The Europeana Data Model
Principles, community and innovation
Antoine Isaac, Nuno Freire
Europeana
2020 DCMI Virtual - Best Practice Presentation
The Europeana Data Model
Principles, community and innovation

Nuno Freire, Antoine Isaac

The Best Practice Day at 2019 Dublin Core Conference
Starting with a few reminders on Europeana
Europeana

60 million digitized objects, from 3,700 institutions in 44 countries
A diverse set of data(sets) and features

- Many different themes and types of object
  Books, newspapers, letters, paintings, maps, photographs, music, fashion, sculpture, 3D objects...
- We must harvest and process metadata that is very heterogeneous
- We ask our providers to give metadata using one data model
- Europeana's aggregation and publication features often evolve
Following established best practices and making it known
Linked Open Data principles

http://vimeo.com/36752317
Data on the Web
Best Practices
Working Group (2014-2016)

- To provide guidance to publishers, promoting the re-use of data;
- To foster trust in the data among developers
- Linked Data, but not only!

https://www.w3.org/2013/dwbp/
BP 15: Reuse vocabularies, preferably standardized ones

- Use terms from shared vocabularies, preferably standardized ones
- Check that classes, properties, [...] used to represent a dataset do not replicate those defined by vocabularies used for other datasets.
  - e.g. using the Linked Open Vocabularies (LOV) repository
- Or if you have to replicate, indicate mappings clearly

[Data on the Web Best Practices](https://www.w3.org/TR/data-on-the-web/) W3C Recommendation
BP 16: Choose the right formalization level

- Accept that (OWL) semantics establish precise specs and can enable automated reasoning but that complex vocabularies require more effort to produce and hamper reuse of data

- Minimize ontological commitment of your vocabulary – or seek to minimize the commitment of others’ vocabularies

- Check that inference does not produce too many statements that are unnecessary for target applications

- Check examples of “softer” specs, e.g. Schema.org or SKOS
Tournoi royal de motos à Londres : changement d'une roue de side-car en marche
Agence de presse Mondial Photo-Presse. 1932, National Library of France
France, Public Domain

Benefiting from community work - and giving back!
EDM: a model to create connected data

Clavecin, Bartolomeo Cristofori
Cite de la Musique,
MIMO - Musical Instruments Museums Online|CC BY-NC-SA
Re-use of vocabularies in EDM

Plus
- Web Annotation
- RDA
- WGS84
- EBUcore
- ccRel
- ODRL
- DOAP
- SVCS
- DCAT
- ADMS

... (sometimes only for one property!)

http://pro.europeana.eu/edm-documentation
Representing full-text/OCR

- A newspapers project for which we had to extend EDM
- We re-used existing patterns from the International Image Interoperability Framework (IIIF) community
  - based on W3C's Web Annotation & Media Fragments
  - community experience was useful, as it's quite complex stuff!
Representing full-text/OCR

- We contributed to the IIIF Newspapers & Text Granularity groups
- We participated to the design of a IIIF extension

1.2 Motivating Use Cases
A number of common workflows can result in annotation sets with discrete levels of text granularity:

- OCR text is frequently available at a range of granularities, from page to character.
- Manual transcription user interfaces may constrain user input to the level of the line, block, or page.
- Transcriptions may be produced without coordinate data and thus have a very coarse level of granularity.

Identification of the level of text granularity in published annotations can facilitate the reuse of their textual highlighting of search result terms in the user interface. Software designed for user interaction and input, such as block-level text to simplify editing.

2. Text Granularity Levels and the textGranularity Property
The textGranularity property identifies the Text Granularity Level of a resource. The value MUST be a single entry from the table below or in the extensions registry.

<table>
<thead>
<tr>
<th>Text Granularity Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>page</td>
<td>A page in a paginated document</td>
</tr>
<tr>
<td>block</td>
<td>An arbitrary region of text</td>
</tr>
<tr>
<td>paragraph</td>
<td>A paragraph</td>
</tr>
<tr>
<td>line</td>
<td>A topographic line</td>
</tr>
<tr>
<td>word</td>
<td>A single word</td>
</tr>
<tr>
<td>glyph</td>
<td>A single glyph or symbol</td>
</tr>
</tbody>
</table>

{ "textGranularity": "line" }
Representing (full-text) rights

● Europeana enforces a controlled system for expressing copyright status and access conditions across the board
  ○ the Europeana Licensing framework

● Available statements come from Creative Commons and RightsStatements.org
Representing more complex rights statements

<edm:WebResource rdf:about="http://www.mimo-db.eu/media/UEDIN/VIDEO/0032195v.mpg">
  <edm:rights rdf:resource="#statement_3000095353971"/>
</edm:WebResource>

<cc:License rdf:about="#statement_3000095353971"/>
  <odrl:inheritFrom rdf:resource="http://rightsstatements.org/vocab/NoC-NC/1.0/"/>
  <cc:deprecatedOn rdf:datatype="http://www.w3.org/2001/XMLSchema-datatype#date">2029-06-01</cc:deprecatedOn>
</cc:License>

● This is an example for a right statements with expiration date

● Re-using bits from the Creative Commons Rights Expression Language (ccREL) and the Open Digital Rights Language (ODRL)
## Representing metadata quality

The [Europeana Publishing Framework](https://www.europeana.eu/portal/) defines a system of metadata quality tiers.

<table>
<thead>
<tr>
<th>TIER</th>
<th>LANGUAGE</th>
<th>ENABLING ELEMENTS*</th>
<th>CONTEXTUAL CLASSES**</th>
</tr>
</thead>
</table>
| A Europeana as a basic search platform.  
  ‘I want to find a specific item that I’m looking for.’ | At least 25% of the provided EDM metadata fields that are relevant have at least 1 language qualified value | At least one enabling element taken from one of the ‘Discovery scenario’ groups | None |
| B Europeana as an exploration platform.  
  ‘I want to browse and explore Europeana even if I’m not sure what I’m looking for.’ | At least 50% of the provided EDM metadata fields that are relevant have at least 1 language qualified value | At least three distinct elements taken from two of the ‘Discovery scenario’ groups | At least one contextual class with all minimum required elements, or link to LOD vocabulary |
| C Europeana as a knowledge platform.  
  ‘I want to search and browse in a more precise way, by named authors, specific subjects or topics.’ | At least 75% of the provided EDM metadata fields that are relevant have at least 1 language qualified value | At least four distinct elements taken from two of the ‘Discovery scenario’ groups | At least two distinct contextual classes with all minimum required elements, or links to LOD vocabularies |

**NB: we have a similar system for content quality but I will not present it today!**
Representing metadata quality

- We use a part of the W3C Data Quality Vocabulary - which we contributed to!
Representing metadata quality

- We keep Europeana-defined elements at a minimum: 3 URIs for (SKOS) concepts that represent our metadata tiers!

- By the way DQV re-uses Web Annotation itself
The acrobatics of metadata modeling
A community sport

- EDM development involves experts from libraries, archives, museums and academics

- Adopting a collaborative, softer form of standardization

Europeana Assembly General Meeting, Rijksmuseum, Amsterdam, 2015

https://pro.europeana.eu/europeana-tech
Lessons worth sharing (again)

- It requires a lot of researching and talking
- Having best practices is useful, yes! And access to resources.
- Flexibility is required:
  - We sometimes changed definitions to avoid semantic overcommitment
  - Re-using is easier when one has a cool-head approach to formal semantics