10th Italian Research Conference on Digital Libraries, IRCDL 2014

Culturaitalia, the Italian National Content Aggregator in Europeana

Sara Di Giorgio*

The Central Institute for the Union Catalogue of Italian Libraries and Bibliographic Information (ICCU), Viale Castro Pretorio 105, 00185 Rome, Italy

Abstract

CulturaItalia is the Portal of Italian Culture, managed by the Central Institute for the Union Catalogue of Italian Libraries (ICCU) the Italian Ministry of Cultural Heritage, Activities and Tourism (MiBACT); it involves cultural institutions from all sectors and levels (national, regional and local). CulturaItalia, as national aggregator, plays an important role in the development of Europeana, making available cooperative networks and agreements and coordinating technical activities leading to the establishment of Europeana. In line with the Europeana strategy of data re-use, CulturaItalia recently presented a section devoted to open data (http://dati.culturaitalia.it), which also makes it possible to search a selected dataset, available in Linked Open Data Format, using SPARQL END POINT. In this way the aggregated metadata can be widely distributed, connected and enriched.

Keywords: System interoperability and data integration; Digital Library; Ontology; Metadata Crosswalks; Metadata standards; Linked Open Data

1. Introduction

This paper presents an overview of the CulturaItalia aggregation workflow describing the pillars of the system and data. CulturaItalia is a pilot project based on Linked Open Data. CulturaItalia (http://www.culturaitalia.it) is based on the same standards as Europeana (http://www.europeana.eu) and it has the same function at the national level as Europeana has on a European scale.

CulturaItalia, online since April 2008, is managed by the Italian Ministry of Cultural Heritage, Activities and Tourism (MiBACT) through the Central Institute for the Union Catalogue of Italian Libraries (ICCU). It is the Portal of Italian Culture, and gives access to a database of "metadata" which gathers and organizes the information arriving

----

* Corresponding author. Tel.: +39 06 49210425; fax: +39 06 4959302.
  E-mail address: sara.digiorgio@beniculturali.it

from the various providers participating in the project. Users can discover different kinds of digital resources which make up the country's extensive cultural heritage (museums, photographs, libraries, archives, galleries, exhibits, monuments, videos, discs, etc.). The Portal is targeted to general users, by offering them a service for retrieving information on Italian culture from one point of access, and to more expert users, such as the operators in the field, who can take advantage of a high-quality showcase to promote their own cultural digital resources.

2. Types of content in CulturaItalia

Through CulturaItalia's content, users can discover different kinds of digital resources which make up the country's extensive cultural heritage (museums, photographs, libraries, archives, galleries, exhibits, monuments, videos, discs, etc.).

CulturaItalia manages a catalog – called Index – which gathers and indexes the information provided by the partners, by means of metadata. The original data remain on the website of the provider, to which the user is guided, via a link, for a complete consultation. For example, in the case of a photograph in the Index, the preview image (thumbnail) is visible along with identifying data and a link to the provider's website, where the user can visualize the photograph in its original format, accompanied by all relevant information and services in order to get the full benefit of the item. CulturaItalia also publishes editorials to help users discover cultural heritage and activities and metadata provided by partners. The project has a presence on the major social networks, including Facebook, LinkedIn, Twitter, and Google+. It represents a response to the needs of an expert audience as well as the needs of the general public. CulturaItalia offers specialised users, such as students, researchers, and those employed in the cultural sector, the opportunity to carry out targeted searches that correspond to very specific interests using cutting-edge software. For general users, such as citizens and tourists, the portal can pique curiosity and provide opportunities to discover or find out more about cultural resources in Italy, thanks in large part to editorial content (thematic itineraries, articles, highlights, events, columns) published to highlight the trove of "metadata" in the website's database. The Italian National Library Service's OPAC and the Europeana API are integrated in CulturaItalia's search functions, providing remote access to the Europeana collections.

CulturaItalia is an "open" system: it grows and develops hand-in-hand with the continuous enrichment of its metadata Index, through metadata harvesting using the OAI-PMH protocol, which allows the harvesting of metadata from content providers to one or more harvesters, adding services such as an indexing system or automatic classification. CulturaItalia ensures that digital resources are interoperable by adopting a cross-domain Application Profile (PICO AP, PICO is the acronym for "*Portale della Cultura Italiana On-line*"), based on the Dublin Core Metadata Initiative technical guidelines. The Portal harvests metadata from different repositories and exports metadata to Europeana in accordance with to the Europeana Data Model (EDM) profile.

After retrieving the metadata of interest in CulturaItalia, the user can directly access the data source by heading to the provider's website or by contacting it via other channels in order to find out more about the resource.

The portal aggregates over 2.4 million metadata from 32 public and private partners included thematic aggregators such as Internet Culturale, the portal of the Italian Libraries, also created and managed by ICCU: the metadata published in Internet Culturale are automatically transferred to CulturaItalia, and then, if the providing libraries agree, to Europeana. As national aggregator, CulturaItalia is the most important of Europeana's Italian content providers: almost 720,000 metadata from various providers have been aggregated in CulturaItalia and sent to Europeana, under the condition of the Creative Commons Public Domain Dedication CC0, which is mandatory for publishing metadata in Europeana pursuant to the Europeana Data Exchange Agreement.

3. Dati.CulturaItalia: a Linked Data pilot project

The pilot project dati.culturaitalia.it started in 2012 in order to establish a Linked Open Data (LOD) Service that will progressively make available open data-sets from the web-portal CulturaItalia. The application was designed by the CulturaItalia team with the technical and scientific support of Scuola Normale Superiore in Pisa (SNS), to allow the resources aggregated by CulturaItalia to be including large semantic networks after displaying, sharing and connecting data according to LOD principles.

An early version of this service has been available on-line since March 2013, as a section of CulturaItalia dedicated to LOD. It presently makes available as LOD the Thesaurus PICO, adopted by the portal for facilitating the browsing of a variety of resources in its domain, and metadata from the following data-sets: Accademia di Santa Cecelia, ArtPast Project, Digibess, ICCU, Internet Culturale, Michael Italia, Polo Museale Fiorentino, Regione Marche and Anagrafe delle Biblioteche Italiane. All data are licensed under CC0 1.0 – Universal Public Domain Dedication. More data-sets will be published as LOD, as soon as they are delivered under an open licence agreement, in compliance with the policies and guidelines recommended by ICCU as the institution in charge of CulturaItalia and in accordance with the Europeana Data Exchange Agreement (DEA).

The CulturaItalia platform extracts the datasets, in XML PICO format, that were submitted by the providers who were allowed to participate in the pilot and convert the PICO metadata into CIDOC standards, in accordance with the mapping document elaborated by Emilia Masci of SNS. The mapping is implemented in an XML stylesheet and the result is an RDF/XML representation of each data provider's metadata. The CulturaItalia repository then allows semantic enrichment with four types of reference resources:

- authority files such as VIAF (Virtual International Authority File: www.viaf.org);
- GeoNames (www.geonames.org/);
- PICO Thesaurus in SKOS;
- DCMI Type vocabulary.

The SPARQL endpoint provides access to RDF metadata structured according to the CIDOC - Conceptual Reference Model in the implementation of Erlangen CRM / OWL. Data can be searched over three querying interfaces, corresponding to three sections of dati.culturaitalia.it:

- Text search: here it is possible to perform free text searches over all triples contained in dati.culturaitalia.it;
- SPARQL query: here users can try their hand at a SPARQL query. There are also some examples of queries;
- iSPARQL query: here there is an even more complex querying interface for advanced users.

In Dati.CulturaItalia is available OAI Provider that makes available XML or RDF metadata structured according to different schemas:

- oai-dc (xml): OAI-PMH schema adopted by Open Archives Initiative Protocol for Metadata Harvesting;
- pico (xml): PICO Application Profile, the CulturaItalia Application Profile;
- edm (rdf): Europeana Data Model, adopted by the portal Europeana EDM;
- cidoc (rdf): CIDOC - Conceptual Reference Model in the implementation of Erlangen CRM / OWL.

In October 2012 Europeana transformed into Linked Open Data a large subset of aggregated metadata that was made available from data.europeana.eu;. the current version of the pilot contains metadata on 20 million texts, images, videos and sounds gathered by Europeana.

4. Conclusions

CulturaItalia and Europeana are sharing a common vision by making Italy's cultural heritage available to all for work, learning or pleasure. They provide their partners with a service to transform their metadata into LOD, on the only condition that they adopt the Creative Commons Public Domain Dedication CC0. Thanks to this project, museums, libraries and archives have started to open their data following common standards and under appropriate licensing conditions. The shared goal to participate in the development of the Linked Open Data Cloud represents an important step for promoting digital cultural heritage, becoming a node in a wider network, and participating in the process of innovation in the use of digital content through metadata integration and enrichment.

The exponential growth of data on the Web is one of the most significant outcomes of web publishing and it opens up new opportunities for library, museum, and archival science professionals who, thanks to LOD technology, can integrate information and facilitate web access to the data produced.

The main benefits for cultural institutions offered by LOD are:

- to steer web users toward aggregated content (for example, through the optimization of search engines), thus increasing traffic to the cultural website ;

- to enable the development of innovative services and new business models, e.g. by providing a new licence framework for the free use of low-resolution images and for the payment of high-resolution images;
- to encourage collaboration between libraries, archives and museums through the sharing of common data;
- to build new services around the LOD, such as APIs or APP for mobile devices, to give more opportunities for learning and for using metadata and content.

CC0 is being increasingly used in the realm of cultural heritage data. This means that museums, archives and libraries have waived all copyrights in their jurisdictions, effectively placing the data in the public domain. Besides, the data is available as linked open data, which means that the data sets are available as RDF (Resource Description Framework) on the web, enabling the data to be linked with other data from different sources.

Tim Berners Lee defines Linked Open Data as Linked Data that is released under an open license; as the word "open" implies, the Linked Open Data approach requires that data be published under a license or other legal tool that allows everyone to freely use and reuse the data. This requirement is one of the most basic elements of the LOD architecture. Additionally, according to Tim Berners-Lee's 5 star scheme, the most basic way of making data available online is to make it "available on the web (whatever format), but with an open licence". However, there still is considerable confusion in the field as to what exactly qualifies as "open" and "open licenses".

But as cultural institutions start to embrace Linked Open Data practices, the intellectual property rights associated with their digital collections become a more pressing concern. Cultural institutions often have problems with rights issues related to the content in their collections, because often they do not hold the copyrights to the works included therein. Instead, copyright often rests with the authors or creators of the works, or intermediaries who have obtained these rights from the authors, so that cultural institutions must get permission before they can make their digital collections available online.

However, the situation with regards to the metadata used to describe cultural collections is generally less complex. Factual data are not generally protected by copyright, and where descriptive metadata records or record collections are covered by rights (either because they are not strictly factual, or because they are vested with other rights such as the European Union's sui generis database right) it is generally the cultural institutions themselves who are the rights holders. This means that in most cases cultural institutions can independently decide how to publish their descriptive metadata allowing them to embrace the Linked Open Data approach if they so choose.

In conclusion it is inevitable to connect the theme of LOD with the management of intellectual property, especially considering knowledge as an economic good: the UE Commission's strategy establishes the principle that all information in the public sector and which does not fall within the exceptions is reusable even for commercial purposes; this involves the adjustment of the rule's framework on the re-use of data. Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 provides a positive re-use right to public documents and limits acceptable charging to only marginal costs of reproduction, provision, and dissemination of the digital cultural content.

Europeana and CulturaItalia believe that if we can make digital cultural heritage available online in open formats, we will start to see benefits for society and the economy. We have a shared responsibility for making this happen and we will support our memory institutions to help them open up their collections as widely as possible by developing standards, embracing new technologies, changing copyright, and developing new business models.

References

1. PICO Application Profile: http://purl.org/pico/1.1/picotype.xsd
2. PICO Thesaurus http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/thesaurus/4.3/thesaurus_4.3.0.skos.xml.
3. Europeana, 2012, Data Exchange Agreement http://pro.europeana.eu/web/guest/data-exchange-agreement.
4. CulturaItalia section devoted to LOD http://dati.culturaitalia.it/
5. http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/documenti/mapping/pico_cidoc/mapping_PICO_CIDOC-CRM_ITA-ENG.pdf
6. Berners-Lee, Tim, 2006 with addition in 2010, Design Issue; Linked Data, discussion document online: http://www.w3.org/Designissues/LinkedData.html