Introduction to metadata quality – the approach of Europeana Collections

Adina Ciocoiu | 22 November
Introduction

Concerto de Amadores | Columbano Bordalo Pinheiro
1882, Museu do Chiado – Museu Nacional de Arte Contemporânea, Portugal, PDM

Introduction to metadata quality – the approach of Europeana Collections
CC-BY-SA
Short story of Europeana

• April 2005  Jacques Chirac tells EC his big idea
• 2007  EDLnet starts building Europeana
• November 2008  Europeana prototype launched
• Summer 2010  Prototype becomes a service
• Sep 2012  Europeana metadata released as CC0
• 2013  Implementing Europeana Data Model (EDM)
• 2013  #AllezCulture campaign
• May 2015  Europeana Digital Service Infrastructure
• November 2018  10 years anniversary
Europeana achievements

- >3,700 European institutions (galleries, libraries, archives and museums) share their collections online
- >54 millions records available online
- more than 40 languages
- CC0 for metadata
- Standardised metadata - Europeana Data Model (EDM) - huge amount of references to places, agents, concepts, time
- All records include a standardised rights statement
- A network of like-minded heritage and technology professionals
- An open data platform (available for reuse via Europeana API)
We want to build on Europe’s rich heritage and make it easier for people to use, whether for work, for learning or just for fun!

OUR PRIORITIES

We have three goals. We want to make it easier and rewarding for cultural heritage institutions to share high-quality collections with a global audience. We want to scale with partners to reach the markets and audiences who can most benefit from cultural heritage content. And we want people to get excited about cultural heritage whether it’s on Europeana’s websites or elsewhere.

MARKETS

ACADEMIC RESEARCH
CREATIVE INDUSTRIES
EDUCATION
CULTURAL HERITAGE INSTITUTIONS

The possibilities for using open cultural heritage in new apps, products and services are almost limitless. Explore our showcase of apps using the Europeana content and API.

Birdie Memory
Europeana Pro | CC BY-SA

Birdie Memory is an educational game which introduces primary school children (and older) to birds and their unique songs using Europeana openly ...

1 minutes to read

Europeana Strategy homepage
Europeana Pro | CC BY-SA

Introduction to metadata quality – the approach of Europeana Collections
CC-BY-SA
Contributing to third party sites - Wikimedia Commons

Media in category "Europeana 1914-1918"

The following 200 files are in this category, out of 996 total.

(previous page) (next page)

"Andrée".jpg 286 KB
"Argonne".jpg 177 KB
"Attestation d'invalidité" van François Van Goidsenoven, item 1.jpg 994 KB
"Eléonore".jpg 291 KB
"Gefallen für das Vaterland". In Erinnerung an den Landwehrmann Jakob Paul, item 3.jpg 759 KB
"Il faut être à deux pour Souvenir".jpg
01 plaque d'identité
112 Shell case trench art
112 Shell case trench art

Must Read:
10 reasons to open up your digital cultural heritage data
by Nicole McNeill

European 1914-1918 content on Wikimedia Commons
Wikimedia Commons | CC BY-SA

Introduction to metadata quality – the approach of Europeana Collections
CC-BY-SA
Aggregation in Europeana Collections

Introduction to metadata quality – the approach of Europeana Collections

CC-BY-SA
Aggregation in Europeana Collections

- Descriptive and technical metadata
- Thumbnails
  
  *As a rule, content is still served from our data partners*

- Content for specific projects
  - newspapers text and images
  - user-generated content
  (Europeana 1914-1918, Europeana Migration)
Metadata quality prerequisites
From this...

... to Europeana Semantic Elements

europeana:type, europeana:dataProvider, europeana:provider, europeana:isShownAt, europeana:isShownBy, europeana:object, europeana:rights
....to...the Europeana Data Model

- Cross-community re-use of data models
- Semantic Web technology allows mixing them
- Collaborative, softer form of standardisation

Introduction to metadata quality – the approach of Europeana Collections

CC-BY-SA
The core classes are:
- edm:ProvidedCHO - the provided cultural heritage object
- edm:WebResource - the web resource that is the digital representation
- ore:Aggregation - the aggregation that groups the classes together

Main contextual classes include:
- edm:Agent - who
- edm:Place - where
- edm:TimeSpan - when
- skos:Concept - what
- cc:license - access and usage

### Properties for ore:Aggregation

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>edm:aggregatedCHO</td>
<td>✓</td>
</tr>
<tr>
<td>edm:dataProvider</td>
<td>✓</td>
</tr>
<tr>
<td>edm:hasView</td>
<td>✓</td>
</tr>
<tr>
<td>edm:isShownAt</td>
<td>✓</td>
</tr>
<tr>
<td>edm:isShownBy</td>
<td>✓</td>
</tr>
<tr>
<td>edm:object</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Properties for edm:WebResource

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc:creator</td>
<td>✓</td>
</tr>
<tr>
<td>dc:description</td>
<td>✓</td>
</tr>
<tr>
<td>dc:format</td>
<td>✓</td>
</tr>
<tr>
<td>dc:identifier</td>
<td>✓</td>
</tr>
<tr>
<td>dc:language</td>
<td>✓</td>
</tr>
<tr>
<td>dc:relation</td>
<td>✓</td>
</tr>
<tr>
<td>dc:rights</td>
<td>✓</td>
</tr>
<tr>
<td>dc:source</td>
<td>✓</td>
</tr>
<tr>
<td>dc:subject</td>
<td>✓</td>
</tr>
<tr>
<td>dc:title</td>
<td>✓</td>
</tr>
<tr>
<td>dct:creator</td>
<td>✓</td>
</tr>
<tr>
<td>dct:coverage</td>
<td>✓</td>
</tr>
<tr>
<td>dct:creator</td>
<td>✓</td>
</tr>
<tr>
<td>dct:creator</td>
<td>✓</td>
</tr>
<tr>
<td>dct:creator</td>
<td>✓</td>
</tr>
<tr>
<td>dct:creator</td>
<td>✓</td>
</tr>
<tr>
<td>dct:creator</td>
<td>✓</td>
</tr>
<tr>
<td>dct:creator</td>
<td>✓</td>
</tr>
<tr>
<td>dct:creator</td>
<td>✓</td>
</tr>
<tr>
<td>dct:creator</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Contextual Classes

#### Properties for edm:Agent

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>skos:prefLabel</td>
<td>✓</td>
</tr>
<tr>
<td>skos:altLabel</td>
<td>✓</td>
</tr>
<tr>
<td>skos:note</td>
<td>✓</td>
</tr>
</tbody>
</table>

#### Properties for edm:Place

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>skos:prefLabel</td>
<td>✓</td>
</tr>
<tr>
<td>skos:altLabel</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Overview of the properties in each class

- **✓** = Mandatory property
- ** ⇒ ** Blue = at least one of the blue properties should be present (and can be used alongside each other)
- ** ○ ** Red = at least one of the red properties should be present (and can be used alongside each other)
- ** ♦ ** Green = at least one of the green properties should be present (and can be used alongside each other)
- ** + ** = recommended property

Detailed characteristics of all properties are given in the full tables following this summary.

**Core Classes**

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dcterms:isReplacedBy</td>
<td>✓</td>
</tr>
<tr>
<td>dcterms:isRequiredBy</td>
<td>✓</td>
</tr>
<tr>
<td>dcterms:isVersionOf</td>
<td>✓</td>
</tr>
<tr>
<td>dcterms:medium</td>
<td>✓</td>
</tr>
<tr>
<td>dcterms:provenance</td>
<td>✓</td>
</tr>
<tr>
<td>dcterms:references</td>
<td>✓</td>
</tr>
<tr>
<td>dcterms:replaces</td>
<td>✓</td>
</tr>
<tr>
<td>dcterms:requires</td>
<td>✓</td>
</tr>
<tr>
<td>dcterms:spatial</td>
<td>✓</td>
</tr>
<tr>
<td>dcterms:tableOfContents</td>
<td>✓</td>
</tr>
<tr>
<td>dcterms:temporal</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:currentLocation</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:hasMet</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:hasType</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:incorporates</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:isNextInSequence</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:isRelatedTo</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:isRepresentativeOf</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:isSimilarTo</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:isSuccessorOf</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:realizes</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:type</td>
<td>✓</td>
</tr>
<tr>
<td>edmc:isPartOf</td>
<td>✓</td>
</tr>
<tr>
<td>owl:sameAs</td>
<td>✓</td>
</tr>
</tbody>
</table>

---

**Introduction to metadata quality – the approach of Europeana Collections**

CC-BY-SA
Europeana Linked Data strategy

EDM supports contextual resources and offers a base for cross linking

EDM allows data providers to contribute their own vocabularies

EDM contributes to the alignment between domain vocabularies
Europeana Semantic Enrichment

We perform automatic **semantic enrichment** to link source data to reference data in order to....

**Improve the user experience**

- **better ways of searching** and navigating through the collections, **eliminating ambiguity** and clarifying the meaning of descriptions
- **better adaptation to the language** of the user

**Improve the quality and interlinking of data**

- **more context** to the objects
- **better multilingual coverage**
- **contribution to build a web of data** that third parties can use to improve their users' experience
Europeana Linked Data strategy

Vocabularies currently provided to Europeana

Introduction to metadata quality – the approach of Europeana Collections

CC-BY-SA
Introduction to metadata quality – the approach of Europeana Collections
Introduction to metadata quality – the approach of Europeana Collections

CC-BY-SA
The Entity Collection - an example of EDM metadata re-use in Europeana Collections

**Enrichment of Provider’s Data**
- A controlled vocabulary to help identify named references to entities

**Crowdsourcing**
- Objects can be annotated with references to entities
- A controlled vocabulary for client applications

**Europeana Collections Portal**
- **Findability**: users can look for entities, not only records (Entity-Based Search)
- **Understandability**: Entity Pages group and present all assertions about an entity
- **Exploration**: Navigation along relationships and browsing through entities becomes possible

**Republication for Re-use**
- Entities can be republished as an open source to the community
The Entity Collection - an example of EDM metadata re-use in Europeana Collections

http://data.europeana.eu/agent/base/59787

Introduction to metadata quality – the approach of Europeana Collections

CC-BY-SA
EDM - an open and living standard

The model is well-documented and openly shared

Different documentation on various profiles of EDM

Allows the identification of new domain specific requirements while maintaining interoperability

Community joint effort
Building block for...

Metadata Application Profile

The DPLA Metadata Application Profile (MAP) is the basis for how metadata is structured and validated in DPLA, and guides how metadata is stored, serialized, and made available through our API in JSON-LD. The MAP is based on the Europeana Data Model (EDM), and integrates the experience and specific needs for aggregating the metadata of America’s cultural heritage institutions. The current version is 4.0.

Download the DPLA Metadata Application Profile, version 4.0.

New to the DPLA MAP? Read our Introduction to the DPLA Metadata Model to get started.

The previous versions can be accessed below:
- DPLA MAP v5.0
- DPLA MAP v5.1

Neu hier?

Willkommen bei der Deutschen Digitalen Bibliothek!
Europeana Collections: Quality Challenges
General metadata issues... to mention just a few

Confusion between information describing the **object** and information describing its **digital representation**

Absence of mandatory fields

Non meaningful information in descriptive fields

Non unique identifiers

Literal values in fields where a reference is expected

Codes instead of full names given in literal values
information about the object vs its digital representation

edm:ProvidedCHO

dc:format

edm:WebResource

dc:format

Title
Portrait of Pierre Sala from BL Stowe 955, f. 17 | Pierre Sala

Description
Portrait of Pierre Sala, at the end of his Petit Livre d'Amour. Image taken from f. 17 of Petit Livre d'Amour. Written in French and Italian (f. 7v).

People
Creator: Pierre Sala
Contributor: Pierre Sala

Classifications
Type: illuminated manuscripts, Catalogue of Illuminated Manuscripts
Subject:
British Library Stowe Collection, BL Stowe 955, f. 17,
Pierre Sala's handwriting in gold on purple (f. 16v in mirror-script)

Properties
Size: 130 x 95 (85 x 70)
Format: JPEG
Language: fre; ita
## Unknown coordinates

Table: Newfoundland: Letter from John Alcock to his sister dated 12th June, 1919 and carried in the cover above.

<table>
<thead>
<tr>
<th>Title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newfoundland: Letter from John Alcock to his sister dated 12th June, 1919 and carried in the cover above.</td>
<td>In 1913 the Daily Mail newspaper offered a prize of £10,000 for the first non-stop crossing of the Atlantic by a heavier than air machine. The 1914-18 War interrupted the attempts and in 1918 the Daily Mail's offer was repeated. Attempts were made from Newfoundland in 1919 and on 14th June the first successful flight with Captain John Alcock and navigator Lt. Arthur Whitten-Brown was made. The image shows a letter in Alcock's hand to his sister in Manchester carried on that first successful flight reading &quot;...This letter will travel with me in the official mail bag, the first mail to be carried over the Atlantic...&quot;</td>
</tr>
<tr>
<td>Location</td>
<td>Location: 0,0, Newfoundland</td>
</tr>
</tbody>
</table>
Language issues

*dc:language provided for non linguistic content*

<table>
<thead>
<tr>
<th>Description</th>
<th>... de adăugat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifications</td>
<td>Type: imagine digitală</td>
</tr>
<tr>
<td>Properties</td>
<td>Language: ro</td>
</tr>
</tbody>
</table>
Digital objects

*Missing direct links*

*Broken links*
Digital objects

Low resolution

Watermarks
Debatable license attribution (country specific)

<table>
<thead>
<tr>
<th>Classifications</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type: Manuscript</td>
<td>Date: 1600 - 1699</td>
</tr>
<tr>
<td>Subject: Manuscript</td>
<td></td>
</tr>
<tr>
<td>Properties</td>
<td></td>
</tr>
<tr>
<td>Language: ita</td>
<td></td>
</tr>
</tbody>
</table>
Our approach to metadata quality
Metadata Quality - what do we do

Dataset analysis

Data Quality Planning

Data quality conversations

Developing Data Assessment Tools

Semantic Enrichment: fetching information from external controlled vocabularies when links are provided

Mezei kocsiút | Vaszary János
Rippl-Rónai Megyei Hatókörű Városi Múzeum
Hungary, CC-BY
Metadata Quality – what can *data partners* do

More meaningful metadata

Use a consistent terminology in fields enabling Semantic Enrichment

Provide links to dereferenceable LOD vocabularies (*Getty AAT, Iconclass, Geonames, VIAF, MIMO, Library of Congress WWI terms, UNESCO Thesaurus*)

Use of language attributes for literal values

More spatial information

Normalisation of dates
....cause when they do....

Provided datasets/vocabularies

ConceptScheme: http://www.mimo-db.eu/instrumentsKeywords

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

skos:prefLabel
- Harpsichord (und)
- Clavicembalo (es)
- 羽管键琴 (zh)
- Klavescyn (pl)
- Clavicembal (ca)
- Cembalo (de)
- Harpsichord (en)
- Clavichin (fr)
- Clavcimbalo (it)
- Klavecimbel (nl)
- Cembalo (sv)

skos:exactMatch
- http://www.mimo-db.eu/HornbostelAndSachs/64/65

skos:closeMatch
- http://dbpedia.org/resource/Harpichord

Iconclass

Outline: Edits - Clipboard

0 Abstract, Non-representational Art
1 Religion and Magic
2 Nature
3 Human Being, Man in General

4 Society, Civilization, Culture

6 art
6c the arts; artists
6c7 music
6c75 making music; musician with instrument
6c753 more than one musician with instrument
6c753(+) (+ variant)
6c753(+)2 (+ artist at work)
6c753(+)3 (+ artist in non-work situation)
6c753(+)4 (4 types of art)
...we and our users are happy...

Introduction to metadata quality – the approach of Europeana Collections

CC-BY-SA
Supporting documentation for quality recommendations
Metadata quality is controlled by a set of processes which ensures that cultural heritage objects can be identified, discovered and seen in context by audiences, in a manner appropriate to the context in which the data provider created them. Metadata must include information on the potential re-use of cultural heritage objects.

Metadata Quality Task Force Report
Europeana Pro | CC-BY-SA
Europeana Publishing Guide structures the relationship between Europeana and its data providers (rights on metadata and content)

Europeana Publishing Framework defines recommendations for high-quality content and encourage institutions to open their data (EDM mandatory elements, accurate rights labels, direct links to object/hires images)
Task Forces & Working Groups

**Creation and Governance of EDM mappings, profiles and extensions**

The Task Force aims to set principles of a future governance model for EDM mappings, profiles and extensions.

**Working Group**

**Copyright Policy Working Group**

The aim of this working group is to work together with those actively engaged in copyright and make sure we align everyone's interests wherever possible.

**Data Quality Committee**

Quality is a key priority for our whole community! The Data Quality Committee works to address key data quality issues over time. The Data Quality Committee is a EuropeanaTech Working Group.

**Europeana Network Association – Task Forces**

Europeana Pro | CC BY-SA