The Europeana Linked Open Data Server

Nicola Aloia, Cesare Concordia, Carlo Meghini
Istituto di Scienza e Tecnologie dell’Informazione – CNR
Pisa
Europeana

• Started in 2007
  – Cluster of projects funded by EU
• 26m (Feb 2013) metadata records (22m+ metadata records as CC0)
  – Paintings, maps, drawings, photographs, music, books, newspapers, journals, diaries…
• 31 languages
• 2200 data providers
• Based in National Library of Netherlands
Europeana & Aggregators
Europeana portal

Roma 1971, edilizia

Description: carta ; b/n ; positivo ; note: Pubblicata su «Rassegna sindacale», n. 204, 1971.; descrizione: Palazzi in costruzione nella periferia romana.


Date of creation: 19710000-19710000; Roma

Type: item

Subject: edilizia - Roma; Roma - periferia

Identifier: http://hdl.handle.net/10978/60082B40-81DF-42D8-83A2-4A58C9636752

Relation: http://www.peoplesheritage.eu

Is part of: http://hdl.handle.net/10978/F3E3752C-FA19-41F9-B8BC-D6444DE7EB6E

Language: ita

Provenance: CGIL

Data provider: Confederazione Generale Italiana del Lavoro

Provider: HOPE - Heritage of the People’s Europe

Providing country: Italy

Auto-generated tags •
Europeana API

**API function:**

<table>
<thead>
<tr>
<th>Search parameters</th>
<th>Request URL</th>
</tr>
</thead>
</table>

**HTTP status code**

200

**JSON response**

```
{
  "apikey": "xxxxxxxx",
  "action": "search.json",
  "success": true,
  "requestNumber": 2444,
  "itemsCount": 12,
  "totalResults": 194,
  "items": [
    {
      "id": "919003CAD63004EA9683DC94835FC81086968294CC323289",
      "provider": [
        "Archives Portal Europe",
        "Archives Portal Europe"
      ],
      "europeanaCompleteness": 5,
      "language": [
        "nl"
      ]
    }
  ]
}
```
Linked Data & Europeana

• Europeana provides integrated access to digital objects of the cultural heritage organizations of all the members of the European Union

• Publishing datasets as Linked Data (LD) can help Europeana to distribute its data and so attract new users and new providers

• Linked Data enables the use of digital representations of cultural artifacts for generating knowledge
Linked Data & Europeana

- Europeana Data Model (EDM) is a suitable data model for publishing Europeana datasets as Linked Data
  - EDM is built with RDF in mind (same metamodel)
  - EDM uses HTTP URIS as resource identifiers
  - EDM re-uses identifiers from authorities for the main entities in metadata (people, places, subjects, etc.), thereby linking to their databases and to the databases of the institutions who do the same
  - EDM re-uses classes and properties from well-known vocabularies in cultural heritage in order to overcome interoperability barriers
Linked Data & Europeana

• Distributing the Europeana datasets as Linked Open Data (LOD) requires:

  – to define an agreement with every data provider to publish their data as open data

  – to process the Europeana dataset to obtain RDF descriptions

  – to build a LD publishing framework
Europeana LD server overall architecture
Europeana LD Server: overall approach

- Convert Europeana metadata dataset into RDF/XML EDM metadata records
  - XML stylesheets, using XSLT 1.0

- Enrich selected metadata fields using controlled vocabularies
  - Annocultur tool (developed at Europeana foundation)

- Link to existing LOD services maintained by Europeana partners (National Library of Hungary, Swedish culture aggregator…)

- Publish LD datasets
  - File download, RDF triple store
Metadata mapping

• Records in dataset were formatted using ESE (unqualified DC + specific fields)
  – Main issues: flat model, values as string, in the same metadata record values belonging to different entities

• EDM designed to open the Europeana information space
  – Key features: distinguish ‘real word objects’ from their digital representations, allow several description for one item, support for complex item representation, re-use and links to existing reference vocabulary reference
  – EDM solves ESE shortcomings

• The mapping workflow:
  – create the EDM records
  – set dereferencable URI id to record’s entities
## ESE record example

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc:identifier</td>
<td>02-Lemba-Lakkous-Lady.tif</td>
</tr>
<tr>
<td>dc:title</td>
<td>Stone figurine</td>
</tr>
<tr>
<td>dc:subject</td>
<td>Cyprus--Antiquities</td>
</tr>
<tr>
<td>dc:description</td>
<td>Stone vessel from Choirokoitia with embossed decoration.</td>
</tr>
<tr>
<td>xml:lang=&quot;en&quot;</td>
<td></td>
</tr>
<tr>
<td>dc:publisher</td>
<td>Cyprus Archaeological Museum</td>
</tr>
<tr>
<td>dc:type</td>
<td>Image</td>
</tr>
<tr>
<td>dc:format</td>
<td>image/tiff</td>
</tr>
<tr>
<td>dcterms:temporal</td>
<td>mid 3rd millenium B.C.</td>
</tr>
<tr>
<td>dc:rights</td>
<td>Cyprus Archaeological Museum</td>
</tr>
<tr>
<td>dc:title</td>
<td>Λίθινο ειδώλιο</td>
</tr>
<tr>
<td>dc:description</td>
<td>Λίθινο ειδώλιο από τη θέση Λέμπα ...</td>
</tr>
<tr>
<td>europeana:provider</td>
<td>Cyprus Library</td>
</tr>
<tr>
<td>europeana:type</td>
<td>IMAGE</td>
</tr>
<tr>
<td>europeana:rights</td>
<td><a href="http://creativecommons.org/licenses/by/3.0/nl/">http://creativecommons.org/licenses/by/3.0/nl/</a></td>
</tr>
<tr>
<td>europeana:dataProvider</td>
<td>Cyprus Archaeological Museum</td>
</tr>
<tr>
<td>europeana:uri</td>
<td><a href="http://www.europeana.eu/resolve/record/0000/E2AAA3C6DF09F9FAA6F951FC4C4A9CC80B5D4154">http://www.europeana.eu/resolve/record/0000/E2AAA3C6DF09F9FAA6F951FC4C4A9CC80B5D4154</a></td>
</tr>
<tr>
<td>europeana:country</td>
<td>cyprus</td>
</tr>
<tr>
<td>europeana:language</td>
<td>gr</td>
</tr>
</tbody>
</table>
EDM example
Europeana EDM record structure

Provider Metadata
xmlns:eulod: "http://data.europeana.eu/"
xmlns:ens = "http://www.europeana.eu/schemas/edm/"
xmlns:ore = "http://www.openarchives.org/ore/terms/"
Mapping: lessons learned

• Europeana URIs identify records rather than resources representing real-world objects
• It is complex to identify the target EDM resource for a given property
  – providers could have not followed Europeana guidelines
• Complex network of resources not easy to ‘consume’ for linked data practitioners
  – We are asking feedback from data consumers
• Enhance navigability between resources
  – Advanced RDF store configuration, new properties
Metadata enrichment

- Metadata enrichment consists of
  
  – replacing values of selected metadata fields with URIs of resources from controlled vocabularies (E.g.: ens:country="Cyprus" becomes ens:country=http://www.geonames.org/146669/)

  – adding meta-level information about the data published (provenance and licensing information)
# Metadata enrichment

<table>
<thead>
<tr>
<th>Entity</th>
<th>Metadata fields</th>
<th>Controlled source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Places</td>
<td>dcterms:spatial, dc:coverage</td>
<td>Geonames</td>
</tr>
<tr>
<td>Concepts (topics)</td>
<td>dc:subject, dc:type</td>
<td>GEMET, DBPedia</td>
</tr>
<tr>
<td>Agents</td>
<td>dc:creator, dc:contributor</td>
<td>DBPedia</td>
</tr>
<tr>
<td>Time</td>
<td>dc:date, dc:coverage,</td>
<td>Semium</td>
</tr>
<tr>
<td></td>
<td>dcterms:temporal, edm:year</td>
<td></td>
</tr>
</tbody>
</table>
LD server implementing architecture
Europeana LD Server: data publishing

• Implemented by a Web Server and by a library of Java servlets

• The Web Server receives a request and redirect it to

  – the download area if a dump file is requested,

  – the servlets library if, instead, a resource is requested.
Europeana LD Server: data publishing

• The servlets implement the 303 URIs dereference strategy

• The implementation algorithm is based on the HTTP server-driven content negotiation mechanism, which enables HTTP clients and servers to negotiate a possible response to a specific request.

  – HTTP “Accept” header
Europeana LD server: URI dereferencing example

Head: /proxy/europeana0947A266EDC70E8C9AE61
Accept: application/rdf+xml
User-Agent: vapour.sorceforge.net

303
Location: http://data.europeana.eu/data/proxy/europeana0947EDC70E8C9AE61

Head: /data/proxy/europeana0947EDC70E8C9AE61
Accept: application/rdf+xml
User-Agent: vapour.sorceforge.net

200
Content-type: application/rdf+xml; charset=UTF8
Vary: Accept
Europeana LOD server

• The Europeana Linked Open Data server publishes 22m+ records
  – Records belonging to providers, who want to make their data available on the web

• The LOD server is separated from the Europeana production server
  – http://data.europeana.eu
Linked Open Data - data.europæana.eu

Linked Open Data is a way of publishing structured data that allows metadata to be connected and enriched, so that different representations of the same content can be found, and links made between related resources.

The metadata for all the objects in the Europeana portal is open, in that it is all licensed under the CC0 Public Domain Dedication under the terms of the Data Exchange Agreement (DEA), and can be freely downloaded via the API.

In October 2012, a large subset of this data was transformed into linked data and made available from data.europæana.eu. The data is represented in the Europeana Data Model (EDM). For more information, see our datasets page.

data.europæana.eu started as an experimental pilot in February 2012 with a small number of data providers who committed at an early stage to Europeana's initiative of promoting more open data. The current version of the pilot contains metadata on 20 million texts, images, videos and sounds gathered by Europeana. They cover a great variety of heritage objects, such as this slovenian version of "O Sole Mio" from the National Library of Slovenia, or Neil Robson's memories of the herring business from the Tyne and Wear Archives & Museums.
Europeana SPARQL endpoint (experimental)

Welcome on the SPARQL end-point of data.europeana.eu!

data.europeana.eu currently contains open metadata on 20 million cultural items, and is currently structured using the Europeana Data Model (EDM) and the terms of the Creative Commons CC0 public domain dedication. Please take the time to check out the list of collections currently included in the SPARQL endpoint.

The terms of use and external data sources appearing at data.europeana.eu are described in the catalogue. Sample queries are available on the sparql page.

Repository overview

Engine: OWLIM SE
Inference ruleset: owl-horst-optimized
Number of entities: 265,799,020
Conclusions & acknowledgements

- Distribute the whole Europeana dataset
  - Agreements with content providers

- Challenges:
  - Licensing: 64% (June 2013) of metadata records does not have clear info about content license
  - Improve metadata record quality
  - Optimizing data for reuse
  - Improve the LOD server performances

- The ESE2EDM mapping approach have been designed by Bernhard Haslhofer and Antoine Isaac