

OSSERVATORIO

The universe of cultural resources: between eureka's and concrete actions

Lightning talks presented to AIB CILW 2016 Conference

edited by the AIB study group on
Cataloging, indexing, linked open data and semantic web (CILW)

Introduction

by Andrea Marchitelli

Last 21st October 2016, the conference room of the National Library of Rome hosted the AIB CILW 2016 Conference “Revamping information resources: granularity, interoperability, and data integration” organized by the AIB study group on Cataloging, indexing, linked open data and semantic web, the Italian Library Association, Lazio Section and the National Library of Rome.

Aim of the conference was to put together experiences, researches and studies on the themes of theoretical, methodological, technological and professional innovation in the area of LAMMS, with specific reference to the LOD technologies and the semantic Web project.

The first part of the conference had the objective to promote an organic and innovative theoretical reflection on the main aspects of the convergence of information resources through the data linked to them.

The second part – which I had the pleasure to moderate – hosted seven talks selected after an open call for proposals. These talks have been presented as ‘lightning talks’ during the conference, and were meant to give scholars the opportunity to explain their ideas and experiences on the proposed themes, highlighting ongoing

AIB CILW STUDY GROUP (Cataloging, indexing, linked open data and semantic web), composed by Roberto Raieli, Marinella Cisternino, Elena Corradini, Antonella Iacono, Antonella Trombone, website <www.aib.it/struttura/commissioni-e-gruppi/cnc>, e-mail cnc@aib.it.

AIB CILW 2016 Conference home page: <<http://www.aib.it/attivita/congressi/c2016/giornata-studi-aib-cilw-2016/aib-cilw-2016-conference/>>.

ANDREA MARCHITELLI, EBSCO information services, e-mail amarchitelli@ebSCO.com.



ing or closed project results related to the conference themes. Six of them are now published in the following pages.

The third part, then, was aimed at the discussion of other projects and experiences able to give evidence of the potential of LOD and Web of data.

'Lightning talk' is a brand new way to present contents at conferences, at least in Italy and in our field, and has several characteristics that make it interesting and with great potential. In fact, lightning talks are brief presentations, during few minutes: at the CILW conference they lasted for 5 minutes. In this short time, the speaker has the chance (and the obligation) to present contents in the most effective way.

The speakers in the lightning talks session interpreted correctly the underlying idea, thus providing convincing and dynamic presentation, with a good complement of images and slides (and videos, in two cases out of seven¹). Due to the reduced amount of available time, each presentation conveyed the main ideas in an effective way to the audience.

The two presentations delivered first (*Criteria for the selection of a semantic repository for managing SKOS data* and *Modelli per la comunicazione dei dati di ricerca in archeologia: il caso dei Getty vocabularies come linked open data*) focussed on one of the key characteristics of semantic Web and LOD, the semantic interoperability manageable with SKOS: this feature was presented as an application to execute simple queries, and as a tangible experience, the one of *Getty vocabularies as linked open data* using LODs.

The two following presentations (*Coordinamento delle biblioteche speciali e specialistiche di Torino: progetto linked open data* and *Gli authority file per l'integrazione cross-domain dei beni culturali: riflessioni su un approccio alla lettura trasversale dei beni culturali della Chiesa cattolica italiana*) showed the potential of ontologies and LOD with respect to the integration and interoperability of data coming from different sources and tools, catalogued and described in different ways and related to very different objects (library and archival heritage, arts, people, institutions...).

The other two talks dealt with ongoing projects, having the common aim to make available specific collections or entire museums to the public (*L'archivio fotografico Zeri in linked open data* and *U. Porto digital museum: towards convergence in university's heritage resources management*²).

The last presentation (*Coming AUTH: per una bonifica e implementazione dell'authority file di SBN*) presented a project aimed at the implementation of SBN authority file with two levels, the first dedicated to the library professionals, the other conceived as an open collaboration with the non professionals, using a Mix'n'match service like the one developed by Wikimedia, inviting users to check the correspondences between entities coming from different sources.

In conclusion, a real richness of contents and experiences shared effectively with the audience... in about half an hour!

¹ Videos and slides used during the lightning talks session are accessible from the web page of the conference program: <<http://www.aib.it/attivita/congressi/aib-cilw-2016-conference/2016/56896-convegno-aib-cilw-2016-programma>>.

² This talk is not published here.

Criteria for the selection of a semantic repository for managing SKOS data

by Ricardo Eito-Brun

Introduction

Libraries, archives, and other organizations dedicated to the preservation and dissemination of the cultural heritage have started paying attention to semantic data technologies. These technologies are part of the linked open data (LOD) strategies, and have become a key component of the institutions' information management infrastructure.

Semantic Web technologies and standards include, but are not limited to, the RDF (Resource description framework) and RDF(S) languages, OWL, RDFa and SKOS (Simple knowledge organization system)¹. The last one is the W3C standard for encoding different types of knowledge organization systems: classifications, thesauri, list of subject headings, etc. Using SKOS, institutions can encode and tag knowledge representation tools and use them in the Web to make access to information easier. SKOS has brought the opportunity of reusing existing vocabularies and thesauri in the context of the new LOD applications. Thanks to this standard, the effort made in the past to design and develop controlled vocabularies can be reused now in the context of innovative applications.

The adaptation of an existing knowledge representation system (i.e. thesauri, list of subject headings or classification schema) to the SKOS format, requires the conversion and the tagging of the data into a specific XML vocabulary. This activity can be subject to automation, and institutions can easily apply text processing techniques (e.g. Python or Perl scripting) to add the necessary tags to existing vocabularies. But the development of a LOD project not only requires the conversion to RDF or to RDF-related schemas of the existing data. It is also necessary to give people the possibility of using data, and in particular, or using these data from other, third party software applications. This is, clearly, one of the most important challenges and overlooked aspects of semantic Web and LOD projects. The purpose of these projects should not be limited to the publication of the data in a different way, using different formats, for a human audience. The aim of semantic Web initiatives is to support the efficient reuse of the data by other software applications.

To achieve this objective, institutions must pay attention and dedicate a significant effort to make the data available in a way that support their dynamic exploitation. Today, most of the LOD initiatives developed by libraries and archives have focused on the conversion of existing data and their publication in RDF format. But these activities will not provide the expected return unless the efficient reuse of these data is properly supported by additional tools and technologies.

RICARDO EITO-BRUN, Universidad Carlos III de Madrid, Madrid, España, e-mail reito@bib.uc3m.es.

¹ World Wide Web Consortium (W3C), *SKOS Simple knowledge organization system: W3C recommendation 18 August 2009*, <<https://www.w3.org/TR/skos-reference/>>.

Technologies for data distribution

Once the data are converted into RDF or RDF-related formats (e.g. SKOS), organizations need to provide tools to query and exploit these data. The portfolio of semantic Web technologies and standards includes the SPARQL language, which is similar to the SQL language used to query relational databases. To query RDF datasets using this query language, common practice is to establish SPARQL end-points that provide URLs where you can direct SPARQL queries against a data set, and get the list of results in different formats (usually XML). SPARQL end-points are built on top of 'semantic repositories' that store the RDF or SKOS data. These semantic repositories offer a set of typical functionalities, including: the bulk import of RDF data, and the availability of SPARQL-end points that can receive, parse and execute SPARQL queries and return the results.

The main problems with the use of SPARQL end-points and semantic repositories are: a) lack of integration of these tools with tools to edit RDF or SKOS data (input files must be edited with specific tools before they are uploaded into the repository); b) access to data is restricted to users of the SPARQL query language; c) not all the data managed by libraries and archives are encoded in RDF or RDF-based formats, for example, we can use other XML-based formats like EAC (Encoded archival context) authority records. The main issue with the approach based on SPARQL end-points is that they are hard to use, as users must know the complex SPARQL syntax in order to integrate 'consumer applications'.

In the case study analyzed in this paper, these reasons led to the selection of a native XML tool to build and give access to a dataset made up of SKOS-encoded data.

Conversion and loading of SKOS data set

The project under analysis had the aim of publishing two thesauri created in the past by Ministerio de Fomento², Spain. These thesauri covered different areas of civil engineering. Initially, their intended use was to provide the concepts used to index the descriptions of civil engineering works and document files. These descriptions were created with the EAD (Encoded Archival Description) standard. Cataloguers in charge of describing the materials were provided with an online tool where they could search and select terms from these thesauri. The access to the thesauri was incorporated through a search feature added to the XML editor used to write the EAD records.

To provide this functionality, the existing paper-based thesauri were digitized and converted into electronic format. Using a set of automated scripts, the process of inserting the SKOS XML tags within the content of the thesauri was completed. This was done using the original thesauri labels for broader terms (BT), narrow terms (NT) or related terms (RT), and replacing them with the corresponding SKOS tags. After encoding the tags according to the SKOS specification, the thesauri were loaded into an XML database in order to provide query and search capabilities to external agents (both human users and computer-based applications).

In a second step, the tagged thesauri were split into separate files. This was done using an XSLT stylesheet with the help of the Altova® MapForce™ tool. Separate XML files were created for each concept in the thesauri. Each file contained the key data about each concept: authorized form, alternative, non-authorized forms, broader, narrow and related terms. To enforce the links between the different concepts in the thesauri, each concept was given a URI (Uniform Resource Identifier). The links and cross-references between broader, narrow and related terms were built using the tar-

² See: <http://www.cehopu.cedex.es/etm/etm_index.htm>.

OSSERVATORIO

get terms' URIs. In this way, it was possible to keep the relationships between the concepts, regardless their storage in separate files.

Finally, the set of files were loaded in a native XML database. A decision was taken to use Oracle® Berkeley DBXML as data storage repository. This database corresponds to the 'embedded XML database tools', which provide a good performance and reliability and minimize the administrative tasks. Although Berkeley DBXML does not support the SPARQL query language, it offers a powerful query language that makes use of the XML tags, and it is feasible to interact with the database using APIs (Application Programming Interfaces) available for different programming languages: JAVA, PHP, C, etc.

In this initiative, the interaction and communication between the external tool (the EAD/ISAD(G) XML editor) and the Berkeley DBXML database was implemented by means of Rest web services implemented with PHP. In order to standardize the query mechanisms, the interaction and the exchange of messages between the database and the external consumer tool was built using the SRU technical protocol published by the Library of Congress.

Using these services and messages, as defined in the SRU technical protocol, users were empowered to query and interact with the digitized thesauri. They were able of searching, browsing, and navigating the network of concepts. Queries also supported different choices: for example, search only the preferred terms, search any term in the vocabulary, search terms with a meaning more specific or more generic than another one, etc.

Conclusions

The use of semantic technologies to distribute and open data to third parties and support the development of innovative, distributed applications has grown in the last years. More and more organizations are being involved in these initiatives, and semantic Web and LOD are bringing new opportunities to leverage the value of existing data sets.

The publication of thesauri – or other data sets – requires a two-step process: the conversion of the existing data into a suitable format, and the deployment of tools to query the data. This paper describes the combined use of SKOS and a native, embedded XML database to create a point of access to digitized thesauri. The approach has proved to be successful in the development of a technical infrastructure for content and metadata management for the libraries of Ministerio de Fomento, Spain.

ABSTRACT

RICARDO EITO-BRUN, Universidad Carlos III de Madrid, Madrid, España, e-mail reito@bib.uc3m.es.

Institutions and organizations have started paying attention to the semantic data technologies as a key component of their information management infrastructure. A key component in traditional information systems that is expected to have a great relevance in the semantic Web and linked open data projects are knowledge organization systems: classifications, thesauri, list of subject headings, etc. W3C technical standards for the semantic Web include the SKOS (Simple Knowledge Organization System) specification for encoding and tagging these knowledge representation tools. The adaptation of an existing knowledge representation system to SKOS requires the conversion and tagging of the data..

Communication patterns for archaeological research data: the case of the Getty vocabularies as linked open data

by Angela Bellia and Fiammetta Sabba

Linked data for cultural heritage

The linked open data (LOD) model, widely established in the use of government and scientific data, is also assuming a central role in the organisation and sharing of cultural heritage data on the Web. Considering the importance of LOD in the promotion of cultural heritage over the Internet and its capacity to merge traditionally distant descriptive practices, this tool now has a prominent place in both the Italian and international debate. The LOD model paves the way not only for unified and global open access, but also for qualified and controlled access to cultural resources, making them available to everyone. The open and unified information environment that the LOD model is building has enormous potential to address some of the issues in the field of cultural heritage, including the interoperability of data and their integration. Through a computational approach that considers the relations among explicit and manageable data, new forms of access to information and new opportunities for interpretation of data and documents have become viable¹. Within a shared technological framework and a new form of representation, the data and metadata of libraries, archives and museums are connected through a methodology that aims to create, share, and reuse them. Thus, LOD are one of the most advanced technological strategies related to the management, cataloguing, promotion and communication of all typologies of cultural heritage. The LOD demands not only a radical review of the approach to traditional data representation, preservation and research, but also additional training for cultural heritage professionals, who need both to upgrade their skills beyond their specific field (Librarianship, Archival Science or Museology), and to be able to redefine their profiles in line with the contemporary cultural issues.

In Italy, some initiatives are spreading to promote training of professionals with both a solid cultural background and technological skills, as to be able to manage cultural heritage data and communication. Among these initiatives, a notable one is the Summer school named Linked data for cultural heritage, directed by Fiammetta Sabba, which was held last May in the Department of Cultural Heritage at the Uni-

ANGELA BELLIA, New York University, Institute of Fine Arts, New York, United States of America; Alma mater studiorum – Università di Bologna, Dipartimento di beni culturali, Ravenna, e-mail angela.bellia@unibo.it.

FIAMMETTA SABBA, Alma mater studiorum – Università di Bologna, Dipartimento di beni culturali, Ravenna, e-mail fiammetta.sabba@unibo.it.

Web sites last accessed: December 1, 2016.

¹ Hilary K. Thorsen; M. Cristina Pattuelli, *Linked open data and the cultural heritage landscape*. In: *Linked Data for Cultural Heritage*, edited by Ed Jones, Michele Seikel. London: Facet, 2016, p. 1-22.

OSSERVATORIO

versity of Bologna, Campus of Ravenna². The Summer school represented a very valuable formative opportunity because it allowed participants to learn about the principles and techniques of linked data, and apply them to cultural heritage³.

The Linked jazz project

Among the projects and case studies presented at the Summer school there was the project Linked jazz and The Networked Catalog at New York Pratt Institute Library: Data Modeling which included the creation of triples from text and metadata. The aim of this project was to use linked data to discover the significant connections between documents and data related to the protagonists and professionals of jazz music, and to develop relevant linked data tools and methods. The Linked jazz project showed how the application of LOD technology to cultural heritage and, particularly, to music resources was realised, initially by filtering a list of approximately 9,300 jazz musicians' names extracted from DBpedia, the LOD version for Wikipedia, and also from Europeana and the Digital Public Library of America (DPLA). Through the intersection and sharing of data formed by triples representing discography, photographs of musicians, performances, and library collections, one of the most significant findings was the possibility to highlight the presence and activity of many talented female jazz musicians who had a strong influence in the training of many of their more famous male peers. Thanks to LOD technology, which allows us to expand up from traditional access to archival content, the memory of these female musicians has not only been recovered, but also made available in a comprehensive and integrated research environment.

A case study: the Getty vocabularies

Museum data has been the least exposed of cultural heritage data. Licensing and provenance information are necessary before these data can be published online, and the correctness of data is extremely important for museums in maintaining their credibility and trustworthiness within the scientific world. However, museums have recently begun to use LOD technology in order to expose their data and collections on the Web. A very interesting example is that of the Getty vocabularies as linked open data of the Getty Research Institute in Los Angeles⁴. This case study aims to explore new models to improve the discovery and interpretation of cultural heritage through the application of LOD technology to digital archives. This large project is part of Getty's ongoing effort to make its knowledge resources freely available to all. The Getty vocabularies as linked open data project is a significant example of LOD creation practices. It contributes not only to the growing search for use of LOD in

² This research can be found within the research project TELESTES, funded by the European programme, Marie Curie actions, International outgoing fellowships for career development at the Institute of Fine Arts at New York University (outgoing) and at the Department of Cultural Heritage of the University of Bologna (incoming): <http://cordis.europa.eu/project/rcn/185858_en.html>. The Summer school took place in Ravenna from 16-20 May 2016: <<https://eventi.unibo.it/linked-data-per-i-beni-culturali>>. The programme for the next event is available online at: <<https://eventi.unibo.it/linked-data-per-i-beni-culturali/news>>.

³ In addition to a theoretical and practical introduction to the latest methods of representation for resources in libraries, archives and other types of cultural institutions, the Summer school covered methods and techniques to increase the visibility and accessibility of the resources.

⁴ <<http://www.getty.edu/research/tools/vocabularies/lod/index.html>>.

libraries, archives and museums, but also makes visible a rich and diverse network of connections. The Getty vocabularies were built to allow to classify, describe, and index cultural heritage objects and information with a multilingual terminology including visual arts, architecture, preservation, archaeology, and archival and bibliographic materials. Moreover, the Getty vocabularies are compliant with international standards, they grow through contribution, and they transform human-defined relationships into machine-readable data sets, embedding them into the evolving semantic Web. By publishing the Getty vocabularies as linked data in an open environment for anyone to freely use, the Getty Research Institute shares the results of over thirty years of research and scholarship with the world.

In order to support the exchange, use and reuse of metadata, the Getty vocabularies adopt a simple language to formalize the concepts expressed in a semantic format, using RDF (Resource description framework) and appropriate ontologies that share the same basic semantic representation with the *Art and Architecture thesaurus* (AAT) (figure 1), the *Thesaurus of geographic names* (TGN) and the *Union list of artist names* (ULAN). All of them have in common a structured set of terminology for archaeology, the arts, architecture, decorative arts, archival and bibliographic materials. The data grows and evolves through contributions in connected and open environments, offering a powerful tool for research. To this we can add that the data comply with international standards, by providing authoritative information for cataloguers, and particularly for researchers and data providers. A description of a complete set of URIs (Uniform resource identifiers) is provided in the *Getty vocabulary semantic representation*.

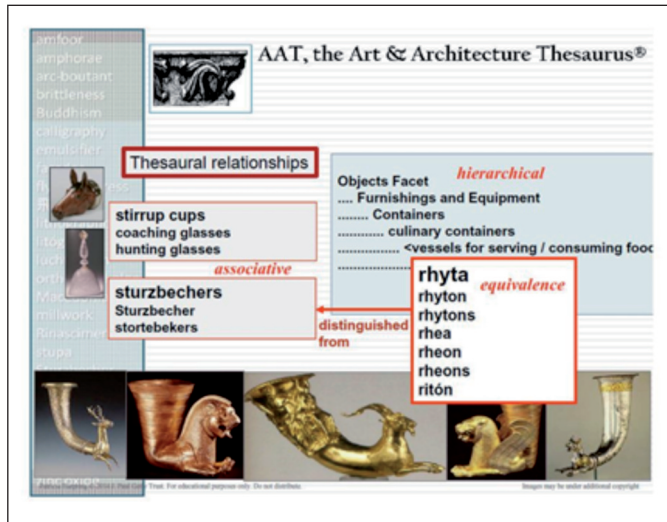


Figure 1 – Relationships in the *Art and architecture thesaurus*

Communication patterns for archaeological research

Some thoughts on archaeological research emerged from the analysis of this case study. The availability of the Getty vocabularies to the research community as LOD can have a transformative and evolutionary impact not only on the field of Archaeology and its related disciplines, but also for archaeological methods of investiga-

OSSERVATORIO

tion, because the LOD model contributes to maintaining the full cooperation of the community through a public discussion forum.

This new relationship between digital data and archaeological methodology, which has not been recently born, is part of a larger phenomenon that aims to build a chain of relationships and virtual interactions, and contributes to the creation of different communicative and totally digital forms. In the field of classical antiquity studies, and particularly in the archaeological field, these solutions aim to share information on a previously unknown level. This is not a simple quantitative expansion of available data on available archives, but a new methodology that takes into account the need to make information understandable to both machines and users. The analysis of the Getty Research Institute experience shows how this approach can contribute to creating new models and practices for the publication of data in archaeology. The goal is to foster collaborative studies, data analysis, and, above all, the sharing of information aimed at a new kind of knowledge transfer.

In conclusion, the Getty project seems to display both a quantitative and a qualitative leap in the study of data, in the aggregation of information and ultimately in their proper reuse. On the one hand, while these project features turn out to be useful and indeed essential to the renewal of archaeological methodology and that of related research fields. On the other hand, they provide appropriate tools and formats to support and develop a shared network of new methods for the conceptualization and organization of knowledge in a broader online collaborative space.

ABSTRACT

ANGELA BELLIA, New York University, Institute of fine arts, New York, United States of America; Alma mater studiorum – Università di Bologna, Dipartimento di beni culturali, Ravenna, e-mail angela.bellia@unibo.it.

FIAMMETTA SABBA, Alma mater studiorum – Università di Bologna, Dipartimento di beni culturali, Ravenna, e-mail fiammetta.sabba@unibo.it.

This paper presents the methods and tools developed within the Getty Vocabularies as Linked open data. The project aims to explore innovative patterns in order to improve the discovery and interpretation of cultural heritage through the application of linked open data (LOD) technology for digital archives. The project is part of Getty's ongoing efforts concerning the free open access to the Getty Research Institute's digital resources. The overall objective of the project is to bring greater visibility to the rich and diverse network of connections. For this reason, the project Getty vocabularies as linked open data is a case study in LOD creation practices: it will contribute to the growing body of research on LOD in libraries, archives and museums. We will discuss how the Getty vocabularies can help in the creation of new patterns and practices for the publication of data in the archaeological field in order to increase the sharing of studies, analysis, and information.

Coordination of special and specialized libraries of Turin: (CoBiS): Linked open data project

by Federico Morando, Emanuela Secinaro and the CoBiS LOD Group

CoBiS and the LOD Project

The CoBiS¹ is the Coordination of Special and Specialized Libraries of Turin. It was born in June 2008 by 10 libraries, in order to create a network for professional training and improvement of user services. It gradually expanded and now consists of 65 libraries and a considerable wealth of books and archival documents.

Users of the CoBiS's libraries are mainly driven by their interest towards the specificity of the documents and often need qualified help in consultation and bibliographical search. Librarians' professional training needs and the aim to enhance the assets of each institution, making them known and accessible, gave rise to the CoBiS's Linked open data (LOD) project.

It is currently the only project in Piedmont in this domain and it can therefore be considered Piedmont's pilot LOD project. As a future goal, we hope to involve other local libraries belonging to different institutions.

How the CoBiS LOD Project was born

The LOD Project started with a training program, in collaboration with Maurizio Vivarelli from University of Turin, which was divided into several related events and topics: copyright, collaboration between libraries and Wikipedia, Open data. It then developed with a focus on LOD applied to cataloging. Seven libraries from the CoBiS decided to participate to the pilot project, with the scientific collaboration of the Nexa Center for Internet & Society at Politecnico di Torino and the technological support of Synapta Srl: Academy of Sciences of Turin, Olivetti Historical Archive Association, the Alpine Club National Library, INRIM – National Institute of Metrological Research, Deputazione subalpina di storia patria, Educatore della provvidenza, INAF National Institute of Astrophysics – Astrophysical Observatory of Turin.

FEDERICO MORANDO, Synapta s.r.l.; Politecnico di Torino, Nexa Center for Internet & Society, Torino, e-mail federico.morando@gmail.com.

EMANUELA SECINARO, Istituto nazionale di ricerca metrologica, Biblioteca, Torino, e-mail e.secinaro@inrim.it.

The CoBiS LOD Group is composed by Barbara Bonino, Elena Borgi, Maria Pia Girelli, Gabriella Morabito, Federico Morando, Emanuela Secinaro, Luisa Schiavone, Anna Maria Viotto.

Web sites last accessed: November 30, 2016.

1 <www.cobis.to.it>.

OSSERVATORIO

Different sources: the data at stake

Below is a brief presentation of libraries and data made available for the pilot.

The Library of the National Institute of Metrological Research (INRIM)² extracted 1,600 registered monographs from its catalog. The cataloguing software is ErasmoNet³ and data were extracted by creating an Excel table.

The Library of the Astrophysical Observatory of Turin⁴ is part of the Institute of Astrophysics – INAF⁵. The INAF library catalogs are merged into a single catalog⁶ managed by BIBLIOWin 5.0web⁷ software. For the project, the bibliographical data are extracted in XML in real time, by selecting only those from the Astrophysical Observatory of Turin⁸.

The Library of the Academy of Sciences⁹ has an estimated holding of nearly 300,000 volumes, of which over 15% old books. It is managed with Sebina open library and it is part of the Bibliographical Research Pole, headed by the University of Turin. Archival documents are cataloged with xDams. Sebina and xDams provide data exports in XML.

The Library of the Olivetti Historical Archive (AASO) collects monographs and 187 Italian and foreign periodicals edited by the company Olivetti S.p.A. The library is part of the Library System of Ivrea and Canavese and adopted the ErasmoNet software for cataloging. For the LOD project, it provides 1,800 titles, mostly composed by gray literature, from the fund of the Mechanics Training Centre (CFM).

The National Library of the Italian Alpine Club (CAI)¹⁰, in Turin since 1863, adheres to SBN since 2004 and from 2014 it coordinates the unique catalog of cultural heritage of the National Museum of the mountain (7,500 news) and the CAI Sections (33,000 news) managed with Clavis.

The library of the Deputazione subalpina di storia patria¹¹ makes available for the project its own publications, starting in 1836 with the publication of the first volume of *Historiae patriae monumenta*, and continuing, to date, with various collections, such as the recent Library of Studies of Italian History and Economic History.

The historical library of the Educatorio della provvidenza¹² consists of two sections: books and photographs. The photographic section includes 716 imagines of

2 <<http://rime.inrim.it/biblioteca/>>.

3 <<http://www.cs.erasmo.it/soluzioni/servizi-beni-culturali/erasmonet/>>.

4 <www.oato.inaf.it>.

5 <www.inaf.it>.

6 <www.bibliowin.it/inaf/>.

7 <<http://www.cgisi.it>>.

8 In this first phase of the project, for which funding is provided by the Piedmont Region, greater prominence will be given to local heritage, without excluding a future broader involvement at the geographic and institutional level.

9 <http://www.accademiadellescienze.it/biblioteca_e_archivio>.

10 <<http://mnm.comperio.it/biblioteche-cai/Biblioteca-Nazionale/>>.

11 <<http://www.deputazione.subalpina.it/biblioteca.html>>.

12 <http://www.cobis.to.it/Web/Biblioteche_aderenti/Biblioteca_storica_dell_Educatorio_della_Provvidenza__Paolo_Girelli_e_Vera_Tua>.

art history from the 1930s and pictures of the period spanning from the late nineteenth century to the 1970s. Some photos and related metadata from this collection have been provided for the pilot.

Technical aspects of the project

The CoBiS LOD Project aims at providing the CoBiS with an infrastructure to publish LOD, creating a pipeline for triplification (transformation of data in linked data), which is easy to automate and replicate. This is done using open source technologies, including the RML mapping language¹³.

The RML mapping language

The RDF mapping language (RML) is a generic language, which can be used to define (using RDF triples) the rules for transforming heterogeneous data (XML, JSON or tabular data) into linked data.

RML is freely available under the MIT license on GitHub¹⁴ and Synapta made some improvements to the RML-Processor to speed-up (by about two orders of magnitude) the triplification process.

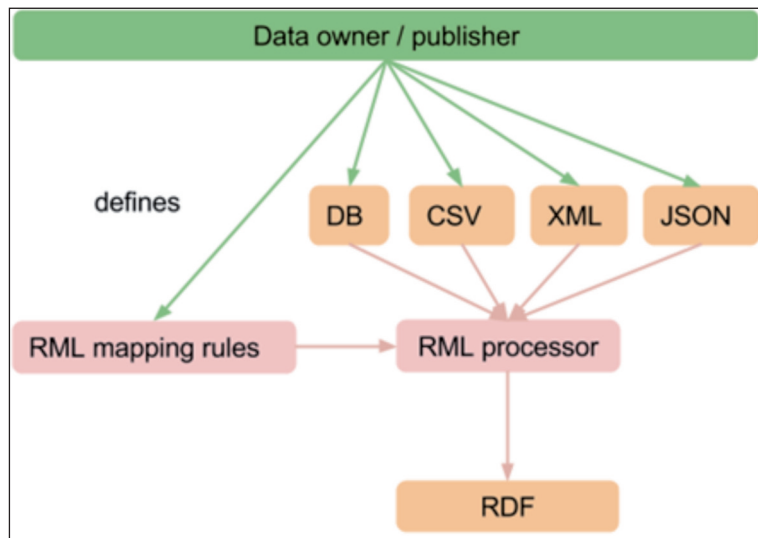


Figure 1¹⁵

Architecture

RML is just one component of the overall software architecture of the CoBiS-LOD project, as represented in figure 2.

¹³ <<http://rml.io/>>, by iMinds, Gand University.

¹⁴ <<https://github.com/RMLio>>.

¹⁵ Retrieved from: <<http://rml.io/>>, © 2013–2016 Ghent University – iMinds.

OSSERVATORIO

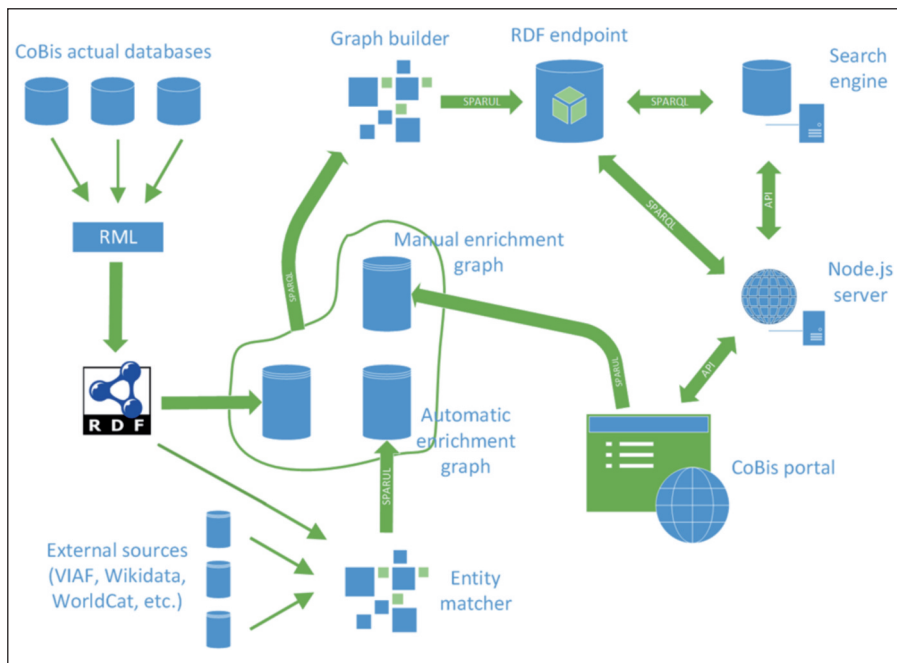


Figure 2

Heterogeneous data from the CoBiS databases are processed using the RML mappings, to produce a first graph of ‘raw’ LOD¹⁶ (i.e., data expressed in RDF, but not connected to other sources), which is saved to a specific database (triplestore). In parallel, we collect data sets from external sources (VIAF, Wikidata, WorldCat, etc.). The component called Entity matcher performs an initial automatic LOD interlinking of the ‘raw’ data. Additional manually defined rules to manage the relationships between the different sources are saved in another graph, which can then be updated from the back-end of the CoBiS portal (e.g., librarians can add or manually correct entity matchings between authors and/or works from different sources). An additional component, the Graph builder, will then create the standard CoBiS URIs and write on a reference triplestore the consolidated version of the data. A search engine will index these data, for a more efficient use by the CoBiS portal (which, in turn, will be powered by a Node.js server, on the basis of the search engine and live SPARQL queries).

All the architecture is based on open source technologies, and the pieces of software developed by Synapta for the CoBiS-LOD project will be released under an open source license¹⁷.

¹⁶ RDF (Resource Description Framework) is the standard for representation of linked open data, see: <<https://www.w3.org/RDF/>>.

¹⁷ All major components will be open source, to facilitate reuse. To optimize costs and performances, we are considering the use of some proprietary technologies, such as SOPHIA semantic search by CELI for the search engine.

Interlinking and reference ontologies

Much of the automatic interlinking activity will be focused on connecting the CoBiS collections to the rest of the world, to enrich cataloging information with additional data. But also the reverse interlinking – namely, the addition of links to the CoBiS platform on third party’s resources – is part of the project, in order to make CoBiS libraries more visible online.

With reference to ontologies, the project is based on reusing the most common standards, in particular the Bibliographic framework initiative (BIBFRAME) launched by the Library of Congress, but also Schema.org, to maximize the impact on search engines.

User interface

The main source of inspiration for the LOD Project is the ‘simplicity’ (at least from an aesthetic point of view) of the WorldCat.org interface, but projects more focused on linked data, such as the SHARE catalogue¹⁸ one, will represent an additional reference.

Concluding remarks and future developments

The roadmap for the pilot project – which represents the first part of a more ambitious project for a CoBiS digital library – is to be online, in beta version, by the second quarter of 2017¹⁹.

ABSTRACT

FEDERICO MORANDO, Synapta s.r.l.; Politecnico di Torino, Nexta Center for Internet & Society, Torino, e-mail federico.morando@gmail.com.

EMANUELA SECINARO, Istituto nazionale di ricerca metrologica, Biblioteca, Torino, e-mail e.secinaro@inrim.it.

The CoBiS LOD Group is composed by Barbara Bonino, Elena Borgi, Maria Pia Girelli, Gabriella Morabito, Federico Morando, Emanuela Secinaro, Luisa Schiavone, Anna Maria Viotto.

Web sites last accessed: November 30, 2016.

The CoBiS is a network of 65 specialized libraries. The CoBiS's Linked open data (LOD) project is the pilot of Piedmont to publish LOD, using open source technologies, such as the RML mapping language. The first challenge of the project is to merge cataloguing data from five different pieces of software (Clavis, ErasmoNet, BIBLIOWin, SbnWeb and Sebina OpenLibrary), also linking to multimedia content and some archival data.

The interlinking focuses on VIAF and Wikidata and the reference ontologies are BIBFRAME and Schema.org.

18 <<http://catalogo.share-cat.unina.it/>>.

19 At the URL: <<http://dati.cobis.to.it/>>.

The authority data for cross-domain integration of cultural heritage: reflections on the cross-reading approach of cultural heritage of the Italian Catholic Church

edited by the Working Group on Authority Control established at the National Office for the Ecclesiastical Cultural Heritage and Housing Worship of the Conferenza episcopale italiana

The purpose of this paper is to present the work methodology for the management and use of authority files within the cross portal Beweb, where the authority data act as turning points.

The reference context of the cross portal of cultural ecclesiastical heritage helps to a better understanding of the identified solutions¹.

Beweb: the cross portal of cultural ecclesiastical heritage

Beweb is the integrated portal of the ecclesiastical cultural heritage, which collects and presents the work of cataloguing and systematic census of all ecclesiastical cultural heritage – started in 1996 and still in progress – carried out by the Italian dioceses and the ecclesiastical cultural institutions.

Historical and artistic heritage, architectural, archival and bibliographical live together in Beweb; databases are constantly updated and increased. But there are also other ‘objects’, which aim to provide tools to approach to the contents of the portal. Among these, glossary words, which are available both at a cross-level and for each domain; the descriptive pages of dioceses and ecclesiastical regions, where there are useful historical information about these ecclesiastical territorial structures; the pages of ecclesiastical cultural institutions, which give value to museums, libraries and archives described in the Register of ecclesiastical cultural institutions of CEI. And there are also authority records of type person, family, organization.

The census and cataloguing projects are promoted by the Office for Ecclesiastical Cultural Heritage and Housing Worship of CEI, adopting a distributed working model: the real cataloguing work is done by individual operators in the national territory (about 3,000 specialists: librarians, archivists, art historians and architects, the only guarantors of the scientific nature of the work), while the central working group on cultural heritage deal with the treatment of these data and their presentation and fruition on the Web.

WORKING GROUP ON AUTHORITY CONTROL ESTABLISHED AT THE NATIONAL OFFICE FOR THE ECCLESIASTICAL CULTURAL HERITAGE AND HOUSING WORSHIP OF THE CONFERENZA EPISCOPALE ITALIANA, composed by Adriano Belfiore, Francesca Maria D’Agnelli, Laura Gavazzi, Claudia Guerrieri, Maria Teresa Rizzo, Silvia Tichetti, Paul Gabriele Weston, website <<http://www.beweb.chiesacattolica.it/it>>, e-mail beweb@chiesacattolica.it.

1 <www.chiesacattolica.it/beweb>.

The complexity and variety of cultural heritage described and catalogued is just recognizable from the home page, where we find a detailed menu according to the specificity of the presented cultural areas: the historical and artistic heritage, architectural, library, archival and cultural institutions.

The web pages dedicated to each domain propose their own research and reading methods on the database: for example, on the pages dedicated to the historical and artistic heritage or to the worship buildings, we can search by chronology, on the library goods page, the research by subject is highlighted, while on the archival heritage one, you can search by creator and custodian.

The real innovation of Beweb is the cross domain approach to the reading of cultural heritage, visible on the home page, where, for example, from a simple research you can query databases of different areas obtaining variegated results with resources that are always related to each other.

The composition of Beweb and its narrative capabilities are highlighted appreciating the circularity of navigation in Beweb through the authority records. For example, looking for general information about the architect Elia Fornoni, you get a composite result.

The research results are listed in a homogeneous way on the right, shown in many ways: list, gallery, map (in order to view the geographical distribution of resources), timeline.

On the result set, on the left, you can filter by the cultural area of interest, seeing only the results related to a specific domain, location, or chronology.

Among highlighted results in the filter by cultural area, we also see the authority file record.

The authority data in Beweb

The authority records in Beweb act as turning point for navigation cross domain of databases which describe different resources. The following working model was adopted for managing the access points: goods are described by professionals in that area (librarians, archivists, architects, art historians) respecting their own domain standard.

The choice was not to adopt a common descriptive standard, in order to guarantee the compliance with the specificity of each area.

During the census and cataloguing projects, the entities person, organization and family are taken over and it would be the case that a same entity 'person' (such as the architect Elia Fornoni) is represented in various cataloguing areas: the solution adopted was to manage centrally all these authority files from different sources, through a clustering system which guides the selection of names that refer to the same entity.

So, also in the case of access points, the choice was not to apply a common syntax for the form of the name, but to adopt the cluster model that we find, for example, in VIAF. In this way, a cluster of equivalent terms of access takes shape, collecting in a single access aggregating point (AF CEI) minimal data and sources.

The access point AF CEI cross domain consists of biographical and historical information (which compose the minimum data set, necessary to describe the access points) and links for further information (such as Wikipedia or the same VIAF). Then, the record is integrated with a 'friendly' and direct name form.

Undoubtedly, the most interesting part of the record is the cross navigation through the contents of the portal:

- connection to the resources through a type of relationship allowing a cross navigation with all resources having the same relationship (such as authorial or subject) with the entity;

OSSERVATORIO

- connection to the resources related by proximity (made explicit with the 'you might be interested in' label);
- and the direct link to the domain databases (made explicit with the label 'you find it in the domains').

The control of the access points in the authority archive allows to have information that are the starting point to begin a dynamic navigation, for example, guiding an art lover to get in touch with the archival materials related to the altarpiece of his interest; and discovering, for example, that a pastoral visit contained in an archival collection of an episcopal curia talks about that altarpiece. Continuing navigation, you can discover the biography of the bishop who guided the visit, and learn about bibliography of him; besides, you might find out that the bishop ordered the restoration of a certain cathedral of which you can browse pictures.

Among solutions to facilitate the cross-reading of the ecclesiastical cultural heritage for a wider audience, not just specialists, there is the simplification.

The specialized language of inventories (which responds to standard and domain vocabularies) was 'translated' by a friendly display of information, conveyed with intelligible labels in a natural language.

Beweb is also accessible from SBN (Italian national library service). If the searched book is found in a library of the Polo biblioteche ecclesiastiche (PBE), the link to the catalogue points directly to Beweb. From the book record to the author record connected to it, where you can discover the presence of different resources related to the research done.

Different developments are under study for the AF CEI management: the increase of new categories of entities (places, topical terms and works will be progressively added to persons, organizations, and families); the increase of new relationships (not only between entities and goods, but also between entities and entities, such as the relationships 'he is student of', 'he is collaborator of', 'he is client of', 'he is successor of', which link two entities person, or the relationship 'he is bishop of' that connects a person entity with an organization entity, or even the relationship 'it is financed by', that binds to a family entity.

ABSTRACT

WORKING GROUP ON AUTHORITY CONTROL ESTABLISHED AT THE NATIONAL OFFICE FOR THE ECCLESIASTICAL CULTURAL HERITAGE AND HOUSING WORSHIP OF THE CONFERENZA EPISCOPALE ITALIANA, composed by Adriano Belfiore, Francesca Maria D'Agnelli, Laura Gavazzi, Claudia Guerrieri, Maria Teresa Rizzo, Silvia Tichetti, Paul Gabriele Weston, website <http://www.beweb.chiesacattolica.it/it>, e-mail beweb@chiesacattolica.it.

Beweb makes it possible to explore the cultural heritage of the Italian dioceses (the historical and artistic, bibliographical and archival heritage, worship-allocated buildings and cultural institutions) through the cross domain research, the thematic in-depth examinations and the constant authority work on entities (person, organization, family) in various ways related to the goods and the other information contents of the portal.

What is the main challenge? It is to identify the access points to browse across the board databases that describe the different cultural heritage, and strengthen connections between objects making them not only capable of interworking but also of illustrate and introduce themselves mutually, in a synoptic and integrated way.

The Zeri photo archive in linked open data

by Marilena Daquino and Francesca Tomasi

The Zeri & LODE project

The Art historical photo archive Federico Zeri¹ of the University of Bologna preserves one of the most relevant European collections of photographs depicting artworks (290,000 exemplars), aside a rich library of art history (46,000 volumes), an important collection of auction catalogs (37,000 catalogs) and the collector's documentation (15,000 archival documents).

In 2003 the Federico Zeri Foundation started cataloguing and digitizing its collection, including a repertoire of artworks, depicted in photographs and fully described in attached documentation. Two national content standards were adopted to achieve the cataloguing process, i.e., Scheda F² (F entry) for describing photographs and Scheda OA³ (OA entry) for describing artworks. Both of the cataloguing rules are issued by the Istituto centrale per il catalogo e la documentazione (ICCD) and each one includes about 300 descriptive fields. The Zeri Foundation adopted a subset of such complex rules – about 120 fields provided by ICCD for the F Entry and about 100 fields for the OA Entry, to ensure that a minimum level of administrative data were documented.

In 2013 the International Consortium of Photo Archives (PHAROS)⁴ asked its 14 partners, including the Federico Zeri Foundation, for an active collaboration, in order to realize a common environment in which images and related metadata could be shared. The aim is to provide final users of photo archives, i.e., researchers and art critics, with a tool that integrates records of the most relevant European and American collections. To overcome difficulties due to the coexistence of different content standards, different technologies, and the gap between partners' data quality, the adoption of linked open data and CIDOC-CRM⁵ was suggested as an efficient solution to achieve the technological and semantic interoperability.

By means of the Zeri&LODE project (cfr. figure 1), the Zeri photo archive is the first Italian photo archive, and the first relevant use case among PHAROS' partners, which started the experimentation on its data to achieve the conversion into linked

MARILENA DAQUINO, Alma Mater Studiorum – Università di Bologna, Multimedia Center CRR-MM, Bologna, e-mail marilena.daquino2@unibo.it.

FRANCESCA TOMASI, Alma Mater Studiorum – Università di Bologna, Dipartimento di filologia classica e italianistica, Bologna, e-mail francesca.tomasi@unibo.it.

1 Federico Zeri Foundation: <<http://www.fondazionezeri.unibo.it/>>.

2 ICCD Scheda F: <<http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografci/Standard/10>>.

3 ICCD Scheda OA: <<http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografci/Standard/29>>.

4 PHAROS: <<http://pharosartresearch.org/>>.

5 CIDOC-CRM: <<http://www.cidoc-crm.org/cidoc-crm/>>.

OSSERVATORIO

open data⁶. About 30,000 catalog entries describing photographs, which reproduce about 19,000 works of modern art, are described according to the above mentioned CIDOC-CRM and two other models known in related communities: the SPAR ontologies⁷ borrowed from the publishing domain to enhance bibliographic descriptions, and the HiCO ontology⁸ to make statements about provenance and methodology underlying attributions.

Two specular ontologies were realized, called respectively the F entry ontology⁹ and the OA entry ontology¹⁰, which are capable to represent the expressiveness of the above described samples of F/OA entries. Such ontologies reuse named models and extend them with new classes, predicates and controlled vocabularies, thus integrating them with CIDOC-CRM terms.

Through the use of these models, several common scenarios may be described: a physical description of the cultural object, and its archival collocation (including the description of the hierarchy); roles and events, part of the life cycle of the object (e.g., commission, reproduction, print, change of property and movings); relations between documents, attached to photographs or cited as sources during the cataloguing process; finally, attributions, supported by technical criteria or bibliographic references.

Catalog entries	OA Entry		F Entry
Ontologies	OA Entry Ontology	CIDOC-crm SPAR HiCO	F Entry Ontology
Mapping Documents	OA Entry to RDF		F Entry to RDF
Data and exemplars	OA Entry RDF/ttl example	Zeri Archive LOD	F Entry RDF/ttl example

Figure 1 – Results of the Zeri & LODE project

Photo archives in the web of data: a model for integrating domains?

The Zeri&LODE project is a first and *in fieri* attempt to concretely realize a semantic integration between cultural heritage domains (galleries, libraries, archives, museums), aiming at representing heterogeneous information in the wide scenario of linked open data.

In this context, by defining as much as possible exhaustive and cross-domain ontologies for the cultural heritage domain we want to provide theoretical and

6 Zeri & LODE project: <<https://w3id.org/zericatalog/>>.

7 SPAR ontologies: <<http://www.sparontologies.net/ontologies/>>.

8 HiCO ontology: <<http://purl.org/emmedi/hico>>.

9 F entry ontology: <<http://www.essepuntato.it/2014/03/fentry/>>.

10 OA entry ontology: <<http://purl.org/emmedi/oaentry>>.

technical bases to several stakeholders. Results presented in this paper are thus applicable and reusable by other cultural institutions than photo archives. Our work can be considered as a mapping of terms from existing vocabularies wherein peculiarities and traditions of each single domain are preserved, so as to enable a comprehensive description of complex scenarios, of which the Zeri photo archive can be considered a representative example. Therefore, aside the above mentioned ontologies, two mapping documents¹¹ and two examples of usage (in RDF/ttl) provide a detailed description of mapped terms from F/OA Entry content standards, as to enable stakeholders to clearly compare their own standards to existing ontological predicates.

Moreover, by publishing a RDF dataset we want to offer another practical support to institutions, who can reuse data in similar projects. The current dataset includes about 11 million triples describing mainly photographs, artworks, artists, people and corporate bodies involved in events, and relations between cultural objects and documents (books, journals, archival documentation, etc.). Main entities were recognized and linked to online authority records (VIAF, Getty ULAN, GeoNames), datasets (from Dbpedia and Wikidata) and web resources (Zeri catalog web pages and Wikipedia pages). Terms referring to materials, artistic techniques and types of objects are aligned to relevant thesauri, like Getty AAT.

The final goal of the project is the realization of new instruments for research and discovery by means of semantic Web technologies, also contributing to highlight information not yet represented in a meaningful way. Indeed, the description of the cataloguing process is enhanced by annotating subjective aspects underlying questionable attributions, as it is one of the most interesting topics for final users of such kind of archives, i.e. focussed on art history. Information like how technical analysis on photographs was performed, which archival classification was chosen, or which opinion of art critics was provided by cataloguers to support attributions, are all fundamental points to reflect on with the perspective of the conversion into linked open data.

In fact, semantic Web enables the coexistence of contradictory statements about the same subject, as provided by different authors that can be considered less or more authoritative. Therefore, we have to manage how to ensure data quality, trustiness and provenance, in order to avoid inconsistencies when merging heterogeneous knowledge bases and to identify which cultural institution provides the most authoritative information.

This issue was taken into account when formalizing the HiCO model, so as to ensure a correct management of contradictory information. In particular, whenever a questionable attribution is considered (e.g. authorship, dating, roles in the life cycle of the object), methodological aspects, motivations, sources and criteria are taken into account too, in order to support authoritativeness of the attribution itself. Such meta-level of information guarantees authoritativeness and highlight quality of data offered by cultural institutions among the other data providers populating the Web of data, and hopefully will ensure a correct information retrieval process.

¹¹ Cfr. Mapping FtoRDF: <<https://dx.doi.org/10.6084/m9.figshare.3175273.v1>>; Mapping OAtoRDF, <<https://dx.doi.org/10.6084/m9.figshare.3175057.v1>>.

ABSTRACT

MARILENA DAQUINO, Alma Mater Studiorum – Università di Bologna, Multimedia Center CRR-MM, Bologna, e-mail marilena.daquino2@unibo.it.

FRANCESCA TOMASI, Alma Mater Studiorum – Università di Bologna, Dipartimento di filologia classica e italianistica, Bologna, e-mail francesca.tomasi@unibo.it.

The Art historical photo archive Federico Zeri of the University of Bologna owns one of the most precious European collections of photographs depicting artworks, aside a relevant library of art history and a collection of auction catalogs. The Zeri & LODE project, as part of the wider PHAROS project, addresses the data publication of the Zeri photo archive as linked open data. By creating two ontologies, F entry ontology and OA entry ontology, the expressiveness of two national content standards, i.e., Scheda F (F entry) and Scheda OA (OA entry) are represented.

Coming AUTH: a project for authority data enrichment and deduplication in SBN

by Agnese Galeffi, Andrea Marchitelli, Patrizia Martini and Lucia Sardo

One of the most common criticism towards SBN is related to the lack of structured, authoritative and updated authority data. In addition to this, there is long-standing issue of data maintenance – both of the catalogue and of the authority file – that has no real and easy solution. Maintenance is hard to manage in a cooperative environment such as the SBN one¹; an example above all are the levels: created with the aim to produce records with a different level of completeness², in some cases they became almost an obstacle to collaboration.

The lack of reliable data creates difficulties to librarians in cataloguing, but most of all it causes difficulties to the end users, being those users the reference librarians, the acquisition librarians, or the real users of the catalogue looking for information or for physical or digital 'objects'.

We are discussing about data and authority records, but we know that the majority of SBN access points are authority records made up of the preferred form of the name and of variant names, where they are available. Editing authors names, however, is not that easy because in order to modify them you must be in a library hub owning one of the records related to those names.

As an example, the table below shows the data related to the number of authors names, with breakdown per type and level:

	A	B	C	D	E	G	R	Totale
AUF	1.895	592	55.515	2.976	94	6		61.078
SUP	2.882	834	118.416	8.679	7.511	5.153	2.523	145.998
MAX	4.145	1.312	134.237	9.643	24.770	9.282	4.492	187.881
MED	5.812	2.112	438.479	25.932	49.770	15.822	21.542	559.469
MIN	32.136	18.951	2.283.040	144.005	229.864	44.261	82.558	2.834.815
REC	5.256	2016	418.306	29.100	89.449	12.668	20.884	577.679
Totale	52.126	25.817	3.447.993	220.335	401.458	87.192	131.999	4.366.920

Figure 1

AGNESE GALEFFI, Scuola vaticana di biblioteconomia, Città del Vaticano, e-mail galeffi@vatlib.it.

ANDREA MARCHITELLI, EBSCO information services, e-mail amarchitelli@ebSCO.com.

PATRIZIA MARTINI, Istituto centrale per il catalogo unico, Roma, e-mail patrizia.martini@beniculturali.it.

LUCIA SARDO, Alma mater studiorum – Università di Bologna, Vicepresidenza della Scuola di lettere e beni culturali, Ravenna, e-mail lucia.sardo@unibo.it.

1 More information in English about SBN (National Library Service): <http://www.iccu.sbn.it/opencms/opencms/en/main/sbn/>.

OSSERVATORIO

What is evident from this data is that on almost 16 millions records only 4.366.920 of them are authors names, and out of them 207.076 are 'authority record' with levels higher than 90. Richer records often lack in structure³ and are not homogeneous with respect to the style, language and length of the biographic note.

SBN developed during the years thanks to the cutting-edge collaborative idea of its very beginning, but then it did not exploit the same kind of working process for data maintenance, judging it as an activity to be managed merely at central level. The development of a unique database editor for the records number managed by SBN was not a feasible solution: during the years, then, ICCU launched a series of improvement projects, from the 90' to the conference in 2002⁴, until the document *Linee guida per la compilazione dell'autorità control - Autore personale in SBN*⁵, available on ICCU website and dated 2015⁶. The document *Sintesi delle priorità emerse dai documenti inviati dai Poli SBN in vista dell'assemblea*⁷ has to be mentioned, too.

Coming AUTH is a recent project, whose idea has been developed during the conference "1986-2016. Trent'anni di biblioteche in rete" to celebrate SBN thirty years, organised by ICCU and held at the National Library of Rome on April 1, 2016. In that occasion, it has become evident that the experience made by and within SBN was strongly related to a specific generation of librarians, and unfortunately not supported by the youngest ones, or by young professionals. An effect of this separation could mean that 'new generations' dealing with SBN in the future (a staging post, otherwise SBN would be closed or outsourced) are not likely to feel the same sense of belonging of those who created it.

In the context of SBN, Coming AUTH proposes a feasible solution with two ways for performing authority work: one at a professional level, the other at a social one.

As said previously, cooperative projects on this subject exist within SBN since many years (one of them is actually ongoing); the innovation brought by Coming AUTH lies in the involvement of professionals as such, not because of their affiliation only.

The first way to perform authority work implies the collaboration between ICCU and AIB Lazio to train specialized people, able to improve the authority data, to the

2 <http://norme.iccu.sbn.it/index.php/Norme_comuni/Codici/Informazioni_di_servizio/Livelli_di_autorit%C3%A0>.

3 To make an example, information about date of birth and date of death are combined in a unique field, while they should be split in two different ones.

4 <http://www.iccu.sbn.it/opencms/opencms/it/main/attivita/naz/pagina_96.html>.

5 <http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/AF_Indicazioni_e_regole_riv_23_luglio2015.pdf>.

6 Being those activities carried out at a central level and then sensitive as they have direct consequences on the collective catalogue and on records coming from different library poles, ICCU enabled these features to the librarians affiliated with the SBN maintenance office. However, the wish is to widen the working group to other poles, according to the information provided by the SBN Technical and Scientific Committee: <http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/catalog_manutenz_cat_sbn/pagina_0001.html>.

7 <http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/Assembleapoli_sintesipriorita.pdf>. See also the pages related to the proposals presented by each library poles: <http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/organi_sbn/pagina_0001.html>; and the proposals by ICCU: <http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/proposteICCU_assemblea_poli.pdf>.

use of specific tools (such as direct interface⁸, cataloguing software like SBN Web, and so on).

The choice of the records to enhance for first will be made according to selection criteria based on the needs of the different library hubs involved in the project, in the view of a regional basis cooperation (we should say, based on subsidiarity). A first hypothesis lead to the possibility to improve the authors names related to librarians, so that to propose a variety that goes beyond the peculiarity that each library hub might have, and that is immediately identifiable, quantifiable and that allows to test the theory underlying it. Thanks to existing information sources, moreover, this work would be made easier⁹.

A preliminary data extraction will enable to verify the validity of the hypothesis, and to define records to work on and the time needed. These conditions make the project scalable and to be dealt with in consecutive steps, when needed.

The other way, the social one, should be performed through a Mix and match service like the one developed by Wikimedia¹⁰, inviting registered user to confirm the correspondence among entities of different kind (names, places, etc.) coming from sources other than librarian' ones. These kind of system leverage on the satisfaction deriving from the gain of benefit or 'badges'. There will be a trust level for those correspondences, for example by validating them only if a certain percentage of users confirmed it.

As an example, Daniel Pitti presented a project called SNAC (Social Networks and Archival Context) at IFLA2016. SNAC is an archival-related project, it uses archival data to describe people in order to retrieve and identify distributed historical resources, so that to create the basis for an international collaboration in this area¹¹. To date, almost 25 millions SNAC entities have their own correspondence with VIAF, and this allows the enrichment of SNAC descriptions with variant names and links to Worldcat bibliographic records.

In the Coming AUTH social Mix and match, an entity is represented by an authority record poor in data, to be compared with another SBN record (for deduplication purposes) or with other sources, either related to the library science disciplines (like VIAF, ISNI, etc.) or not (Wikipedia, Wikidata).

In Wikipedia, the majority of the entries related to people contains IDs and link to catalogues and authority records of different kinds, in a box at the end of the entry itself¹². The system underlying those links is based on Wikidata properties: an ad-hoc template¹³ adds authority data to encyclopedia entries.

Such a project could foster the creation of a virtuous cycle because, through the enrichment of SBN authority file, all the related projects can benefit from this, both libraries and Wiki ones; as a consequence, SBN can use those data to enhance its

8 <http://www.iccu.sbn.it/opencms/opencms/en/main/sbn/Catalog_manut_catalogo_SBN/pagina_0001.html>.

9 To make an example, see the online version of the *Dizionario bio-bibliografico dei bibliotecari italiani del XX secolo*, edited by Simonetta Buttò: <<http://www.aib.it/aib/editoria/dbbizo/dbbizo.htm>>.

10 <<https://tools.wmflabs.org/wikidatagame/distributedhttps://tools.wmflabs.org/mix-n-match/>>.

11 In order to have a clearer idea on the project, have a look at the prototype search: <<http://socialarchive.iath.virginia.edu/snac/search>>.

12 <https://it.wikipedia.org/wiki/Aiuto:Controllo_di_autorit%C3%A0>.

13 <https://it.wikipedia.org/wiki/Template:Controllo_di_autorit%C3%A0>.

OSSERVATORIO

authority records. To make an example, through the link to Wikidata, SBN could be enriched with data coming from different authority files (i.e. ORCID).

The social way to authority control would have the beneficial effect to open SBN to users, then favouring the participation of non authorised personnel and improving the visibility of SBN itself and of the projects related to it, with the chance of an active involvement within the Wikimedia world, too.

To sum up, Coming AUTH would enhance SBN authority file and have the following added advantages:

- with reference to the professional level: the project is scalable and can be organised in consecutive steps, provided that specific dataset to work on are selected, and that working groups can count on people turnover;

- as for the social level, then: the added value lies in the gained visibility of SBN, ICCU and its projects with other players, with the likely involvement of Wikimedia. The promotion of SBN outside the ranks of the people working on it is not a marginal advantage, and opens to the need of free of costs promotion and marketing strategies.

The innovation that Coming AUTH brings is to deal with an old issue using new approaches and solutions.

ABSTRACT

AGNESE GALEFFI, Scuola vaticana di biblioteconomia, Città del Vaticano, e-mail galeffi@vatlib.it.

ANDREA MARCHITELLI, EBSCO information services, e-mail amarchitelli@ebSCO.com.

PATRIZIA MARTINI, Istituto centrale per il catalogo unico, Roma, e-mail patrizia.martini@beniculturali.it.

LUCIA SARDO, Alma mater studiorum – Università di Bologna, Vicepresidenza della Scuola di lettere e beni culturali, Ravenna, e-mail lucia.sardo@unibo.it.

The Coming AUTH project is about the enrichment and deduplication of authority data in SBN catalogue; the project will have both a professional procedure and a social one. Thanks to the cooperation between ICCU and AIB Lazio, the first one involves young professionals, and the choice of the data to be enriched will be decided later on.

The second one will take advantage of services like Wikimedia Mix'n'match, and will be open to everyone; anyone could register and then try to match the names of different entities. In doing so, we will be able not only to clean and enrich the authority data, but also to improve the visibility of the SBN catalogue and the participation of users to its improvement.

Ai fini della citazione, utilizzare esclusivamente il testo in lingua italiana, che presenta il doi, la paginazione, l'abstract e gli altri dati ufficiali.

When citing, please always refer to the Italian translation only, complete with doi, page numbers, abstract and other data.

[Associazione italiana biblioteche. Gruppo di studio Catalogazione, indicizzazione, linked open data e web semantico, *L'universo delle risorse culturali: lampi di genio e azioni concrete. Lightning talks presentati al Convegno AIB CILW 2016.*

AIB studi, vol. 57 n. 1 (gennaio/aprile 2017), p. 91-117. DOI 10.2426/aibstudi-11569]