Using semantic web technologies for the digitization of heterogeneous university collections in the long run
The research project “Objekte im Netz” (2017-2020)

Goals:
Development of a common uniform digital documentation, storage and presentation (incl. digital infrastructure)
→ standards for digital documentation
→ best practices, guidelines & tools

Basis until 2020:
● 6 representative collections: graphics, medicine, music, geology, school history, prehistoric archaeology
● WissKI as research infrastructure and presentation tool
● CIDOC CRM as reference ontology
ICOM CIDOC Conceptual Reference Model (ISO 21127)

- Ontology for Documentation in the Cultural Heritage Domain
- Lingua Franca of the scientific fields
- Support of data exchange and Interoperability
- Long-term Interpretability
- approx. 90 classes (e.g. Physical Things, Actors, Places, Concepts, Events...) and 150 relations
- Special feature: event-centered
- Expandable
• CIDOC CRM is a paper based Document - in the Semantic Web the format for data models like the CIDOC CRM typically is RDF or OWL
• First Implementation in OWL at the University of Erlangen-Nürnberg in 2007
• One and only actively maintained OWL implementation of the CIDOC CRM
• The current ISO standard is from 2014
• Ontology used by Objekte im Netz: ECRM 170309 / CIDOC-CRM 6.2.2

http://www.erlangen-crm.org
https://github.com/erlangen-crm/ecrm
Common Ontology and Metadata schema

Main Instances:
- **S1 Collection Object** (sub: E84 Information Carrier)
- S53 Subcollection (sub: E78 Curated Holding)
- E21 Person
- S86 Organisation (sub: E40 Legal Body)
- S39 Location (sub: E53 Place)
- S40 Geographical Place (sub: E53 Place)
- E57 Material
- S93 Collection Object Classification (sub: E55 Type)
- E38 Image
- S68 Authority File (sub: E32 Authority Document)
Collection specific Ontologies and Schemas
WissKI

Scientific Communication Infrastructure
(Wissenschaftliche KommunikationsInfrastruktur)

WissKI is ...

• a virtual research environment for scientific research data
• based on idea of the Wiki
• accessible online via web browser from everywhere
• open source and free for download
• compatible to ISO 21127 (CIDOC CRM)
• a lot of interfaces
• support for usage of authority files like Getty TGN and PND
• ideal system for Linked Open Data

http://wiss-ki.eu
http://www.facebook.com/wissskiproject
WissKI System Architecture

WissKI is...

- a module set for the CMS Drupal
- using all features that drupal provides e.g. user management, web pages, forums, wysiwyg editors, ...
- storing its data in a triple store
<table>
<thead>
<tr>
<th>Object</th>
<th>Group [ecrm:E84_Information_Carrier]</th>
<th>edit</th>
<th>delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inventory number</td>
<td>ecrm:E84_Information_Carrier -&gt; ecrm:P1_is_identified_by -&gt; ecrm:E42_identifier</td>
<td>✓</td>
<td>edit delete</td>
</tr>
<tr>
<td>Collection</td>
<td>ecrm:E84_Information_Carrier -&gt; ecrm:P50_has_current_keeper -&gt; ecrm:E40_Legal_Body -&gt; ecrm:P1_is_identified_by -&gt; ecrm:E82_Actor_Appellation</td>
<td>✓</td>
<td>edit delete</td>
</tr>
<tr>
<td>Label</td>
<td>ecrm:E84_Information_Carrier -&gt; ecrm:P1_is_identified_by -&gt; ecrm:E41_Appellation</td>
<td>✓</td>
<td>edit delete</td>
</tr>
<tr>
<td>Title</td>
<td>ecrm:E84_Information_Carrier -&gt; ecrm:P1_is_identified_by -&gt; ecrm:E35_Title</td>
<td>✓</td>
<td>edit delete</td>
</tr>
<tr>
<td>Dimensions</td>
<td>ecrm:E84_Information_Carrier -&gt; ecrm:P30l_wasMeasured_by -&gt; ecrm:E10_Measurement -&gt; ecrm:P40_observable_dimension -&gt; ecrm:E54_Dimension</td>
<td>✓</td>
<td>edit delete</td>
</tr>
<tr>
<td>Place</td>
<td>ecrm:E84_Information_Carrier -&gt; ecrm:P108_lWasProduced_by -&gt; ecrm:E12_Production -&gt; ecrm:P7_located_place_entity -&gt; ecrm:E50_Place -&gt; ecrm:P1_is_identified_by -&gt; ecrm:E49_Place_Name</td>
<td>✓</td>
<td>edit delete</td>
</tr>
<tr>
<td>Inscription</td>
<td>ecrm:E84_Information_Carrier -&gt; ecrm:P128_carees -&gt; ecrm:E34_Inscription</td>
<td>✓</td>
<td>edit delete</td>
</tr>
</tbody>
</table>
Semantic data modelling

Albrecht Dürer

Nürnberg

E82 Actor Appellation

P131 is identified by

E21 Person

P14 carried out by

E84 Information Carrier

P108i was produced by

E12 Production

P7 took place at

E53 Place

P87 is identified by

E48 Place Name

P3 has note

„Albrecht Dürer“

„Nürnberg“
WissKI in use at “Objekte im Netz”

- as data entry and presentation system for each of the 6 collections
- as portal for cross-collection presentation and research

http://objekte-im-netz.fau.de/portal/
Dataflow between the single systems and the portal

- **common context**
  - Collection Portal FAU
  - Data Graphical Collection
  - Data Medical Collection

- **collection-specific context**
  - WissKI Graphical Collection
  - WissKI Geological Collection
  - WissKI...

→ cross-collection presentation and research based on the common schema

→ collection specific documentation and presentation based on the common and specific schema
Managing the data schemas - WissKI Pathbuilder

Pathbuilder Overview of the Geological Collection

Pathbuilder, Common Schema

Pathbuilder, Extension Geology
Continuation of OiN: Adding the Herbar

The „Herbarium Erlangense“ (https://objekte-im-netz.fau.de/herbar/) is one of the first early adopters of WissKI at the FAU

- The Erlangen CRM is an older version

- The OiN Common Ontology is derived from the Herbar-Ontology but:
  - The language differs
  - Some classes and properties differ
  - Some paths in the pathbuilder are not existing in the one or the other case, have different semantics or even are structurally different
Continuation of OiN: Adding the Herbar

Problem 1: What should go where?

- Spotting which concepts of the Herbar ontology already are represented by concepts of the OiN common ontology
- Is a special Herbar extension needed and which classes/properties should exist there?
- Do we have to make any changes to the OiN common ontology?
Continuation of OiN: Adding the Herbar

Problem 2: Old ECRM-Version

- Herbar was built with ECRM 120111
- OiN uses ECRM 170309
- Luckily the changes were just minor and only additive – in contrast to the steps that will have to be taken to move to most recent versions of ECRM
- So search and replace of 120111 to 170309 did the job. This was easy and took nearly no time.
Continuation of OiN: Adding the Herbar

Problem 3: Differences between Herbar ontology and OiN common ontology

- Differs in language and other 1:1 mapping of concepts and properties
  -> search and replace class by class and property by property

- Structural equivalent mappings and Paths that became shorter
  -> Insert the new triples via SPARQL and delete the old triples afterwards

- Paths that become longer:
  - Creation of new URIs via SPARQL is complicated… but you can cheat it.
Continuation of OiN: Adding the Herbar

After doing these steps we could use the existing pathbuilder for the OiN common ontology and build a small one specifically for the Herbar extension. Next step will be the addition of the data to the University Collections Portal.

We will have to do this workflow with every already existing collection. Can we shortcut it in any way?
Download

- WissKI-Software
  → https://www.drupal.org/project/wisski

- Ontologies (draft)
- Path Templates (draft)
- Guidelines for Editing (draft)
- WissKI Manual for Collection Staff
  → http://objekte-im-netz.fau.de/projekt
Thank you for your attention!