



Metadata framework and application profiles in the global structure of catalogs and digitization projects of the Vatican Library

Paola Manoni

The issue of technical interoperability unfolds in the informatics field of the Vatican Library in an area covering the aggregation of bibliographic data (derived from systems with different metadata structures) and the interconnection of structured data (provided through sharing of cooperative programs) within the recent digitization project regarding collections of manuscripts and incunabulum. This presentation aims to show the context, as well as the application models and systems currently in use in the Vatican Library, focusing attention on the metadata framework involved in the overall structure of the new Library's catalogues inaugurated into the web on the 16th of May 2012. Moreover, there will be illustrated the organization of the Digital Library to be soon published on the Web. As a preliminary discourse it would be appropriate to mention something about the current state of informatics and cataloging development of the Vaticana Library, in order to locate more precisely the library environment into a broader design scenario of the global interoperability of the library sector. Starting from the



present and going backwards, the BAV (Biblioteca Apostolica Vaticana)¹ has warned the need to interact with different bibliographic databases, stipulating an aggregation method for heterogeneous data and syntax regarding the description of various collections composing Vaticana's patrimony. It was therefore designed a system to establish connections between various catalogs, in order to enable an integrated (and not federated) search within the platform of a general catalog. Without aiming to retrace the history of Vaticana's electronic catalogs but only to give some introducing information, here it is appropriate to mention that at the BAV the consultation databases are available, particularly, in relation to:

- publications (monographs and text periodicals): the first electronic catalog built in the middle of 80's years of the past century on the first group of current cataloguing, continued with the complete retrospective data entry and then subjected to revision (especially with regard to the access points);
- graphic materials (prints, drawings, photographs): the project "Stampe on line" ("Prints online"), started in 1998, includes the analytical cataloguing of prints as well as drawings. Since 2001, it has been proceeded with the digital scanning of images and their hypertext links together with the bibliographical cards;
- coins and medals: the project in progress that has been started in 2001. This project includes numismatic descriptions until the sixteenth century. As for graphic materials, almost for all data digital images are available: for each unit description the photo shots of the obverse and reverse of a coin are got;
- incunabula: newly established electronic catalog (available since 2009) that combines full descriptions of the inventory

¹<http://www.vaticanlibrary.va>.

published in 1997 and maintained by P. William Sheehan. The project BAVIC (Bibliothecae Apostolicae Vaticanae Incunabulorum Catalogus) consisting of analytical cataloging of specimens is in progress.

The mentioned above catalogs are structured under the MARC21 record syntax. Regarding the organization of electronic catalogs of manuscripts and archival documentary collections (the latter ones are in the Archives Section of the Department of Manuscripts), we refer to another type of computer processing and cataloging regarding both data format, system architecture and their interoperability.

- Manuscripts: since 2002 a project of retrospective conversion of catalogs and inventories of paper fonds has been operated. The project was divided into several phases. Preliminarily, the definition of criteria for data processing has been chosen according to the standard TEI (Text Encoding Initiative Consortium). The electronic cataloging is therefore conforms to the elements set out in the form of TEI-MS - specifically for the description of manuscripts. From a management perspective, a database able to handle its structure in XML was designed. The system, developed by the Coordination of computer services of the Vaticana is called InForMA (Informatics For Manuscripts and Archives), which manages the production and research data on manuscripts and on archive's papers.
- Archival material: formalized in the same XML language, but according to the XML EAD (Encoding Archival Description) standard. In effect, the system can handle different collections of data or documents that refer to different metadata schemas, at the same time allowing them to be managed separately (different formats for different cases) or in the related way: in cases of shared lists of authority items (for names, titles and se-

mantic descriptors) and the connections between bibliographic records and files of digital images.

Starting from 2011 there has been initiated a study for the adoption of a new management system OPAC able not only to aggregate in a general catalog the contents of individual catalogs but also to manage structures of separated and referenced indexes to each type of catalog. OPAC enables to link bibliographic records to other information resources available in the Library and to manage information useful for researchers, alongside the mere querying of the catalog. The tool which was built allows the regular population of data from different catalogs to a repository in which the different data formats or MARC21, TEI-MS and EAD are stored, each for the semantics of reference. The system conforms to the interoperability protocols such as OAI-PMH and OpenURL. Now it would be appropriate to mention a few words on the use of these two protocols at the BAV, with particular regard to the first experience of the aggregation of data for the general catalog. As it is known, the main goal of the protocol is to allow two or more archives of separated data to exchange metadata. OAI-PMH defines the communication protocol through rules and methods for transferring metadata, established according to the two classic basic entities: data providers and service providers. The data provider provides the OAI-PMH to publish its metadata. The service provider sends through the OAI-PMH requests (via http) to data providers and collects metadata. The data provider responds with an XML message. Considering the problems of data harvesting (given the extensive testing on which it stands) rather than the use of MARCXML for cataloging data in MARC 21 (which obviously did not present any problems), it is worth dwelling on the handling of metadata for archives and manuscripts involved in managing Vaticana's catalog. The regular population of the general catalog for descriptions of manuscripts and archives encoded

in EAD and TEI formats occurs according to a conversion format designed to maintain similar structural organization and granularity, thus giving way to interaction via OAI-PMH, considering the difficulty in managing crosswalk producing satisfactory results, which avoid the reduction or loss of information in data representation. One can think about the hierarchical organization of the descriptive units in EAD, about the richness of attributes and complexity of nested elements in both EAD and TEI-MS encodings. The literature on the topic abounds with the assessment on the Dublin Core/DC (also in qualified version) wherewith the protocol represents the aggregated data. Anyway, the DC format is too general to mediate the types of referred metadata. In the construction of the catalog, there was made a computers choice of the hybrid type. Particularly, besides the harvesting protocol functioning for MARC 21 catalogs there was implemented a Web service according to the peer-to-peer interaction model transferring the EAD and TEI-MS data from the InForMA system to the general catalog, ensuring the interoperability between heterogeneous systems. Beyond the harvesting technique, the system (based on the technological solutions offered by Infor) controlling the new catalogs offers an indexing system that allows the uniform representation of the results of a search. In practical terms, it means that the interoperability framework for communication among the catalogs has the function of the device for the collection of heterogeneous data converging within the same system in which, in real time, the bibliographic descriptions in different formats and from the specific catalogs are gathered. In other words, the interaction occurs within a database, whose conceptual model is based on the notions of information objects, classes, attributes and methods. The classes and attributes are used to describe the structural aspects (the metadata for each application domain), while methods are used to represent the functions of objects (information

units treated) derived from different cataloging representations. If within the problem of the interaction it would be desirable to situate other types of aggregation of information, in the evaluation of the front-end structure of new Vaticana's system one can see how – for each information unit dedicated to the various catalogs – the OPAC querying is provided, through an interface between various proposed widgets offering textual, visual, multimedia documentaries relevant to different document areas. The application presents to users a functional organizational complex that can be related to the model of a portal, consisting of web pages dedicated to each cataloging activity, to OPAC queried through searchable indexes and separated for each type of catalog and material typology. The search presents also the possibility to perform post-coordinated selections as the association of similar categories and concepts, through an automatic network of links offering to the researcher documents available of the same author or publisher in another language, another edition or another support, as well as techniques for scrolling through lists of results via progressive refinement categories, properties and attributes. Moreover, there are also searches on fuzzy² logic allowing the analysis of the search term in its different parts (roots, prefixes, suffixes, spelling variants, etc.) thus suggesting to the researcher performing a search other allied results. The search, starting from a bibliographic record, is further extended to the collection of electronic databases available at the BAV or other resources available via the web through a link resolver that conforms to the standard OpenURL. This last is a protocol for metadata exchange, aimed at managing services of the so-called "linking in context", widely used in bibliographic databases, in academic citation systems and in the open archive. It's also possible to manage the interaction mecha-

²Polyvalent Logic (literally 'fuzzy'), extension of Boolean logic or the categorization of a continuous variable in the subbranks.

nisms with RSS feed (for now not activated), as well as the function of word cloud. Both of them will be eventually allowed to meet the specific user information needs permitting the access to the system with authentication (after the activation profile is determined). The option for consultation of the application by entering user's credentials will also allow the user to customize the information display and OPAC functionalities (e.g. elimination of widgets, sorting or reduction of the search indexes, etc.). From a point of view of data aggregation and interoperability between systems at the BAV, there can be further extended the treatment with the interaction between catalogs and the RFID implementation created at the Vaticana as well as with the security system³ tracking the movement of books and persons within the possible paths in the Library. Anyhow, in order to not abandon the central theme of the present discussion, the work is propelled to analyze the issue of interoperability in view of digitization projects of the BAV. Regarding this topic, the Library is putting into practice the digitization both in a view of long-term storage, and in the implementation of a digital library accessible via the Web and through links to the catalogs. Leaving aside the first aspect, whose organization or specific use of metadata is evidently not directly involved with the subject discussed in this section, it would be appropriate to consider the structure of the second goal. Hereby there should be immediately added that, while the first project (concerning the long-term preservation) has already started, the second one has not yet produced any public evidence, even the presentation of the first group of data and digital objects in the Web over the next years is expected. Without any fear of contradiction, there can be said that treating a library of digital objects means, necessarily, handling of its metadata. The choices, which have been already accomplished in this direction at the BAV, are related to the

³System for internal use, which protects the privacy.

management of digital objects according to the METS (Metadata Encoding and Transmission Standard) model which represents an information container to encode in the XML metadata required for managing digital library objects. In addition, to each digital object accessible through METS profile a persistent URL is assigned. For the projects originated within cooperative agreements, stipulated with other international institutions considering the importance of the creation of digital libraries shared with the BAV, the digital objects are expressed in RDF graphs. These last represent an approach for structural interoperability, aiming at managing a connections between digital objects belonging to collections in various libraries. From this scenario there can be obtained a description of digital composite objects formed by a set of distributed resources in the web independent of their allocation. For this task each resource aggregation will be described in a resource map expressing the semantic relationships between existing aggregations of resources. It is expected that each map has an associated identifier (URI) that will be invoked (differentiated) with an http request, thus providing the serialized representation according to the RDF/XML standard format. In the first instance, these technical specifications make reference to the important work of digitization of the Vatican Library in partnership with the Bodleian Libraries of Oxford. This activity is made possible thanks to the contribution of the Polonsky Foundation committed to support initiatives providing access to and knowledge of cultural heritage of mankind preserved in the world's great library collections. Within this specific work plan, in a five-year period, it is expected the digitization of a million and a half pages in total regarding volumes (Greek and Hebrew manuscripts, and incunabula) chosen by both institutions. Thus, among the first treasures connected via the web differentiated URI there will be (among the incunabula) the famous *De Europa* di Pio

II Piccolomini, and the Latin Bible of 42 lines of Johann Gutenberg. While for the Hebrew manuscripts of the Vaticana, as for one of the most important collections of the existing Jewish codes (even if not the most extensive) there will be chosen the *Sifra*, written between the late ninth and mid tenth century and, probably, representing the oldest Jewish code come up to our days. Moreover, there will be chosen a whole Bible written in Italy around 1100 as well as biblical commentaries, Halakhah and Kabbalah, Talmudic commentaries, and writings from the liturgy, philosophy, medicine and astronomy fields. The earliest Greek manuscripts that will ultimately enter into the inferential logic of Linked Data will be important witnesses of the works of Homer, Sophocles, Plato, Hippocrates, as well as codes of the New Testament and the Fathers of the Church, many of which are richly decorated with Byzantine miniatures.

PAOLA MANONI, Biblioteca Apostolica Vaticana.
manoni@vatlib.it

Manoni, P. "Metadata framework and application profiles in the global structure of catalogs and digitization projects of the Vatican Library". *JLIS.it*. Vol. 4, n. 1 (Gennaio/January 2013): Art: #5516. DOI: [10.4403/jlis.it-5516](https://doi.org/10.4403/jlis.it-5516). Web.

ABSTRACT: This paper will focus on the application profiles recently implemented in the new Vatican Library's discovery tool that interacts with interoperability standards and manages different metadata. The presentation will also take into consideration what the Library is planning for accessing web-based digitized manuscripts collections..

KEYWORDS: Vatican Library-Manuscripts collections; Library linked data

Submitted: 2012-06-01

Accepted: 2012-08-31

Published: 2013-01-15

