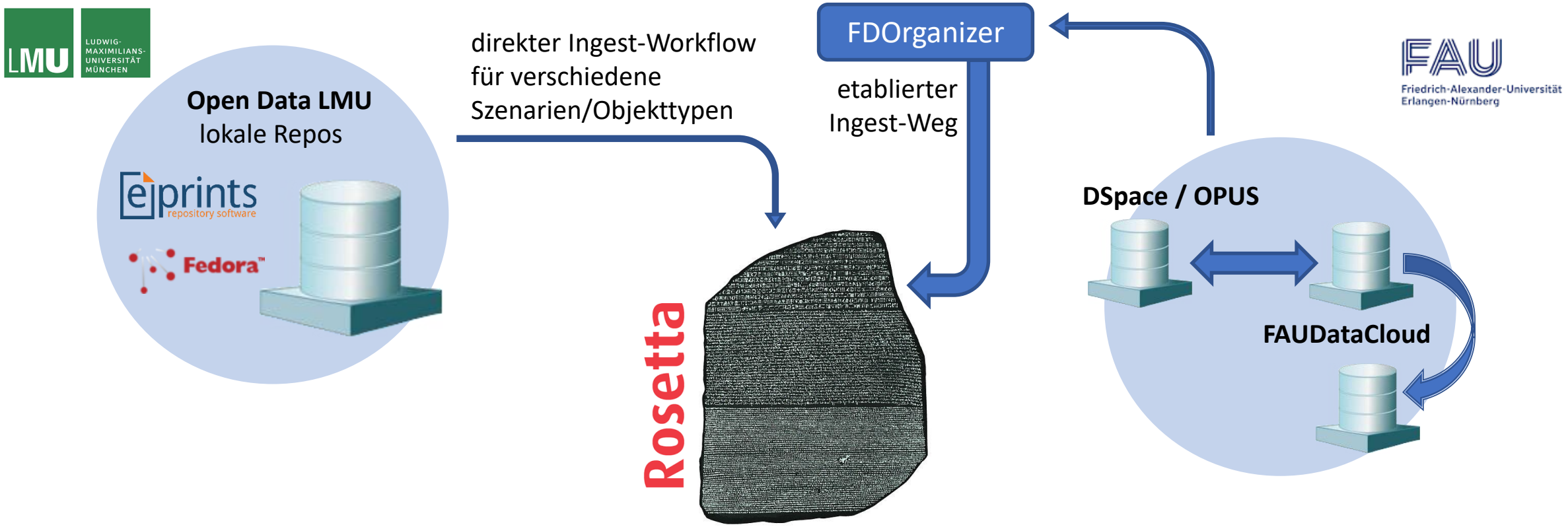
- zwei Softwarelösungen im Einsatz an der UB
- Daten und Metadaten liegen ***gebündelt*** vor
- diverse Daten (Datenstruktur, Objekttypen, …)

- verschiedene Systeme in mehreren Einrichtungen
- Metadaten und Daten liegen ***nicht immer gebündelt*** vor
- diverse Daten (Datenstruktur, Objekttypen, …)

mit Dank an Dr. Jürgen Rohrwild, FAU / LaVe

3. LaVe – Vorbereitung von Daten für externe Ingests am Beispiel Rosetta

Weitergabe an externe Systeme (LZ-Archiv, etc.) möglichst über einen zentralen Dienst an der Einrichtung



mit Dank an Dr. Jürgen Rohrwild, FAU / LaVe

4. NFDI4Memory

Nationale Ebene

BSB Co-Applicant in NFDI4Memory

- Task Area 2 Data Connectivity

NESTOR

DFG-Projekt OstData (Forschungsdaten Osteuropa) der BSB (FID)

Regionale Ebene

Projekt „Digitale Langzeitverfügbarkeit im Bibliotheksverbund Bayern“

Lokale Ebene

Rosetta DPS von ExLibris (BVB-weit nutzbar)

- Datenbestand 1.150 TB
- 380 Mill. Images
- Infrastrukturpartner: LRZ München

mit Dank an Dr. Klaus Ceynowa

5. NFDI4Memory und LZV im BVB – Blick in die Zukunft

BSB:
- born-digital Objekte
- Retrodigitalisate

BVB (10 UBs, 17 HAWs):
- Retrodigitalisate
- OA-Materialien
- Forschungsdaten

NFDI4Memory:
Forschungsdaten der historisch arbeitenden Geisteswissenschaften

Rosetta: Mandanten-fähige Multiserver-Installation

Storage: NAS, Archivierung auf Tape

Partner-Bibliotheken haben Systemzugriff auf ihre Einrichtung

BVB-Verbundzentrale: Betreuung des Gesamtsystems; Unterstützung und Erfahrungsaustausch

LRZ: Betrieb der Server- und Speicherinfrastruktur

Delivery je nach Workflow und Rechten:

dark archive oder *light archive*

mit Dank an Dr. Klaus Ceynowa

6. Ausblick

- **Ausweitung**: Integration weiterer Institutionen in die ZDI im Sinne der IT-Strategie der bayerischen Hochschulen sowie des „Kooperativen Leistungsverbundes“.

- **Verstetigung**: Personelle und finanzielle Absicherung zur Gewährleistung einer dauerhaft operierenden, effektiven Forschungsdaten- und LZV-Infrastruktur.

- **Vertiefung**: Intensivere Kooperationen mit parallelen, Rosetta-basierten LZV-Projekten zur Nutzung von Synergien.

➢ *Absicherung einer **Kernaufgabe** bayerischer Bibliotheken der Zukunft.*

7. Weiterführende Links

- *OAIS*: http://www.oais.info/
- *Rosetta*: https://knowledge.exlibrisgroup.com/Rosetta
- *nestor*: https://www.langzeitarchivierung.de/Webs/nestor/DE/Home/home_node.html
- *NFDI* auf der DFG-Website: https://www.dfg.de/en/research_funding/programmes/nfdi/index.html; *NFDI* Website: https://www.nfdi.de
- *LaVe* und *eHumanities-interdisziplinär* über: https://www.fdm-bayern.org/projekte/
- *NFDI4Memory*: https://4memory.de/
- *OstData*: https://www.ostdata.de/ostdata-client/