Integrating Library Metadata in a Semantic Web Research Environment for University Collections

Martin Scholz, University Library of Erlangen-Nürnberg (FAU)
University & academic collections

- > 1000 collections in Germany
- very heterogeneous material, conditions & documentation
- ~ 60% not digitally accessible
- ~ 40% with high-quality digital image

https://portal.wissenschaftliche-sammlungen.de/kennzahlen, CC-BY-NC 3.0
Collections at the University of Erlangen-Nürnberg

- > 20 collections
- heterogeneous material, size, condition and documentation
- scattered (historically and administratively)
  ⇒ till now no common presentation
  ⇒ central custodial agency
  ⇒ digitization strategy

https://www.fau.de/universitaet/das-ist-die-fau/sammlungen-der-fau/
The project “Objekte im Netz” (2017-2020)

Goals:
➢ Common standards for (digital) documentation
➢ Best practices, guidelines & tools

Means:
➢ 6 pilot collections: graphics, medicin history, mineralogy, music, prehistoric archaeology, school history
➢ WissKI as common indexing and research tool
➢ CIDOC CRM as common data model

http://objekte-im-netz.fau.de
WissKI (Wissenschaftliche KommunikationsInfrastruktur)

- virtual research environment for cultural heritage documentation
- for complex, network-like data
- data stored natively as CIDOC CRM / OWL
- wiki-like aggregation of information
- XAMP - Drupal - WissKI

http://wiss-ki.eu
WissKI approach: ontology paths

Backend:
- Data stored as RDF triples
- Local & external sources

Frontend:
- Aggregated view
- Mixed media (tabular, textual, image, …)
WissKI approach: ontology paths

Path patterns are used to aggregate information from the triple data
Collection model

Common top ontology based on CIDOC CRM
Domain ontologies for collection specifics

Class “Collection object” as main entry point
The graphics and prints collection

Small but renowned collection: paintings, graphics, prints, maps, …

~5000 prints, thereof:
2162 are catalogued according to bibliographic rules and available online
12 digitized images available

Sisis / local ⇒ item information
Aleph / library network ⇒ expression / work information
Graphics Collection as part of Objekte im Netz

case study: how to integrate bibliographic metadata into the collection model / database?

piloting with ~2000 prints

data accessible via OAI-PMH + SRU in MARCxml
Data integration workflow (first approach)

1. fetch data from OAI-PMH and SRU on demand ⇒ MARCxml records
2. convert MARCxml to BibFrame with marc2bibframe2 (xslt scripts) ⇒ RDF triples
3. provide (rudimentary) LOD-REST-API
4. align BibFrame with CIDOC CRM (with help of FRBRoo): ⇒ build congruent ontology paths
5. integrate library data as external “authority” ⇒ authority data dynamically enriches local WissKI data

“correct & neat” from LOD perspective
Data integration workflow (current approach)

1. periodically fetch data from OAI-PMH and SRU ⇒ MARCxml records
2. store records in SQL table
3. convert MARCxml to CIDOC CRM using WissKI SQL Import feature ⇒ build triples directly according to local model & mapping file
4. import library data into local WissKI data ⇒ library data becomes part of local data and is periodically updated

“quick & dirty” from LOD perspective
Why not first approach?

Mainly practical issues…

Incomplete / incorrect / inconvenient conversion to BibFrame
  ⇒ special fields, deviating semantics; blank nodes

Ontological “mismatches” between BibFrame and CIDOC CRM
  ⇒ BibFrame is less verbose ⇒ missing intermediate nodes / resources
  ⇒ virtual mismatches due to conversion

Fetch-on-demand or import / Authority data or local data
  ⇒ affects performance and search
Further observations

Technical hindrances: half-conforming APIs for OAI-PMH and SRU
  client libraries (e.g. phpoaipmh) fail
Missing URIs: no officially coined URIs for items or expressions by library network
  ⇒ own URIs (as with other collections)
Unique objects vs. serial production / item vs. work
  ⇒ other collection domains don’t apply FRBR concepts ⇒ divergent models

BibFrame is used in the background to evaluate the local modelling / mapping
Thank you!