The Europeana Data Model and Europeana Libraries

Robina Clayphan

Overview

1. How delighted I am to be here

2. The Europeana Data Model
   • What is it for?
   • What does it look like?
   • First implementation
     • Identification of classes and properties (EDM and other)

3. Europeana Libraries
   • Applying library data to EDM
How delighted am I to be here?
Europeana and EDM

→ Europeana is a service that aggregates data from the cultural heritage sector in Europe.
  • libraries, museums, archives and audio-visual archives
  • [http://www.europeana.eu/](http://www.europeana.eu/)
→ Provides a portal for users to access that data
  • Metadata, previews and links to source
→ Will make the metadata freely available for anyone to re-use
  • under Creative Commons Zero (CC0) -public domain dedication
→ Enriches data, provides tools
  • Link to data from other sites, embed on wikipedia, API
→ Makes data available as Linked Open Data (experimentally)
  • [http://data.europeana.eu/](http://data.europeana.eu/)
EDM Requirements

1. Distinguish between the real world object (painting, book, program) and its digital representation

2. And the object and the metadata record describing the object.

3. Allow multiple records for the same object,
   - containing potentially contradictory statements about an object

4. Support for objects that are composed of other objects

5. Standard metadata format that can be specialized

6. Standard vocabulary format that can be specialized

7. EDM should be based on existing standards
EDM basics

OAI ORE for organization of metadata about an object
➡ Requirements 1 to 4 - distinguishing between objects

Dublin Core for descriptive metadata representation
➡ Requirement 5

SKOS for conceptual vocabulary representation
➡ Requirement 6

OAI ORE, Dublin Core and SKOS together fulfil
➡ Requirement 7
EDM property taxonomy (without ESE properties)
<table>
<thead>
<tr>
<th>Réponse n° 1</th>
<th>Domaine</th>
<th>peinture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type d'objet</td>
<td>tableau</td>
<td></td>
</tr>
<tr>
<td>Titre</td>
<td>PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE</td>
<td></td>
</tr>
<tr>
<td>Auteur/exécutant</td>
<td>LEONARDO DI SER PIERO DA VINCI ; VINCI Léonard de (dit)</td>
<td></td>
</tr>
<tr>
<td>Précision auteur/exécutant</td>
<td>Vinci, 1452 ; Amboise, 1519</td>
<td></td>
</tr>
<tr>
<td>École</td>
<td>Italie</td>
<td></td>
</tr>
<tr>
<td>Période création/exécution</td>
<td>1er quart 16e siècle</td>
<td></td>
</tr>
<tr>
<td>Millésime création/exécution</td>
<td>1503 entre ; 1506 et</td>
<td></td>
</tr>
<tr>
<td>Genèse</td>
<td>ouvre en rapport ; reproduit en gravure</td>
<td></td>
</tr>
<tr>
<td>Historique</td>
<td>commandé par le florentin Francesco del Giocondo, époux de Mona Lisa entre 1503 et 1506 ; nombreuses copies dont une conservée au Louvre ; gravé par Fauchery, par Filhol, par Landon</td>
<td></td>
</tr>
<tr>
<td>Matériaux/techniques</td>
<td>peinture à l'huile ; bois</td>
<td></td>
</tr>
<tr>
<td>Mesures</td>
<td>77 H ; 53 L</td>
<td></td>
</tr>
<tr>
<td>Sujet représenté</td>
<td>portrait (Mona Lisa, femme, à mi-corps, de trois-quarts, assis, accoudé, loggia, Italien) ; fond de paysage (montagne, rocher, cours d'eau, pont, plaine, route)</td>
<td></td>
</tr>
<tr>
<td>Date sujet représenté</td>
<td>1479-1528</td>
<td></td>
</tr>
<tr>
<td>Lieu de conservation</td>
<td>Paris ; musée du Louvre département des Peintures</td>
<td></td>
</tr>
<tr>
<td>Statut juridique</td>
<td>propriété de l'Etat ; musée du Louvre département des Peintures</td>
<td></td>
</tr>
<tr>
<td>Anciennes appartances</td>
<td>François 1er ; Couronne de France</td>
<td></td>
</tr>
<tr>
<td>Numéro d'inventaire</td>
<td>INV 779</td>
<td></td>
</tr>
<tr>
<td>Commentaires</td>
<td>légère diminution du tableau sur les côtés (environ 7 mm) ; acheté vraisemblablement vers 1519, après la mort de l'artiste</td>
<td></td>
</tr>
<tr>
<td>Bibliographie</td>
<td>HEYDENRICH 6 ; OTTINO DELLA CHIESA 31 ; VILLOT I 484 ; HAUTECOEUR 1601 ; C.S.I. 1981, P 192</td>
<td></td>
</tr>
<tr>
<td>Copyright notice</td>
<td>© Musée du Louvre, © Direction des Musées de France, 1999</td>
<td></td>
</tr>
<tr>
<td>Crédits photographiques</td>
<td>© Réunion des musées nationaux ; © Hervé Lewandowski ; © Thierry Le Mage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>commande reproduction et/ou conditions d'utilisation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>renseignements sur le musée</td>
<td></td>
</tr>
<tr>
<td></td>
<td>000PE025604</td>
<td></td>
</tr>
</tbody>
</table>
Example 2 from the Louvre

Leonardo di ser Piero DA VINCI, dit Léonard de Vinci
Vinci, 1452 - Amboise, 1519

Portrait de Lisa Gherardini, épouse de Francesco del Giocondo,
dite Monna Lisa, la Gioconda ou la Joconde
Vers 1503 - 1506
Peint à Florence
Bois (peuplier)
H. : 0,77 m. ; L. : 0,53 m.

Acquis par François Ier en 1518
Département des Peintures

INV. 779
Aggregation organizes data of a provider

Provenance of representations and description

"Direction des musées de France"@fr

edm:dataProvider

ore:Aggregation
ex1:aggregation/000PE025604

digital representation

edm:WebResource
http://www.culture.gouv.fr/Wave/image/joconde/0372/m503604_00-010164_p.jpg

edm:WebResource
http://www.culture.gouv.fr/Wave/image/joconde/0372/m503604_00-010401_p.jpg

edm:ProvidedCHO
ex1:object/000PE025604

Provided cultural heritage object
Proxy - context-specific metadata for an object

context-specific “view” on object
One provider = one aggregation + one proxy
Moving to implementation

Implementation will be an incremental process

→ Experiment with what is doable
→ Take partners with us
→ Start with a subset of the classes and properties in the model

Three core classes
→ edm:ProvidedCHO, edm:WebResource, edm:Aggregation

Four contextual classes
→ edm:Agent, edm:Place, edm:Timespan, skos:Concept

Europeana enrichment
→ Use of proxies
Moving to implementation

- Allocation of all existing properties to the classes
  - ESE elements and new EDM elements
- Creation of default mapping from existing ESE to EDM
- Working templates created listing these
  - Giving characteristics of each property
    - Obligation and type of value
- Identification of properties from other namespaces for the contextual entities
  - Mostly these did not already exist in ESE, some had been created in EDM

- [http://europeanalabs.eu/wiki/EDMOBJECTTEMPLATESPROVIDERS](http://europeanalabs.eu/wiki/EDMOBJECTTEMPLATESPROVIDERS)
For introduction and general documentation on this page, see http://europeanalabs.eu/wiki/EDMXMLSchema

The corresponding XML schema can be accessed at http://europeanalabs.eu/svn/europeanalabs/trunk/ROOT/src/main/webapp/schemas/edm/

Note: please be aware that some elements mentioned here are NOT be implemented in the first production version of EDM. See Dev Team doc

Note: in the following all resources have identifiers, which are left implicit in the documentation. These identifiers could be "external" identifiers, e.g. HTTP URIs.

Core EDM resources

edm:ProvidedCHO

Quite often this may be a resource of type edm:PhysicalThing as well. This should be indicated using a specific rdf:type statement.

All descriptive metadata for the ingested object will be represented attached to the proxies in the Europeana information space. Only the owl:sameAs, meant to enable providers to point to their own (linked data) representation of the object (e.g., for the Swedish aggregator’s linked data or the Nat. Lib of Hungary), would remain attached to the central CHO node.

Info sources:

• ESE: EDM mapping rules give a distribution of ESE properties between proxies and aggregations

| owl:sameAs | reference | min 0, max unbounded |
| rdf:type   | reference | min 0, max unbounded |
| dcterms:contributor | literal or reference | min 0, max unbounded |
| dcterms:coverage   | literal or reference | min 0, max unbounded |
| dcterms:creator    | literal or reference | min 0, max unbounded |
| dcterms:date       | literal or reference | min 0, max unbounded |
Classes for contextual resources

edm:Agent

Info sources:

- some relevant datasets, with their own schemas or extensions: VIAF, Amsterdam Museum persons, DNB authority files as linked data, constructs using data.bnf.fr
- vocabularies: FOAF, RDA group 2 elements and relations, MADS/RDF

Properties that can be applied:

<table>
<thead>
<tr>
<th>Property</th>
<th>Type</th>
<th>Min 0, Max 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>edm:hasPresentAt</td>
<td>reference</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>edm:hasMet</td>
<td>reference</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>edm:isRelatedTo (for generic relations to other agents, especially)</td>
<td>reference</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>skos:prefLabel</td>
<td>literal</td>
<td>min 0, max 1 per lang tag</td>
</tr>
<tr>
<td>skos:label</td>
<td>literal</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>skos:note (for e.g., biographical notes)</td>
<td>literal</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>owl:sameAs</td>
<td>reference (of an Agent)</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>foaf:name</td>
<td>literal</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>dc:date</td>
<td>reference</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>dc:identifier</td>
<td>reference</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>rdaGr2:dateOfBirth</td>
<td>literal</td>
<td>min 0, max 1</td>
</tr>
<tr>
<td>rdaGr2:dateOfDeath</td>
<td>literal</td>
<td>min 0, max 1</td>
</tr>
<tr>
<td>rdaGr2:dateOfEstablishment</td>
<td>literal</td>
<td>min 0, max 1</td>
</tr>
<tr>
<td>rdaGr2:dateOfTermination</td>
<td>literal</td>
<td>min 0, max 1</td>
</tr>
<tr>
<td>rdaGr2:gender</td>
<td>literal</td>
<td>min 0, max 1</td>
</tr>
<tr>
<td>rdaGr2:professionOrOccupation</td>
<td>literal or reference</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>rdaGr2:biographicalInformation</td>
<td>literal</td>
<td>min 0, max unbounded</td>
</tr>
<tr>
<td>edm:begin</td>
<td>literal</td>
<td>min 0, max 1</td>
</tr>
<tr>
<td>edm:end</td>
<td>literal</td>
<td>min 0, max 1</td>
</tr>
</tbody>
</table>

General "begin" and "end" properties are being used to indicate start date and end date.
Classes and properties for the first implementation

http://pro.europeana.eu/edm-documentation

→ Europeana Data Model Mapping Guidelines
  • About the first implementation
  • With example data mapping

→ Europeana Data Model Definition
  • Full specification

→ Europeana Data Model Primer
  • The story of EDM

→ Other documents relating to EDM
  • Case studies showing edm data from different providers
The Europeana Libraries Project and EDM
Applying library data to EDM

→ The European Library (TEL)
  • Aggregates metadata from the national libraries of Europe
  • Positioned to be the library domain aggregator for Europeana

→ The Europeana Libraries Project (EL)
  • A project to bring research libraries into TEL and Europeana
    • Digital material only
  • WP5 – aligning library metadata to EDM
WP5 – library data to EDM

→ Metadata Working Group
  • National libraries and research libraries

→ Sample data (MARC21, UNIMARC, MODS, DC)

→ Focussed on monographs, multi-volume works and serials
  • Other materials tackled later

→ Report on findings D5.1 EDM for Libraries
  • http://www.europeana-libraries.eu/outcomes

→ A validation process is taking place and the report will be revised
  • Feedback from Europeana
  • Results of validation exercise
Does the model work for library material?

This model is really simple and flexible and has the following advantages:

→ It can be used for both digitized and born digital objects.
→ It can be easily expanded for further entities i.e the contextual resources
  • edm:Event, edm:Agent, edm:Place, edm:PhysicalThing, edm:TimeSpan, and skos:Concept
→ It can be used to describe the complex structure of multi-volume works.
→ It can be used to model the hierarchical and sequential structures of serials
Monograph diagram

edm:WebResource

edm:hasView

ore:Aggregation

edm:ProvidedCHO

edm:aggregatedCHO
Does it work for serials and multi-volume works?

Serials have a structure with
→ hierarchical relationships (dcterms:isPartOf)
  • Article, issue, volume, journal title levels
→ sequential relationships (edm:isNextInSequence)
  • Volume 1, volume 2 etc
→ The structure is not the same across all providers

Multi-volume works can have (in addition) a complex nested structure
  • Complete works of Shakespeare including the three separate volumes of Henry VI.
Serial - article

edm:WebResource
http://www.digizeitschriften.de/dms/img/?PPN=PPN385489110_0024&DMDID=DMDLOG_0203

edm:hasView

edm:Agent
http://www.digizeitschriften.de/

edm:ProvidedCHO
http://digizeit/PPN129542113_0024/i21/p251

edm:aggregatedCHO

dcterms:isPartOf

"Naturwissenschaftliche Rundschau, v. 24, l. 21"

"article"

"Die Dinosaurier"

"Arldt, Th."

dct:creator

dct:title

dct:type
Serial hierarchy

- edm:WebResource
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0203

- edm:ProviderCHO
  - http://digizeitschriften.de/PPN129542113_0024/281

- edm:Agent
  - http://www.digiztschriften.de/

- ore:Aggregation
  - http://www.digiztschriften.de/PPN185489110_0024/DMLOG_0202

- edm:hasView
  - http://resolver.sub.uni-goettingen.de/publ/PPN1854893116

- edm:hasView
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0202

- edm:Agent
  - http://www.digiztschriften.de/

- edm:WebResource
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0202

- edm:Agent
  - http://www.digiztschriften.de/

- edm:WebResource
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0202

- edm:Agent
  - http://www.digiztschriften.de/

- edm:WebResource
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0202

- edm:Agent
  - http://www.digiztschriften.de/

- edm:WebResource
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0202

- edm:Agent
  - http://www.digiztschriften.de/

- edm:WebResource
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0202

- edm:Agent
  - http://www.digiztschriften.de/

- edm:WebResource
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0202

- edm:Agent
  - http://www.digiztschriften.de/

- edm:WebResource
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0202

- edm:Agent
  - http://www.digiztschriften.de/

- edm:WebResource
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0202

- edm:Agent
  - http://www.digiztschriften.de/

- edm:WebResource
  - http://www.digiztschriften.de/dms/img/PPN=PPN185489110_0024&MODID=DMLOG_0202
Issues encountered - 1

Defining the CHO more precisely (edition vs item):

→ In the context of published textual resources (library material) this is the edition, i.e. the entirety of all identical copies of a text produced in the same process of publishing.

→ Any information about the digital representation(s) of an edition – whether it is born digital or digitised - can be provided using the class edm:WebResource.

→ But – how to model rare books?
Possible model for rare books
Issues encountered - 2

→ dcterms, edm terms and non-literal values
  • In many places both the specifications require non-literal values
  • These don’t necessarily exist in the data (ESE data)
  • For the time being we will have to accept this anomaly

→ Use the event class for e.g. publication data
  • Event class not included at first, (would appropriate values be available?)

→ Non-inclusion of desired properties in first implementation
  • Two pronged approach – pragmatic for first implementation
  • Optimal proposal for the future.
Thank you

Robina Clayphan

robina.clayphan@kb.nl
OAI ORE
Open Archives Initiative Object Reuse & Exchange
Specification:

http://www.openarchives.org/ore/1.0/toc.html

Specified with an RDF model

Four key notions (RDF classes)

→ **Object**: the book/painting/program being described

→ **Aggregation**: organizes object information from a particular provider (museum, archive, library)

→ **Digital representation**: some digital form of the object with a Web address

→ **Proxy**: the metadata record for the object