Europeana Creative. EDM Endpoint. Custom Views.

Nikola Ikonomov, Boyan Simeonov, Jana Parvanova, Vladimir Alexiev

Ontotext AD, 47A Tsarigradsko Shosse, Fl.4, 1504, Sofia, BG

Abstract. The paper discusses the Europeana Creative project which aims to facilitate re-use of cultural heritage metadata and content by the creative industries. The paper focuses on the contribution of Ontotext to the project activities. The Europeana Data Model (EDM) is further discussed as a new proposal for structuring the data that Europeana will ingest, manage and publish. The advantages of using EDM instead of the current ESE metadata set are highlighted. Finally, Ontotext's EDM Endpoint is presented, based on OWLIM semantic repository and SPARQL query language. A user-friendly RDF view is presented in order to illustrate the possibilities of Forest - an extensible modular user interface framework for creating linked data and semantic web applications.

Keywords: Europeana, EDM, ESE, Semantic Web, RDF, SKOS, URI, OWLIM, Semantic Repository, SPARQL, Query, Endpoint.

1. Introduction

Europeana is Europe's multilingual digital library, museum and archive. It gives people free access to millions of books, paintings, films, sounds, museum objects and archives that have been digitised throughout Europe[1].

Europeana provides access to heterogeneous content from a wide variety of cultural institutions. The digital objects in Europeana are not hosted in a central database, but remain with the cultural institutions and are stored on their networks. Europeana is in fact a metadata sharing service which collects descriptive information about each digital object. Users search this huge database and once they find what they are looking for, they can get access to the full content of the object by visiting the original site of the cultural institution where it is being preserved.

Different types of cultural heritage organisations – libraries, museums, archives and audio-visual collections – catalogue their content using different structures and domain specific standards. Approaches also vary in different countries. To make the information in Europeana searchable, it has to be mapped to a single common standard, known as the Europeana Semantic Elements (ESE). This metadata standard at present takes a lowest common denominator approach to the integration of different types of heterogeneous digital content. However, the introduction of a richer metadata standard, the Europeana Data Model (EDM), will help to give users more and better information[2].

---

[1]http://pro.europeana.eu/documents/900548/1229554/Virtual+Exhibitions+-
+Press+Release+Europeana+Version.doc

[2] http://en.wikipedia.org/wiki/Europeana

2. Europeana Creative

Currently, Europeana supports only search and display of metadata with no access to the content referenced by metadata. The actual Europeana Licensing Framework[3] applies only to metadata. The re-use of cultural heritage resources by creative industries, organisations and individuals is hindered not only by the lack of licensing framework for content, but also by a number of different obstacles –organisational, technical, and cultural.

The Europeana Creative[4] project will allow Europeana to facilitate the re-use of cultural heritage metadata and content by the creative industries. It will establish an Open Laboratory Network as a sustainable environment for experimentation with cultural content and stakeholder engagement and will create Europeana Content Re-use Framework. The Framework will be implemented in co-ordination with Europeana Cloud[5] and will provide a permission management infrastructure allowing content providers to communicate conditions for content re-use scenarios. The project will further implement all needed technical infrastructure and services to resolve the issues of storage, ingestion, legal infrastructure and re-use. Key areas of service enhancement include: online console to experiment with application programming interfaces, scalable storage and caching architecture, taxonomy services to improve metadata lookup&linking, geographic mapping&transformation algorithms to support mobile and augmented reality apps, image similarity services.

The project will create pilot applications in five thematic areas (History Education, Natural History Education, Tourism, Social Networks and Design) and then conduct open innovation challenges to identify, incubate and spin-off viable projects into the commercial sector. It will also undertake an extensive stakeholder engagement campaign promoting the benefits of cultural heritage content re-use to creative industries and to memory institutions.

Europeana Creative is co-funded by a 30-months long European CIP PSP project that started in February 2013 and will finish in July 2015. It involves 26 organizations ranging from museums and libraries, to innovation and creative hubs, and proven technical partners.

Ontotext[6] as project consortium partner is participating in the following tasks:

- Creation of Content Retrieval Services as part of the Content Re-use Framework. The aim of this task is to implement a sophisticated platform for the retrieval of Europeana metadata and related digital content objects in order to enable re-use from applications. The OWLIM[7] platform will be deployed for metadata storage and search, continuing Ontotext on-going activities on the experimental Europeana SPARQL[8] endpoint. This is a core backend component of the architecture: Europeana Creative will make use of this triple store

---

[3]http://pro.europeana.eu/documents/858566/7f14c82a-f76c-4f4f-b8a7-600d2168a73d

[4]http://pro.europeana.eu/web/europeana-creative

[5]http://pro.europeana.eu/web/europeana-cloud

[6] http://www.ontotext.com/

[7]http://www.ontotext.com/owlim

[8]http://www.w3.org/TR/rdf-sparql-query/

as a central data integration repository to provide the backbone of the backend architecture.

- Geographic mapping and transformation algorithms, to be used for the creation and management of cultural routes as required by the Tourism pilot.
- Linking to external web resources (e.g. Freebase, DBpedia, Wikipedia, VIAF, Getty, Geonames).
- Services for geo-referencing of data (based on entity extraction) that are required by the Social Networks Pilot led by HistoryPin. Geo-referenced cultural objects can also be used for location-based services.

3. The Europeana Data Model (EDM)

The Europeana Data Model (EDM) is a new proposal for structuring the data that Europeana will be ingesting, managing and publishing. The Europeana Data Model is a major improvement on the Europeana Semantic Elements (ESE), the basic data model that Europeana began life with.

Each of the different heritage sectors represented in Europeana uses different data standards, and ESE reduced these to the lowest common denominator. EDM reverses this reductive approach and is an attempt to transcend the respective information perspectives of the sectors that are represented in Europeana – the museums, archives, audio-visual collections and libraries. EDM is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate the range and richness of particular community standards such as LIDO[9] for museums, EAD[10] for archives or METS[11] for digital libraries.

EDM will replace ESE which currently underlies the data space of Europeana. EDM offers greater expressivity and flexibility and allows for richer and truly semantic representations of the millions of objects from all kinds of cultural heritage communities in Europeana. In comparison with prior data models EDM realizes a very high level of abstraction. It is the most radical generalization of metadata properties in the cultural heritage area so far and it does not bind the representation of ingested metadata to one common schema. EDM carefully integrates well-established ontologies like SKOS[12], Dublin Core[13], and FOAF[14] in order to allow for rich and interoperable descriptions of Europeana objects. As a common top-level ontology it allows for integrating the distinct information perspectives and needs of the various communities providing data to Europeana and to preserve the original richness of community standards like LIDO, CIDOC CRM[15], MARC[16] or EAD. The EDM uses RDFS as it's

---

[9]http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/lido-technical/specification/

[10]http://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-description-ead/encoded-archival-description-ead

[11]http://www.jisc.ac.uk/uploaded_documents/tsw_02-05.pdf

[12]http://www.w3.org/2004/02/skos/

[13]http://dublincore.org/documents/dces/

[14]http://xmlns.com/foaf/spec/

[15]http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf

[16]http://www.loc.gov/marc/

meta-model and URIs to identify structured information about cultural heritage objects. The structural modeling framework for the EDM ontology is provided by the OAI Object Reuse & Exchange (OAI-ORE) specifications. This open architecture of the EDM makes Europeana compatible with the Semantic Web paradigm and enables it to become part of the emerging Linked Open Data community. In fact, the EDM provides a migration path for cultural heritage institutions from their currently mostly closed information architectures to open, linked environments – for the benefit of both these institutions and the WWW community.

The EDM framework will allow different participants to structure their data in a way that suits their original data and their desired functions. Data providers may create simple datasets or more complex ones depending on the structure of their source data. Similarly, Europeana will manipulate the data internally to perform its aggregation and enrichment functions. Basic structures that are likely to be of interest to data providers are shown below on Fig.1 and 2.

EDM has three core classes of resources that result from the package of data provided to Europeana:

   • the "provided cultural heritage object" itself (a painting, a movie, a music score, a book…) (edm:ProvidedCHO)

   • one or more accessible digital representations of this object, some of which will be used as previews (the digital picture of the painting.) (edm:WebResource)

   • an aggregation to represent the result of this provider's activity.
(ore:Aggregation)

Each of them can be subject to corresponding metadata, leading to three core "metadata sections". The EDM standard defines further which properties can be used for each of these three classes.
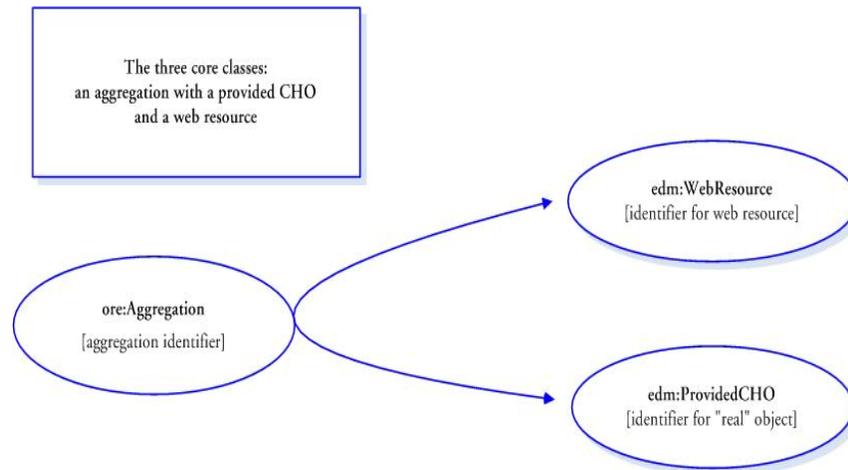

Fig.1 Basic data model
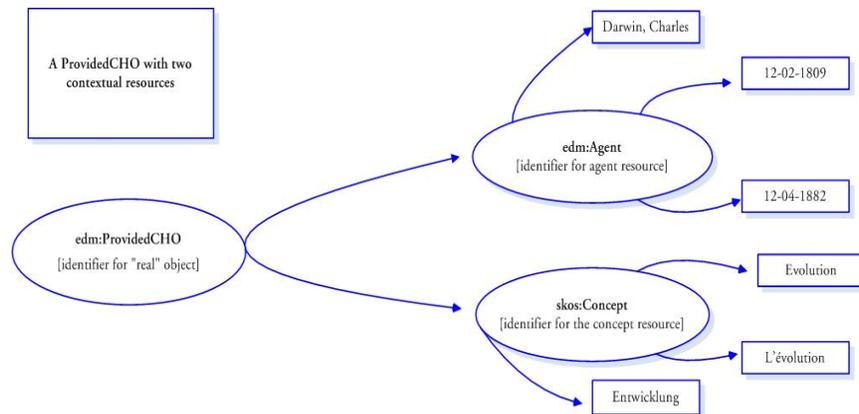
Fig 2 Two contextual classes

4. EDM Endpoint

A few months ago Ontotext launched the Europeana's EDM Endpoint, which allows cultural heritage data from data.europeana.eu to be accessed from applications and to be explored through the User Interface of the Forest framework. An Endpoint is a specific location for accessing a web service using a specific protocol and data format. At its simplest a SPARQL endpoint is a URI to which queries can be sent, and which returns answers to the queries as a response.

4.1. OWLIM

The backbone of the EDM Endpoint is OWLIM - a family of semantic repositories, or RDF database management systems, with the following characteristics:

- native RDF engines, implemented in Java;
- delivering full performance through both Sesame and Jena;
- robust support for the semantics of RDFS, OWL 2 RL and OWL 2 QL;
- best scalability, loading and query evaluation performance.

As announced, one billion explicit statements have been loaded in the OWLIM repository and with inference they stack up to about 4 billion distinct retrievable statements. Ever since Ontotext launched the Endpoint in December 2012, the positive reactions from the community and the number of visitors constantly increased.

OWLIM comes in three editions: OWLIM-Lite, OWLIM-SE (Standard Edition) and OWLIM-Enterprise. The editions share the same inference mechanisms and semantics (rule-compiler, etc.). OWLIM-Lite is a fast, in-memory repository, whereas OWLIM-SE uses a scalable, file-based indexing mechanism. OWLIM-Enterprise is a clustered infrastructure based on OWLIM-SE.

4.2. Forest – the UI to the EDM Endpoint

Ontotext Forest is an extensible modular UI framework for creating linked data and semantic web applications. It is a lightweight web interface that brings together tools for execution of SPARQL queries, visual exploration of RDF triples, full text search and integration of popular third-party data visualization components. It is the front end for linked data web services offered by Ontotext AD.

Forest offers an optimized generic interface for accessing semantic repositories and a foundation for creating custom applications. The look and feel it provides is designed for easy customization and branding.

Forest is an optimal platform for exploring the data in an OWLIM repository because the interface takes advantage of various proprietary capabilities exposed by OWLIM and supports connectivity using both Sesame and the Ontotext ORDI middleware component.

Forest enables simplified access to OWLIM features like RDF rank, query result count, sameAs filtering, preferred URI labels and snippets. Additionally it offers a seamless integration with a number of third-party tools.

In addition to the EDM Endpoint, Forest has been used in Ontotext's two public reason-able view (RDF warehouse) services (FactForge and Linked Life Data), in several commercial projects with companies such as AstraZeneca, and in the new OWLIM Workbench.

FactForge[17] aims to allow users to find resources and facts based on the semantics of the data. It enables users to easily identify resources in the Linking Open Data[18] (LOD) cloud. It provides efficient mechanism to query data from multiple datasets and sources, considering their semantics. FactForge is designed also as a use-case for large-scale reasoning and data integration.

Linked Life Data[19] (LLD) is a data-as-a-service platform that provides access to 25 public biomedical databases through a single access point. The service allows writing of complex data analytical queries, answering complex bioinformatics questions, simply navigate through the information, or export subsets like "all approved drugs and their brand names".

4.3 Custom views

Forest provides a flexible architecture that offers various extension points for development of custom functionalities like user-friendly RDF views. Fig. 3 illustrates a typical custom view with following options:

- All triples of all nodes comprising the CHO are shown. Download can be done in JSON, RDF, Turtle, NTriples.
- A view from the Europeana portal is shown, including links to the original (source).
- A CHO Graph and connections between the nodes are shown. URLs are shortened. Data is compressed into the node.

---

[17]http:// www.factforge.net

[18]http://www.ontotext.com/factforge/linked-data

[19]www.linkedlifedata.com

Fig. 3

Dictionnaire De La Noblesse / 4

Source: http://data.europeana.eu/item/03486/D16390DE3412850D1AA872E6B901F450625C56GF82
Description: Ausgabe : Seconde Edition

Prefixes

a   http://www.w3.org/1999/02/22-rdf-syntax-ns#
cc   http://creativecommons.org/ns#
data   http://data.europeana.eu/
dc   http://purl.org/dc/elements/1.1/
dct   http://purl.org/dc/terms/
edm   http://www.europeana.eu/schemas/edm/
eu   http://www.europeana.eu/
foaf   http://xmlns.com/foaf/0.1/
html   http://www.w3.org/1999/xhtml#vocab
ore   http://www.openarchives.org/ore/terms/
static http://europeanastatic.eu/api/image?uri=

data.proxyprovider/03486...62F82
  a ore:Proxy
    dc:title   Dictionnaire De La Noblesse / 4
    dc:description   Ausgabe : Seconde Edition
    dc:creator   Aubert de la Chesnaye Des Bois, Fran?ois-Alexandre
    dc:subject   Nobility
    dc:subject   CS587
    dc:spatial   929 74
    dc:spatial   France
    edm:type   TEXT
    dc:type   Druck
    dc:date   1772
    dc:date   Paris
    dc:publisher   XVI, 677 S.
    dc:extent   Signatur: 4 Genal. 18-4
    dc:language   fre
    dc:identifier   bvb-id : BV035513043
    dc:identifier   oclc : 645307537
    dc:identifier   urn: urn:nbn:de:bvb:12-bsb10358421-4
    dc:created   (1772)

data.aggregationprovider/03486...62F82
  a ore:Aggregation
    edm:dataProvider Bayerische Staatsbibliothek
    edm:provider   Bayerische Staatsbibliothek
    edm:rights   http://creativecommons.org/publicdomain/zero/1.0/

data.item/03486...62F82
  a edm:ProvidedCHO
    foaf:isPrimaryTopicOf
    foaf:primaryTopic

data.resmap/03486...62F82
  a ore:ResourceMap
    dc:contributor   Bayerische Staatsbibliothek
    dc:publisher   Europeana
    html:license   http://creativecommons.org/publicdomain/zero/1.0/
    cc:useGuidelines http://creativecommons.org/public/docs/metadata-usage-guidelines/

data.aggregation/europeana/03486...62F82
  a edm:EuropeanaAggregation
    dc:creator   Europeana
    edm:country   germany
    edm:language   de
    edm:collectionName CS486_L_DE_BSBMuenchen.xml.gz
    cc:useGuidelines eu:rights/pd-usage-guide/

data.proxy/europeana/03486...62F82
  a ore:Proxy
    edm:hasMet   http://www.w3.org/item/1772
    edm:hasMet   http://semium.org/item/1772
    edm:hasMet   http://sws.geonames.org/2921044/
    edm:hasMet   http://sws.geonames.org/2951232/
    edm:year   1772

eu:resolve/record/03486...62F82
  http://www.mdz-nbn-resolving.de/
  urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb10358421-4

http://www.bsb-muenchen-digital.de/
  ~web1035%2Fbsb10358421%2Fimages
  /bsb10358421/image/bsb10358421_00001.jpg

static: http://daten.digitale-sammlungen.de%2F~web1035%2Fbsb10358421%2Fimages%2Fbsb10358421_00001.jpg&size=FULL_LOO&type=TEXT

edm:landingPage eu:portal/record/03486...F82.html

5. Conclusions

Semantic technologies allow for an unprecedented ease of integration of heterogeneous data sources. Such technologies are already adopted in pharmaceuticals and publishing industries. The cultural heritage domain and Europeana, in particular, can become a useful usecase for the application of semantic technologies through the combination of facts and knowledge from different datasets, thus creating convincing real life use cases demonstrating the benefits of these technologies.

The use of semantic models is not a matter of only keeping up with the semantic web technologies. There is also the side of the users and their needs and expectations. Europeana did a series of user studies over the years; and they demonstrated that the users wish to be able to understand better the relationships within digital objects when they are using the cultural portal. Summarising the findings of one of these international user studies, Dobreva and Chowdhury (2010) highlight that "There was a call for more linking between items to show relationships."

As a leading developer of Core Semantic Technology, Ontotext is especially active in its efforts to make cultural heritage metadata and content in Europeana re-uasable by the creative industries, while at the same time enriching semantically this content. It is no chance that Jill Cousins, Executive Director of the Europeana Foundation in an interview for a major Bulgarian daily describes among the most successful Bulgarian contributions to Europeana Ontotext's establishment of an Europeana semantic Data End-point.

6. References

1. Europeana Data Model Primer,
http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5
2. Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., Van de Sompel, H. (2010), The Europeana Data Model (EDM)
http://conference.ifla.org/past/ifla76/149-doerr-en.pdf
3. Gradmann, S. Knowledge = Information in Context: on the importance of Semantic Contextualisation in Europeana, (2010),
http://pro.europeana.eu/c/document_library/get_file?uuid=cb417911-1ee0-473b-8840-bd7c6e9c93ae&groupId=10602
4. Dobreva, M., Chowdhury, S. (2010). A User-Centric Evaluation of the Europeana Digital Library. In: Chowdhury, G., Khoo, C., Hunter, J. (Eds.) The Role of Digital Libraries in a Time of Global Change. Proc. Of the 12th International Conference on Asia-Pacific Digital Libraries, ICADL 2010, Gold Coast, Australia, June 21-25, 2010, LNCS 6102, 148-157.