



Wege ins Langzeitarchiv: Workflows für Digitalisate und Open-Access-Journals für den Erhalt mit Rosetta

Bibliocon 2024 | 05. Juni 2024

Karl-Arnold Bodarwé, UB Regensburg

Michael Huber, BSB/BVB



LZV im BVB - Das Projekt

- Aufbau einer zentral-dezentralen Infrastruktur für die Langzeitverfügbarkeit von verschiedenen Objektklassen
- Projektlaufzeit: 2019 - 2025
- Projektziel: LZV-Angebot an alle bayerischen Hochschulen und ihre Bibliotheken
- Weitere Informationen unter lzv-bayern.de



Rosetta - How we keep stuff safe

- Bitstream preservation
- Anreicherung/Verwaltung von Metadaten (Ingest und im Archival Storage)
- redundante Speicherung im Leibniz-Rechenzentrum
- Definierter Rechtlicher Rahmen
 - Preservation Policy (work in progress)
 - Datenübernamevereinbarung
- Zertifizierung durch das CoreTrust Seal (Antrag in Vorbereitung)
- ...





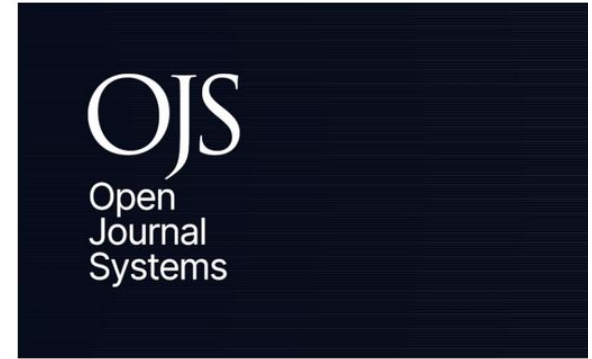
Archivierung von OJS Artikeln in Rosetta

Ein Export Workflow



OJS - Open Journal Systems

- Software zum Verwalten und Publizieren von wissenschaftlichen Zeitschriften
- Kompletter Entstehungsprozess von Einreichung bis Veröffentlichung kann abgebildet werden
- Vielzahl an vorinstallierten Plugins für Metadaten, Anbindung an Crossref, DOAJ etc.



Zielsetzung

Archivieren der veröffentlichten Artikel aller Zeitschriften in Rosetta

- Ein Eintrag pro Artikel, Dokument, Metadaten
- Sämtliche Entstehungsdaten des Artikels werden mitgespeichert
 - Submission Files
 - Entscheidungen
 - Diskussionen / Reviews
 - Daten der Beteiligten: Name, ORCID, soweit in OJS angegeben.
- Nachnutzbarkeit des Workflows durch andere Einrichtungen.



Datenstruktur in OJS

Frontend	Backend
Abstract	Submission Files
Metadaten	Reviews / Diskussionen
Artikelfahnen (PDF)	Entscheidungen
Daten der Beteiligten	Daten der Beteiligten



Export via OAI Schnittstelle

+ In OJS vorinstalliert



Frontend

Abstract



Metadaten



Artikelfahnen (PDF)



Daten der Beteiligten

Backend

Submission Files



Reviews / Diskussionen



Entscheidungen



Daten der Beteiligten



+ OAI Harvester in
Rosetta vorhanden



Nativer XML Export von OJS

- + In OJS vorinstalliert
- + Mehr Inhalte als OAI
- Im nativen OJS Format
(kein DC, nicht von
Rosetta zu verarbeiten)

Frontend	Backend
Abstract	Submission Files
Metadaten	Reviews / Diskussionen
Artikelfahnen (PDF)	Entscheidungen
Daten der Beteiligten	Daten der Beteiligten



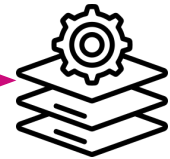
Workaround: Angereicherter XML Export (2023)

Python Script

- Nativer XML Export wird gestartet
- Script greift auf die OJS Datenbank zu und extrahiert zusätzliche Daten
- Zusätzliche Daten werden im XML eingepflegt



OJS



Rosetta



Workaround Nachteile

Umständliche Installation:

- Script läuft auf dem System und muss manuell eingerichtet werden.

Unzureichender Output / Manuelle Archivierung

- Das angereicherte XML ist nicht im DC Format und muss daher einem bestehenden Eintrag in Rosetta zugewiesen werden.

Nicht Zukunftssicher:

- Mit Version 3.4.x ist dieser Workaround nicht mehr nutzbar.



Neuer Workflow: OJS Plugin (WIP)

- Als Plugin für OJS Programmiert
 - PHP
 - einfache Installation
- Output:
 - mets.xml zum direkten Import nach Rosetta
 - Alle im Laufe des Entstehungsprozesses hochgeladenen Dateien
 - Diskussionen und Entscheidungen im XML Format
 - Direkter Push der Exporte nach Rosetta



Offene Punkte

- XML Schema der Diskussionen / Entscheidungen
- Einreichung nach Rosetta noch genau zu definieren



Vielen Dank

Bei Interesse an dem Thema freue ich mich über eine Email:

karl-arnold.bodarwe@ur.de





Retrodigitalisate

Alles irgendwie gleich?!



Rechnungsbücher des Katharinenspital

- Umfang:
 - 124 Rechnungsbücher
 - ~ 180.000 Scans
 - ~ 23 TB Speicherplatz
- Anzeige bereits in Bavarikon
- eigenes Metadatenformat (FAUST)
- Zusammenführen
 - der Strukturdaten (Fortsetzung zum Digitalisierungsprojekt)
 - der Metadaten aus der “Faust”-Datenbank
 - der digitalisierten Bilder (TIFF)



Handschriftendigitalisate der UB Erlangen

work in progress (2 Projektphasen bereits abgeschlossen)

- Umfang (Phase 2 und 3):
 - 241 Pergamenthandschriften
 - ~ 90.000 Scans
 - 219 Papierhandschriften
 - 116.00 Scans
 - > 5 TB
- bereits etablierter Workflow in Kitodo 3 nach Digitool → nach Rosetta
- Anzeige aus Rosetta (IIIF) via eigener Mirador-Instanz



Zeitungenprojekt der UB Erlangen-Nürnberg

work in progress

- Internes Digitalisierungsprojekt wegen Säurefraß
- Digitalisierungsworkflow in Kitodo 3
- Darstellung via Zeitungenportal der deutschen digitalen Bibliothek (DDB)



Vielen Dank für Ihre Aufmerksamkeit!

Bei Fragen wenden Sie sich bitte an:

karl-arnold.bodarwe@ur.de

und

michael.huber@bsb-muenchen.de

