Universität Rostock

Traditio et Innovatio

DORO ein Repositorium zur Forschungsdatenarchivierung

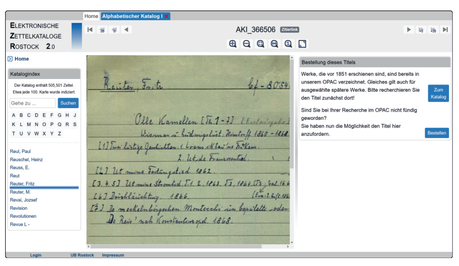
Ausgangssituation

Mit **DORO** stellt die Universitätsbibliothek Rostock Wissenschaftler_innen eine Plattform zur Verfügung, um Forschungsdaten langfristig zu archivieren und dauerhaft im Internet zu publizieren.

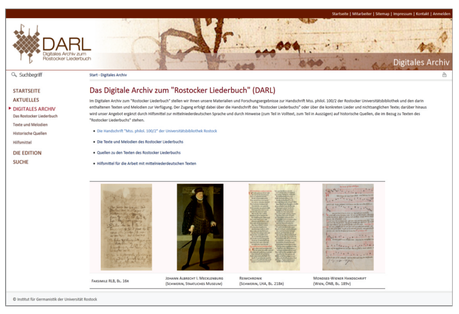
DORO dient als **Backend-Lösung** für Spezialanwendungen, die auf die Daten und Metadaten über Schnittstellen zugreifen und diese für die Präsentation nutzen können.
Gleichzeitig dient es als **Fallback-Lösung**: Kann eine Spezialanwendung aus verschiedenen Gründen (begrenzte Projektlaufzeit / technischer Alterung / Forscher_innen, die die Universität verlassen / ...) nicht mehr betrieben werden, bleiben die Daten über ihren Persistent Identifier zitier- und zugreifbar und eine vereinfachte Präsentation über DORO wird gewährleistet.

Projekte (Datenlieferanten)

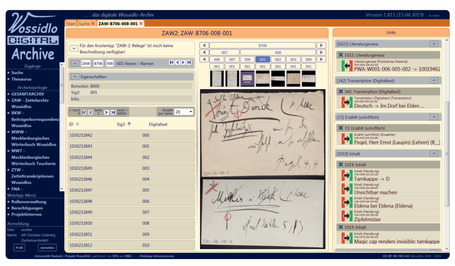
IPAC - Elektronische Zettelkataloge der Universitätsbibliothek Rostock
In einem Digitalisierungsprojekt wurden verschiedene Zettelkataloge der UB Rostock mit insgesamt ca. 550.000 Zetteln digitalisiert und im Internet zugänglich gemacht.



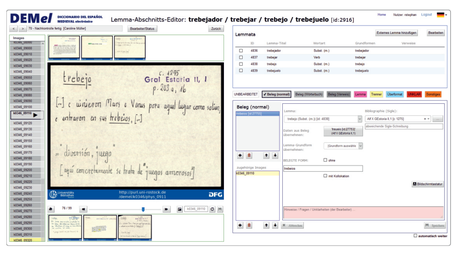
DARL - Digitales Archiv zum Rostocker Liederbuch
Das Rostocker Liederbuch ist eine mittelalterliche Handschrift mit heute noch 44 Blättern. Das Archiv enthält digitalisierte Materialien (Abbildungen, Parallelüberlieferungen, Publikationen), die einen Bezug zu diesem Werk haben.



WossiDiA - Wossidlo Digital Archive
Die handschriftliche Sammlung des mecklenburgischen Feldforschers Richard Wossidlo (1859-1939) umfasst ca. 2.3 Mio Belege mit Feldforschungsdaten, Korrespondenzen, Excerpten aus Publikationen,... die im Rahmen eines DFG-Projektes digitalisert und in einem Webportal veröffentlicht wurden.



DEMel - Diccionario del Español Medieval electrónico
In diesem laufenden DFG-Projekt wird ein umfangreiches Datenarchiv zum mittelalterlichen Spanisch digitalisiert und publiziert. Grundlage bilden ca. 900.0000 Karteikarten, die Wortbelege aus ca. 600 Werken bzw. Textsammlungen des 10. bis 15. Jh. verzeichnen.



Schnittstellen

REST-API	für das Lesen und Schreiben von Metadaten und Dateien sowie für die Suche (Proxy auf den integrierten SOLR-Server)
OAI-PMH	Konfiguration mehrerer OAI-PMH-Endpoints für den Austausch mit Nachweissystemen
IIIF-Image-API	für die Bereistellung von Bilddaten Auslieferung von Bildkacheln für verschiedene Zoomstufen

powered by <MyCoRe/>

Formate und Standards

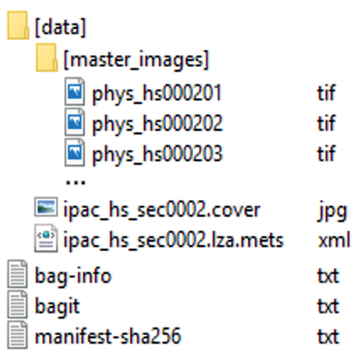
MODS - Metadata Object Description Standard
Die Forschungsdatensätze werden mit einem minmalen MODS-Metadatensatz beschrieben. Dieser enthält mindestens einen internen Objekt-Identifier, DOIs (wenn vergeben), Titel, eine Objektklassifikation und die abliefernde Einrichtung.

METS - Metadata Encoding and Transmission Scheme
Die Beschreibung der Binnenstruktur eines Forschungsdatensatzes erfolgt mit METS. Damit lassen sich beschreibende, administrative und strukturelle Metadaten der Objekte sowie wichtige Datei-Metadaten (wie Größe, Änderungsdatum, Checksumme) speichern.

Persistent Identifier
Als Persistent Identifier werden derzeit DOIs und PURLs unterstützt. Diese werden i.d.R. von der internen ID des Forschungsdatensatzes abgeleitet.

S3-Storage (Archivierung)
Die dauerhafte Speicherung der Forschungsdatensätze erfolgt redundant in je einem S3-Storage beim IT-und Medienzentrum (ITMZ) der Universität Rostock und der Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG) in Göttingen.

BagIt (Binnenstruktur des S3-Objektes)
Die Daten werden gemäß dem BagIt-Standard abgelegt und mit den dafür notwendigen Metadaten- und Checksummen-Dateien angereichert.



BagIt-Dateistruktur

```
Source-Organization: University of
  Rostock, University Library
Organization-Address:
  Universitätsbibliothek Rostock,
  Albert-Einstein-Str. 6, 18059 Rostock
Organization-Identifier:
  http://d-nb.info/gnd/25968-8
Contact-Email: digibib.ub@uni-rostock.de
Bagging-Date: 2019-04-02
External-Identifier: http://purl.uni-
  rostock.de/ipac/hs/sec0002
External-Description: Digitized Work
  from the Rostock University Library
  collections (uncompressed TIFF images)
Bag-Size: 1,657 GB
Payload-Oxum: 1778858218.135
```

bagit.txt

ZIP64 uncompressed
Der ZIP-Standard wurde 1989 entwickelt und ist heute Public Domain. In der Erweiterung ZIP64 lassen sich 2^{64} Dateien mit einer Gesamtgröße von 16 ExaByte speichern. Schaltet man die Komprimierung aus, lassen sich einzelne Dateien (sofern ihre Byte-Position bekannt ist) ohne Kenntnis des ZIP-Standards entnehmen, da ihr Byte-Strom unverändert übernommen wurde.

```
<mets:fileGrp ID="STORAGE" USE="STORAGE">
  <mets:file ID="LZA_S3_0000" USE="LZA_S3_ITMZ" CHECKSUMTYPE="MD5"
    CHECKSUM="0c3dedc4f00abcd2f2898569cd19e09f" SIZE="1778892863"
    CREATED="2019-04-02T10:12:33Z" MIMETYPE="application/zip" >
    <mets:FLocat LOCTYPE="URL" xlink:href="https://[s3.storage.url]
      /ub-arc-ipac/ipac/hs/ipac_hs_sec0002.zip"/>
    <mets:file ID="LZA_S3_0000.MASTER_IMAGES_phys_hs000138"
      CHECKSUM="1f317776" CHECKSUMTYPE="CRC32" MIMETYPE="image/tiff"
      SIZE="13466875" BEGIN="66" END="13466941" BETYPE="BYTE"
      OWNERID="data/master_images/phys_hs000138.tif" />
    [...]
  </mets:file>
</mets:fileGrp>
```

METS-Filegroup eines ZIP-Containers (mit enthaltenen Dateien)

Fazit

- Erste Erfahrungen mit dokumenten-/bildbasierten Forschungsdaten sind positiv.
- Für naturwissenschaftliche und technische Daten fehlen derzeit noch praktische Erfahrungen.
- Es ist offen, ob und wie weitere Systeme (relationale Datenbanken, XML-Datenbanken, TripleStores) für eine performante Recherche integriert werden sollen.
- Herausfordernd bleibt die Zusammenstellung (Bildung von Intellectual Entities) der Forschungsobjekte - Wie granular sollen die Daten gespeichert werden?

UNIVERSITÄTSBIBLIOTHEK ROSTOCK
Robert Stephan, Karsten Labahn
digibib.ub@uni-rostock.de

Universitäts Bibliothek Rostock