Specialising the EDM for Digitised Manuscripts

Kai Eckert¹, Steffen Hennicke², Evelyn Dröge², Julia Iwanowa², Violeta Trkulja²

¹Universität Mannheim, ²Humboldt-Universität zu Berlin

Semantic Web in Libraries - Hamburg, 27.11.2013

co-funded by the European Union
Digitised Manuscripts to Europeana

• EU-funded Europeana satellite project
• Duration: Three years (2012 – 2015)
• Partners from Germany, Austria, Norway, Greece, UK and Italy

• DM2E works on:
  – a tool-chain for data migration to Europeana and the LOD Web (OMNOM),
  – a digital research environment for the Digital Humanities (PUNDIT),
  – an open community of cultural heritage professionals (OPENGLAM)
DM2E: Infrastructure
DM2E: Provided Content

- Metadata about manuscripts:
  - Described by: TEI, MAB2, MARC, EAD, METS/MODS
    Database content
  - In different languages
  - 118,000+ items
  - 20,006,930+ pages
DM2E: Data Model

- Semantically and structurally heterogeneous data
  - e.g. EAD, METS, TEI, MARCXML and MAB2, relational databases, proprietary schemas

- The Europeana Data Model (EDM) is made for this scenario!
  - provides a generic semantic interoperability layer
  - enables the definition of “applications profiles” which may address the needs of specific communities

- The DM2E Data Model (DM2E)
  - is an “application profile” of the EDM for the domain of handwritten manuscripts
  - retains rich descriptions by specialising the EDM
DM2E: Specialisation approach

- RDF(S) allows the specialisation of EDM classes and properties
  - use of rdfs:subClassOf
  - use of rdfs:subPropertyOf

- An “application profile” typically also includes
  - additional ontological restrictions
  - documentation
DM2E: Specialisation Guidelines

• Empirical analysis of provided source metadata
• Iterative mappings to the EDM
• Close cooperation with data providers
  – agree on shared conceptualisations
• Create rich and connected representations
  – retain original semantics as much as possible
  – use existing URIs of resources
  – assign a class to the resources (rdf:type)
DM2E: Interoperability approach

• Create new classes or properties in the DM2E-Namespace only if there is no other suitable option available
  – reuse existing namespaces (ontologies)
  – mind existing semantics (scope notes, domains, ranges)

• Types, roles and relations between agents
  – Friend-of-a-Friend (FOAF) [FOAF] (types of agents)
  – Publishing Roles Ontology (PRO) [SPAR] (roles of agents in the publication process)
  – VIVO [VIVO] (types of agents)

• Detailed semantics on bibliographic entities
  – FRBR-aligned Bibliographic Ontology (FaBiO) [SPAR]
  – Citation Typing Ontology (CiTO) [SPAR]
  – Bibliographic Ontology (BIBO) [BIBO]
DM2E Model: Class-Specialisation

- 23 new or reused classes, mainly for
  - physical and conceptual parts of a handwritten manuscripts
  - as found in our source metadata
  - different types of Agents
Physical and tangible aspects of handwritten manuscripts.
Different types of agents.

- edm:Agent
- foaf:Person
- foaf:Organisation
- dm2e:Archive
- vivo:Library
- vivo:Museum
- vivo:University

is-a
http://www.europeana.eu/schemas/edm/
http://xmlns.com/foaf/0.1/
http://onto.dm2e.eu/schemas/dm2e/1.0/
http://vivoweb.org/ontology/core#
DM2E Model: Properties-Specialisation

- **Property-centric modelling**
  - more than 50 new properties

- **Documentation for the DM2E Data Model contains only EDM properties which are utilized**
  - to keep the documentation clear
  - e.g. dcterms:replaces, dc:source, or dc:conformsTo are not used

- **Domain and Range Restrictions**
  - some OWL-Restrictions on properties in order to encourage the use of specific resources of a specific type, e.g.
    - CHO hasPart CHO
    - WebResource hasPart WebResource

- **Some EDM-Properties are mandatory in DM2E**
  - dc:type: at least one of the physical (e.g. dm2e:Page) or logical (e.g. dm2e:Paragraph) aspects
  - dc:subject: ideally an URI from a controlled vocabulary
Example: Adding new properties as subproperties for `dcterms:creator`.
Outlook: Uncertain Statements

Part of the next model version: How to deal with uncertain timespans and presumably creators?

- Problem: Confidence declarations for RDF-statements need Named Graphs or Reification

- Solution:

<table>
<thead>
<tr>
<th>Agents</th>
<th>Timespans</th>
</tr>
</thead>
<tbody>
<tr>
<td>„The creator of the CHO is presumably Goethe.“</td>
<td>„The timespan was somewhere in the 1920ies and lasted 2 years.“</td>
</tr>
</tbody>
</table>

res1 dc:creator presumableAgent1.
presumableAgent1 a PresumableAgent;
isPresumably goethe;
confidence 0.8.

timeSpan1 a edm:TimeSpan.
uncertainBegin 1920;
uncertainEnd 1929;
duration 2.

Confidence is optional
Duration is optional
The PDF and the OWL representations can be accessed via the project’s website:

dm2e.eu/document/#DM2EModelSpecification
• Human & machine readable
• Version 1.0
Summary

• The DM2E Data Model is an application profile of the EDM for the domain of Manuscripts

• DM2E v1.0: Latest and first operational version

• DM2E v1.1: Next version under development

• Work is on-going and feedback welcome!
Thank you for your attention!

Questions and Feedback:
Steffen Hennicke,
Julia Iwanowa,
Evelyn Droege.

vorname.nachname@ibi.hu-berlin.de