Europeana Data Model

Stefanie Rühle (SUB Göttingen)
Slides by Valentine Charles

08th Oct. 2014, DC 2014
Europeana: A broad, heterogeneous network

2,300 galleries, museums, archives and libraries
Europeana: A broad, heterogeneous network

→multiple data manipulation and transitions from one data model to another
  • Data provider can directly map their original data to the Europeana Data Model
  • Aggregator usually developed intermediary domain profiles

→Mapping exercise differs due to
  • The nature of the metadata standards used by the provider
  • The nature of the EDM Profile used by an aggregator
Europeana Data Model (EDM)

→ A data model that re-uses several existing Semantic Web-based models
  • OAI ORE (Open Archives Initiative Object Reuse & Exchange)
  • Dublin Core
  • SKOS (Simple Knowledge Organization System)
  • CIDOC-CRM

→ It represents more granular metadata
  • links e.g. between objects and context entities (persons, places)
  • multilingual & semantic linked data for contextual resources (e.g. Concepts)
Different semantic grains

→Adopts Semantic Web principle of specialising classes and properties

→ Enables extensions, “applications profiles”, based on needs and best practices from specific sectors or domains
EDM rationale: requirements

1. Distinguish “provided objects” (painting, book, movie, etc.) from their digital representations
2. Distinguish object from its metadata record
3. Allow multiple records for a same object, containing potentially contradictory statements about it
4. Support for objects that are composed of other objects
5. Support for contextual resources, including concepts from controlled vocabularies
EDM main classes

Groups of things that have common properties, e.g., web resources.
EDM main properties

Attributes or characteristics of resources e.g., a member of the class Agent will have a name.

EDM conceptual classes

Representing (real-world) entities related to a provided object as fully fledged resources, not just strings
Contextual entities

Representing (real-world) entities related to a provided object as fully fledged resources, not just strings

**edm:Agent**
- foaf:name
- skos:altLabel
- rdaGr2:biographicalInformation
- rdaGr2:dateOfBirth

**skos:Concept**
- skos:prefLabel
- skos:altLabel
- skos:broader
- skos:related
- skos:definition

**edm:TimeSpan**
- skos:prefLabel
- dcterms:isPartOf
- edm:begin
- edm:end

**edm:Place**
- wgs84_pos:lat
- wgs84_pos:long
- skos:prefLabel
- skos:note
- dcterms:isPartOf
EDM basic pattern

→ A data provider submits to Europeana a “bundle” of an object and its digital representation(s)
An example

edm:WebResource

edm:ProvidedCHO

Clavecin

Description: 2 clavières : C / E à C''', 45 notes * 3 rangs de Cordes : 2 x 8'' + 1 x 4'' * Table en épicea (?) * T * Collection Geneviève Thibault de Chambure

Creator: http://www.mimo-db.eu/InstrumentMaker/Person/593; Cristofori

Coverage: http://sws.geonames.org/3176959/

Date: fin 17e

Type: http://www.mimo-db.eu/InstrumentsKeywords/2251; http://www.mimo-db.eu/HornbostelAndSachs/6461

Identifier: #CM:0161930

Data provider: Cité de la musique

Provider: MIMO - Musical Instrument Museums Online

Providing country: MUL

ore:Aggregation
Modeling data in EDM

→When modeling, mapping data to EDM providers have to make sure
  • the semantics of the original data is kept
  • the constraints from the EDM model are respected
  • the data are syntactically correct

→In order to solve those issues, providers take specific decisions on how to implement the mapping.
  • Schema (or ontology) level decisions: decisions that affect any records formatted using the schema.
  • Instance level decisions: decisions that affect specific (sets of) metadata records to be transformed with the mapping.
Example: define the ProvidedCHO

edm:ProvidedCHO
#CM:0161930

"Clavecin"@fr
"Cristofori"

dc:title

dc:creator

dc:description

"2 claviers : C / E à C"', 45 notes * 3
rangs de Cordes : 2 x 8" + 1 x 4" * Table
en épicéa (?) * T * Collection Geneviève
Thibault de Chambure"@fr

skos:Concept
http://www.mimo-db.eu/
InstrumentsKeywords/2251

skos:prefLabel
"Clavecin"@fr

→ Contextual resources – multilingual &
semantic linked data for Concepts
Example: Provide richer metadata using contextual resources

Dolní Kounice

From Wikipedia, the free encyclopedia

Dolní Kounice (Czech pronunciation: [dɔːlniː kʊɲɪtsɛ]; German: Kanim) is a small town in the South Moravian Region of the Czech Republic. It has around 2,400 inhabitants.
Mapping, extensions and refinements of EDM

Many projects and partners have worked on enabling metadata interoperability with Europeana

- either directly mapping their metadata to EDM according to the guidelines
- or create specialisations of EDM for representing specific use-cases and minimising the loss of metadata from their original metadata format to EDM
Mapping to EDM

- A mapping is any kind of specification of relationships and equivalences between two metadata formats or models
  - on a structural and a semantic level

- Correspondences can be defined at
  - schema-level correspondences (definition of constraints)
  - Instance-level which specifies functions that allow interoperability of the content values (e.g. combining the first and last name into one field that is called creator)

- A mapping can be also referred to as crosswalk
  - enable elements defined in one metadata format to be available to communities using related metadata formats.
METS/MODS2EDM – Main Classes

```xml
<edm:WebResource>
  http://digital.lb-oldenburg.de/ihd/id/44803
</edm:WebResource>

<edm:file MIMEType="image/jpeg" CREATED="2010-09-24T04:02:29Z" ID="IMG_DEFAULT_44816">
  <edm:WebResource>
    http://digital.lb-oldenburg.de/ihd/download/webcache/1004/44816
  </edm:WebResource>
</edm:file>
```

```
<edm:object>
  <edm:aggregatedCHO>
    <edm:ProvidedCHO>
      <edm:WebResource>
        http://digital.lb-oldenburg.de/ihd/id/44803
      </edm:WebResource>
    </edm:ProvidedCHO>
    <edm:isShownAt>
      http://digital.lb-oldenburg.de/ihd/id/44803
    </edm:isShownAt>
  </edm:aggregatedCHO>
</edm:object>
```
EDM refinement or application profile

A refinement is any kind of specialisation of EDM to meet specific needs of the data provider.

- a set of guidelines or rules are applied to the classes or properties
- the property or class being refined will be used in a narrower, but still compatible, sense

A refinement can be associated to an application profile

- set of metadata properties, policies and guidelines defined for a specific application with specific requirements.
  - describes what a community wants to accomplish with its application
  - characterizes the types of things described by the metadata and their relationships
  - enumerates the metadata elements to be used and the rules for their use
  - defines the machine syntax that will be used to encode the metadata.
Refinement

Specification of the Europeana Data Model

3.3.24 Provenance

<table>
<thead>
<tr>
<th>Property name: provenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>edm:ProvidedCHO dcterms:provenance dcterms:ProvenanceStatement rdf:value rdfs:Literal</td>
</tr>
</tbody>
</table>

3.3. Unterfacette Eigentümer

Definition: Personen und Organisationen, die das Analoges Objekt oder Digitales Primärobject besitzen oder zur Verfügung stellen (z. B. das Museum, dem eine Skulptur gehört oder die Bibliothek, die ein Buch besitzt).

<table>
<thead>
<tr>
<th>Obligation &amp; Occurrence</th>
<th>Optional (Minimum: 0, Maximum: unbounded)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>A philanthropist may have donated a work of art to a museum.</td>
</tr>
<tr>
<td>Comment</td>
<td>This relates to the ownership and custody of the original analog or born digital object.</td>
</tr>
</tbody>
</table>

Europeana Data Model Mapping Guidelines

dcterms:provenance literal or reference min 0, max unbounded
EDM extension

→ An EDM extension is also a type of refinement of EDM, but is characterised by any addition of classes or properties (implemented by XML attributes and elements) to the original EDM specifications.

→ An EDM extension is required when existing EDM classes and properties cannot represent the semantics of providers’ metadata with sufficient details.

→ Define a new set of classes and properties that are declared as specialisations of the existing ones.
  • new (sub-)classes or (sub-)properties can be taken from other existing namespaces
  • or newly created as part of an ontology or schema.
EDM extension

Digital Manuscripts to Europeana project has developed a EDM extension for contemporary handwritten manuscripts.
The specialisation of the EDM classes and properties is done by using constructs from RDF Schema:
- \textit{rdfs:subClassOf} to state that all the instances of one class are instances of another.
- \textit{rdfs:subPropertyOf} to state that all resources related by one property are also related by another.

New classes or properties were created only if there is no other suitable option available from existing ontologies.
EDM extension
Conclusion

The main issues are
• express constraints in your specific EDM profile,
• validate data according to these constraints,
• provide data in different flavors (DM2E-EDM, DDB-EDM, …)
• access to these data.

We want to be able to state:
• … how values are represented exactly,
• … which values are mandatory,
• … which values are mutually exclusive,
• … user guidelines of our EDM profile
• … and much more.

To optimize the quality of our EDM-metadata !!!!!!!