

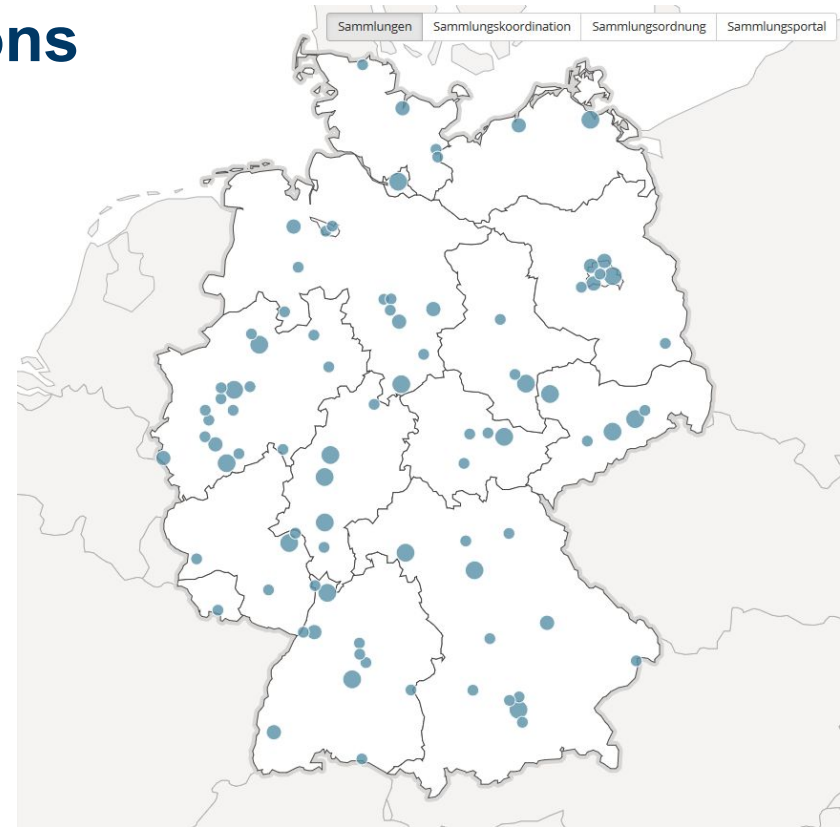
Integrating Library Metadata in a Semantic Web Research Environment for University Collections

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, ...)

The screenshot displays the Patrimonium.net interface. At the top, there's a navigation menu with options like 'Content management', 'Site building', 'Site configuration', 'User management', 'Reports', and 'Help'. The main header reads 'Patrimonium.net Digital 3D Reconstruction in Virtual Research Environments'. Below this, a specific entity is highlighted: 'FOTO: Paul von Hindenburg on a visit at Friedrichstein manor'. A network graph shows this entity connected to various other entities, including 'Paul von Hindenburg', 'Friedrichstein', 'Jan Lutt', and 'Circ - Carum'. On the right side, there's a search bar and a list of project pages. The footer features logos of partner institutions: Herder Institut, UAM, Justus-Liebig-Universität Gießen, PSNC, FAU, and Germanisches Nationalmuseum.

WissKI approach: ontology paths

Path patterns are used to aggregate information from the triple data

Menu Item	Path	Enabled	Operations
Ungrouped		<input checked="" type="checkbox"/>	
Museumsobjekt	Group [ecm:E84_Information_Carrier]	<input checked="" type="checkbox"/>	edit delete
Inventarnummer	ecm:E84_Information_Carrier -> ecm:P1_is_identified_by -> ecm:E42_Identifier	<input checked="" type="checkbox"/>	edit delete
Abteilung	ecm:E84_Information_Carrier -> ecm:P50_has_current_keeper -> ecm:E40_Legal_Body -> ecm:P1_is_identified_by -> ecm:E82_Actor_Appellation	<input checked="" type="checkbox"/>	edit delete
Bezeichnung	ecm:E84_Information_Carrier -> ecm:P1_is_identified_by -> ecm:E41_Appellation	<input checked="" type="checkbox"/>	edit delete
Titel	ecm:E84_Information_Carrier -> ecm:P1_is_identified_by -> ecm:E35_Title	<input checked="" type="checkbox"/>	edit delete
Herstellung	Group [ecm:E84_Information_Carrier -> ecm:P1081_was_produced_by -> ecm:E12_Production]	<input checked="" type="checkbox"/>	edit delete
Hersteller	ecm:E84_Information_Carrier -> ecm:P1081_was_produced_by -> ecm:E12_Production -> ecm:P14_carried_out_by -> ecm:E21_Person -> ecm:P1_is_identified_by -> ecm:E82_Actor_Appellation	<input checked="" type="checkbox"/>	edit delete
Zeit	ecm:E84_Information_Carrier -> ecm:P1081_was_produced_by -> ecm:E12_Production -> ecm:P4_has_time-span -> ecm:E52_Time-Span	<input checked="" type="checkbox"/>	edit delete
Ort	ecm:E84_Information_Carrier -> ecm:P1081_was_produced_by -> ecm:E12_Production -> ecm:P7_look_place_at -> ecm:E53_Place -> ecm:P1_is_identified_by -> ecm:E48_Place_Name	<input checked="" type="checkbox"/>	edit delete
Material	ecm:E84_Information_Carrier -> ecm:P1081_was_produced_by -> ecm:E12_Production -> ecm:P32_used_general_technique -> ecm:E57_Material -> ecm:P1_is_identified_by -> ecm:E75_Conceptual_Object_Appellation	<input checked="" type="checkbox"/>	edit delete
Technik	ecm:E84_Information_Carrier -> ecm:P1081_was_produced_by -> ecm:E12_Production -> ecm:P33_used_specific_technique -> ecm:E29_Design_of_Procedure -> ecm:P1_is_identified_by -> ecm:E75_Conceptual_Object_Appellation	<input checked="" type="checkbox"/>	edit delete
Inhalt Beschreibung	ecm:E84_Information_Carrier -> ecm:P701_is_documented_in -> ecm:E31_Document	<input type="checkbox"/>	edit delete
Inschrift	ecm:E84_Information_Carrier -> ecm:P128_carries -> ecm:E34_Inscription	<input type="checkbox"/>	edit delete
Ikonografie	ecm:E84_Information_Carrier -> ecm:P82_depicts -> ecm:E1_CRM_Entity	<input checked="" type="checkbox"/>	edit delete
Literatur	ecm:E84_Information_Carrier -> ecm:P701_is_documented_in -> ecm:E31_Document	<input checked="" type="checkbox"/>	edit delete
Bilder	ecm:E84_Information_Carrier -> ecm:P65_show_visual_item -> ecm:E98_Image -> ecm:P1_is_identified_by -> ecm:E42_Identifier	<input checked="" type="checkbox"/>	edit delete

Content management Site building Site configuration User management Reports Help

Patrimonium.net
Digital 3D Reconstruction in Virtual Research Environments

Leibniz Association Project Tasks Logout

Create Navigate Find

FOTO: Paul von Hindenburg on a visit at Friedrichstein manor

View Create and Link Tool Delete Edit Form Graph Log Merge Network Paths Triples XML Devel

Southwest facade... Avant-corps main... Standard View

Paul von Hindenburg... Friedrichstein... 3-422-06593-8... Entry... 03.12.2014 13:52:12... Title Set... Paul von Hindenburg bei einem seiner Besuche in Friedrichstein... Paul von Hindenburg... Paul von Hindenburg on a visit at Friedrichstein manor... Coverage Object... Paul von Hindenburg... Photographie... CIRC - CIRCUM...

Photo → R26 documents → Hindenburg Hindenburg → P131 is identified by → Name Name → P3 has note → „Paul von Hindenburg“

HERDER INSTITUT
JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
PSNC FAU
GERMANISCHES NATIONALMUSEUM

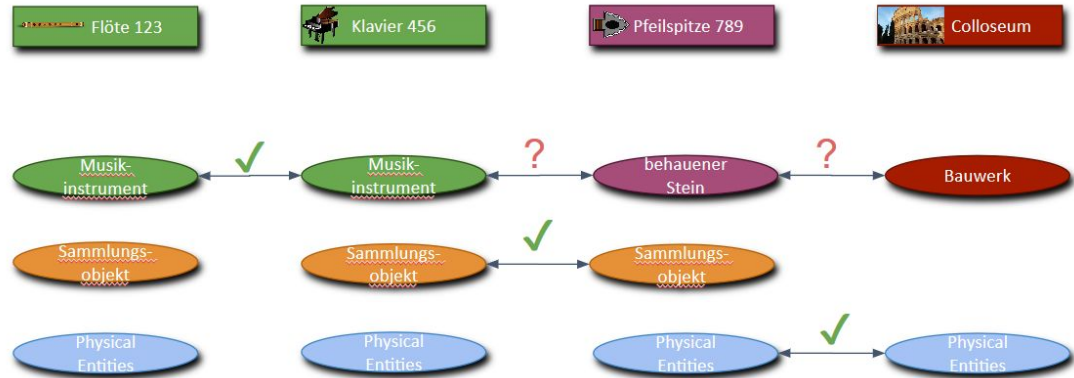
http://www.patrimonium.net

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



Albrecht Altdorfer: Das Urteil des Paris, 1511,
Signatur: H62/AH 13

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

```
<datafield tag="100" ind1="1" ind2=" " >
  <subfield code="a">D&#xFC;rer, Albrecht</subfield>
  <subfield code="d">1471-1528</subfield>
  <subfield code="e">Verfasser</subfield>
  <subfield code="0">(DE-588) 11852786X</subfield>
  <subfield code="4">aut</subfield>
</datafield>
<datafield tag="240" ind1="1" ind2="0">
  <subfield code="0">(DE-588) 4540702-2</subfield>
  <subfield code="a">&#x98;Das&#x9C; Rhinoceros</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">&#x98;Das&#x9C; Rhinoceros</subfield>
  <subfield code="c">Albrecht D&#xFC;rer</subfield>
</datafield>
<datafield tag="246" ind1="1" ind2="3">
  <subfield code="a">Dessein du Rhinoc&#xE9;rus</subfield>
</datafield>
<datafield tag="246" ind1="1" ind2="3">
  <subfield code="a">Rhinoceros</subfield>
</datafield>
<datafield tag="264" ind1=" " ind2="0">
  <subfield code="c">[1515]</subfield>
</datafield>
```

```
SELECT (CONCAT_WS(' ', MARC100$a, MARC240$a, MARC245$a, MARC246$a)) AS title,
       MARC100$d AS date,
       MARC100$e AS author,
       MARC240$0 AS subject,
       MARC245$c AS subject2,
       MARC264$c AS date2
FROM marcxml
WHERE MARC100$a LIKE 'D&#xFC;rer, Albrecht'
AND MARC240$a LIKE '%Rhinoceros%'
AND MARC245$a LIKE '%Rhinoceros%'
AND MARC246$a LIKE '%Dessein du Rhinoc&#xE9;rus%'
AND MARC246$a LIKE '%Rhinoceros%'
AND MARC264$c LIKE '[1515]'
```



WissKI SQL Import

```
<marc:datafield tag="049" ind1=" " ind2=" " >
  <marc:subfield code="a">DE-29</marc:subfield>
</marc:datafield>
<marc:datafield tag="100" ind1="1" ind2=" " >
  <marc:subfield code="a">Dürer, Albrecht</marc:subfield>
  <marc:subfield code="d">1471-1528</marc:subfield>
  <marc:subfield code="e">Verfasser</marc:subfield>
  <marc:subfield code="0">(DE-588) 11852786X</marc:subfield>
  <marc:subfield code="4">aut</marc:subfield>
</marc:datafield>
<marc:datafield tag="240" ind1="1" ind2="0">
  <marc:subfield code="0">(DE-588) 4540702-2</marc:subfield>
  <marc:subfield code="a">&#x98;Das&#x9C; Rhinoceros</marc:subfield>
</marc:datafield>
<marc:datafield tag="245" ind1="1" ind2="0">
  <marc:subfield code="a">&#x98;Das&#x9C; Rhinoceros</marc:subfield>
  <marc:subfield code="c">Albrecht Dürer</marc:subfield>
</marc:datafield>
<marc:datafield tag="246" ind1="1" ind2="3">
```

WissKI

Graphische Sammlung

[My account](#) [Log out](#)

Home
Objekte in Bearbeitung
Find
Navigate
Create

Home » Navigate » Sammlungsobjekt

H62/AH 266: Das Rhinoceros

View
Edit
Delete
Triples
Devel

<p>Inventarnummer H62/AH 266</p> <p>BV.Nr. BV044521316</p> <p>Bezeichnung/Titel Das Rhinoceros</p> <p>Normdaten http://d-nb.info/gnd/4540702-2</p> <p>Motiv http://d-nb.info/gnd/4540702-2</p> <p>Herstellung Hersteller (Person) Dürer, Albrecht</p> <p>Herstellungsdatum Datum (informativ) [1515]</p> <p>Frühestmöglicher Startzeitpunkt (maschinell codiert) 1515-01-01</p> <p>Spätestmöglicher Endzeitpunkt (maschinell codiert) 1515-12-31</p> <p>Material Holzschnitt</p> <p>Technik Holzschnitt</p> <p>Messung Messergebnisse Größe 21.2 x 29.8 cm</p> <p>Provenienz (Referenztitel) Provenienz: Graphische Sammlung der Markgrafen von Ansbach</p> <p>Provenienz Vorbesitzer (Institution) Graphische Sammlung der Markgrafen von Ansbach</p> <p>Datum Datum (informativ)</p>	<p>Literaturvermerk</p> <p>Kurztitel Bartsch 7</p> <p>Seite VII 147, 136</p> <p>URL http://digi.ub.uni-heidelberg.de/diglit/bartsch1808bd7/0151</p> <p>Kurztitel Geisberg 2</p> <p>Seite Band II, 684, 721</p> <p>Kurztitel Kassler-Luhde 1</p> <p>URL http://nbn-resolving.org/urn:nbn:de:hbz:6-bv-009445844-0108-8</p>
---	--

WissKI Linkblock

Objekte desselben Herstellers (Person)

- H62/AH 266
- H62/AK 849/850
- H62/AK 851 a
- H62/AK 851 b
- H62/AK 852 a
- H62/AK 852 b
- H62/AK 853
- H62/AK 858
- H62/AK 859
- H62/AK 860 a
- H62/AK 860 b
- H62/AK 861 a
- H62/AK 861 b
- H62/AK 862
- H62/AK 863
- H62/AK 871
- H62/AK 872
- H62/AK 873
- H62/AK 874
- H62/AK 875
- H62/AK 876 a
- H62/AK 876 b
- H62/AK 877
- H62/AK 878
- H62/AK 879
- H62/AK 880
- H62/AK 897
- H62/AK 899
- H62/AK 900 a
- H62/AK 900 b
- H62/AK 901
- H62/AK 902 a
- H62/AK 902 b
- H62/AH 279
- H62/AH 280

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!