Biblissima's Choices of Tools and Methodology for Interoperability Purposes

Biblissima: selección de herramientas y de metodología para fomentar la interoperabilidad

Eduard Frunzeanu / Régis Robineau / Elizabeth MacDonald[*]
Biblissima
Cité des Humanites et des Sciences Sociales
Campus Condorcet. Paris. Aubervilliers

DOI: http://dx.doi.org/10.20318/cian.2016.3146

Recibido: 21-04-2016
Aceptado: 24-05-2016

Abstract: Biblissima *(Bibliotheca bibliothecarum novissima)* is a digital humanities project which aims to create a federated access point for approximately 40 partner databases dedicated to the history of manuscripts and early printed books, to their circulation and their readers, from the 8th to 18th centuries. These databases contain different data types and use different database systems, which are added complications when building a unique access point. This contribution will focus on the various challenges to be faced in standardising the data and achieving image interoperability, as well as on the technical solutions that have been adopted: choosing to define an ontology based on CIDOC-CRM and FRBRoo in order to encompass all the data types found in the partner databases, building a thesaurus for the concepts indexed in the databases (especially for scientifc terminology and iconogra-

Resumen: Biblissima (*Bibliotheca bibliothecarum novissima*) es un proyecto de humanidades digitales cuyo objeto es la creación de un punto de acceso federado para aproximadamente 40 bases de datos asociadas sobre la historia de los manuscritos y los antiguos libros impresos, su circulación y sus lectores entre los siglos VIII y XVIII. Estas bases de datos contienen datos distintos y usan sistemas de bases de datos diferentes, lo que complica la labor de construcción de un único punto de acceso. Esta contribución se centra en los desafíos varios que supone estandarizar los datos y alcanzar la interoperabilidad de las imágenes, además de las soluciones técnicas que se han adoptado. Dichas soluciones incluyen la selección de una ontología basada en CIDOC-CRM y FRBRoo con el fin de abarcar todos los tipos de datos que se encuentran en las bases de datos asociadas; la construcción de un diccionario de

[*] eduard.frunzeanu@biblissima-condorcet.fr; regis.robineau@biblissima-condorcet.fr; elizabeth.macdonald@biblissima-condorcet.fr.

phic descriptors) and an XML-TEI authority file for all non-concept data. The article will also present the tools and methods used to align different data types (people, corporate bodies, places, etc.) with international resources such as the Bibliothèque nationale de France authority file, VIAF, DBpedia and GeoNames, as well as the standards adopted for image interoperability and the scientific contributions that image annotation tools based on IIIF standards can provide.

Keywords: cultural heritage, manuscripts, early printed books, Middle Ages, Renaissance, Early Modern Period, semantic web, data interoperability, ontology, data alignment, thesaurus, Linked Open Data, image interoperability, IIIF, Mirador, image viewer.

sinónimos para los conceptos indexados en las bases de datos (en especial para la terminología científica y los descriptores iconográficos); y un archivo de autoridad XML-TEI para los datos no conceptuales. El artículo también introduce los instrumentos y métodos empleados para aliñar diferentes tipos de datos (gente, corporaciones, lugares, etc.) con recursos internacionales como el archivo de autoridad de la Biblioteca nacional de Francia, VIAF, DBpedia y GeoNames, así como los estándares adoptados para la interoperabilidad de las imágenes y de las contribuciones científicas que las herramientas de anotación de imágenes basadas en estándares de IIIF pueden proporcionar.

Palabras clave: Patrimonio cultural, manuscritos, antiguos libros impresos, Edad Media, Renacimiento, temprana Edad Moderna, web semántica, interoperabilidad de datos, ontología, alineación de datos, diccionario de sinónimos, datos abiertos enlazados, interoperabilidad de imagen, IIIF, Mirador, visualizados de imágenes.

Biblissima (Bibliotheca bibliothecarum novissima) is an observatory for medieval and Renaissance written cultural heritage, developed with funding from the French government programme Equipements d’excellence, and is part of the Investissements d’avenir[1]. Led by the Campus Condorcet, the Biblissima project brings together eight French partner institutions in the fields of research, teaching and cultural heritage, including the BnF (National Library of France) and the IRHT (Institut de recherche et d’histoire des textes). One of the aims of the project is to create a federated access point[2] for almost 40 partner resources[3] dedicated to the history of manuscripts, libraries and

[1] The project was designed and is directed by Anne-Marie Turcan-Verkerk (EPHE, Paris). Further information about the project, its executive committee and its scientific advisory board can be found on Biblissima’s website: http://www.biblissima-condorcet.fr

[2] This component of the observatory is carried out by a team supervised by Matthieu Bonicel (BnF) and composed of Elizabeth MacDonald (Programme Coordinator), Régis Robineau (Web Coordinator), Stefanie Gehrke (Data Coordinator), Pauline Charbonnier (Data Specialist), Eduard Frunzeanu (Expert on Authority Files), Marie Muffat (Functional Specialist), Kévin Bois (Web Developer). In this article, “we” always refers to the whole Biblissima team whose work is summarised here.

[3] The complete list and the description of these resources is available here: http://www.biblissima-condorcet.fr/en/resources/biblissima-resources

texts, whose main chronological span is the Middle Ages and the Renaissance. These resources have been developed and enriched since the 1990's, and use different database systems. They contain different types of metadata that are relevant to different research fields: codicology[4], library resource cataloguing[5], history of manuscript and book collections[6], iconography[7], prosopography[8], Latin textual corpora[9], etc. These metadata fall mainly into the categories traditionally used to index and describe documents, some of which are also used in library cataloguing: authors and other kinds of protagonists (translators, owners, scribes, illuminators, etc.), works (original works or adaptations thereof, translations or commentaries), physical objects (manuscripts, early printed books), collections (private or public), codicological and palaeographical elements (book bindings, scripts), places (of production for a manuscript or an illumination, of publication, of conservation). Several Biblissima resources contain images which have a research-oriented status (as is the case with the iconographic databases where illuminations are the main object of study), while others are documentary in nature (for example, in the codicological databases) or are surrogate, allowing the user to see a digital reproduction of the original documents. In sum, the Biblissima team must deal with a huge amount of metadata and images that are quite varied in nature and purpose.

1. Metadata Interoperability

One of the first steps in achieving interoperability between the partner resources was to analyse their data, the metadata categories they use, and the relationships between these categories. First of all, we noticed that they have different objects of study, with some being person-oriented (focussing on the identity of a person and his relationships to other people and objects), while others are object-oriented (a material item, be it a manuscript or an early

[4] Reliures (BnF, Paris) http://reliures.bnf.fr; Codicologia (IRHT, Paris) http://codicologia.irht.cnrs.fr

[5] BnF Archives et Manuscrits (Paris) http://archivesetmanuscrits.bnf.fr; Pinakes (IRHT, Paris) http://pinakes.irht.cnrs.fr

[6] Bibale (IRHT, Paris) http://bibale.irht.cnrs.fr; Esprit des livres (Ecole des chartes, Paris) http://elec.enc.sorbonne.fr/cataloguevente

[7] Mandragore (BnF, Paris) http://mandragore.bnf.fr; Initiale (IRHT, Paris) http://initiale.irht.cnrs.fr

[8] BUDE (IRHT, Paris-CESR, Tours) http://bude.irht.cnrs.fr

[9] Bibliothèques Virtuelles Humanistes (CESR, Tours) http://www.bvh.univ-tours.fr; Sermones (CIHAM, Lyon) http://www.sermones.net

printed book), event-oriented (the history of the transmission of a material item), or concept-oriented (the elements identifiable in an image). Besides these conceptual differences, there are many disparities of a technical, semantic, scientific, structural and epistemological nature.

1. Technical issues: beyond the fact that the databases use various administration systems and formats (MySQL, Access, XML-TEI, XML-EAD, UNIMARC), they do not all provide unique identifiers for each type of data. For instance, one may find databases which have created IDs for the works they have catalogued but not for their authors, or databases which have assigned IDs to the authors and their works but not to the owners to whom the books had previously belonged. As for the digital images, they are kept in silos and do not use standards that enable their sharing and reuse outside of the hosting digital repositories.
2. Semantic issues: Each database has structured its metadata and labelled the data types in its own way so that for a particular cultural entity one can find several labels: for example, the author of a text can be found labelled as 'author' or 'creator'. However, some databases have chosen to simplify the structure of their data types and, given the fact that an individual could potentially have had several cultural roles, have decided to create a broader category labelled 'person' or 'protagonist' whose various functions are then specified by assigning predefined roles ('author', 'scribe', 'translator', 'editor', 'owner' and so on). Another apparent point of difficulty with the partner resources is the fact that a given category might refer to the subject of a cultural heritage item (for example, the creator or owner: Dante, the author of the Divine Comedy) or to its object (the concept identifiable in an illumination: Dante as represented in the manuscripts of the Divine Comedy). It is a feature that one also finds in library information processing, whose cataloguing standards distinguish between named authority files and subject headings.
3. Scientific issues: some objects appear in two or more resources but they are sometimes described in different terms, and even in conflicting ways. For example, such conflicts may occur with regards to the date of production of a manuscript or the identity of a manuscript's illuminator.
4. Structural features: depending on the level of granularity defined, each database has stored information that is likely to answer several scien-

tific questions. However, none of the databases have adopted an exhaustive perspective of their object of study, with some of them leaving aside metadata like a person's sex or the language of a text. Even when such information is available, it cannot be used without caution. For instance, the category of corporate bodies labelled "previous owner" often includes religious institutions as well as old libraries: while one could indicate the sex for the first (a monastic community might have been composed of men or women), it is not relevant to do so for the second. This means that when querying the sex of former institutional owners, one must be aware of the limits to the relevance of such information and take it into account when reading and interpreting the query results. From another point of view, that of the physical description of manuscripts, the databases have rarely encoded information on any damage or mutilation sustained by the manuscripts, which complicates the study of dispersed manuscripts and fragments or, on a larger scale, that of violence to cultural heritage objects[10].

5. Epistemological aspects: since the resources do not have identical objects of study, they are unable to respond to the same kinds of queries. This has had an impact on the way different objects of research have been defined, as well as on the way data have been collected, structured, and processed in varying degrees of detail. For some databases, a person stands alone, while for others she is considered a part of a network. The description of an object is always limited to its materiality or to any physical marks that can be observed on it (with or without the help of technical devices, as is the case with erased portions of text or with palimpsests). This approach does not take into account any still visible signs of planned work that was never accomplished, as happens for instance in several illumination cycles where zones were delineated on the page for images that in

[10] For instance, in the Initiale database, there are notes about approximately 280 cut out miniatures or damaged manuscripts; however these notes are not encoded. For some of these cuttings, it was possible to find their present location and, with the technologies available nowadays, one can virtually reconstruct the original manuscript. Manuscript 5 at the Châteauroux Bibliothèque municipale provided Biblissima with an interesting case study for this kind of virtual reconstruction: half of the miniatures of this manuscript (14 of 28) were cut out at an unknown date, but fortunately 11 of them are currently held at the BnF. Once the manuscript held at Châtearoux and the miniatures held at the BnF had been digitised, it was possible to reconstruct the original manuscript using image interoperability technologies, and thus to better understand the relationship between text and illumination: http://demos.biblissima-condorcet.fr/chateauroux/

> the end were never drawn or painted. Consequently, the process of manuscript or incunabula production can only be partially studied.

These are but a few of the numerous differences or missing elements we observed when analysing the partner resources. In order to face Biblissima's challenge of achieving interoperability between these various scientific and documentary resources of metadata and images, it was necessary to propose a new structure for the data that is based mainly on the categories the resources have in common. This means, however, that it will not be possible to answer all the same kinds of questions that the project resources can answer individually[11].

As it was necessary to find technical solutions to federate different database formats and reconcile different ways of organising and indexing data, to ensure long-term usability and offer new ways of exploring the partners' resources, we chose to adopt semantic web standards. This technology implies the use of a common conceptual model for information modelling (RDF, or Reference Description Framework) that makes interoperability possible with other projects that share the same standards, and opens Biblissima's data to reuse by other projects. In addition, it allows the user to make complex queries by leveraging the graph-oriented representation of the data, which is based on a set of subject–predicate–object expressions called «triples». The SPARQL query language makes it possible to build complex searches in order to answer questions such as "all the authors active during a specific time span", "all the translations of a text that were owned by private collectors" or "all the printers who had a printing shop in more than two cities". In considering the option of using semantic web technologies, we looked into what several national libraries (for example, the Bibliothèque nationale de France, the Biblioteca nacional de España) were doing in this field. We also looked at projects like those supported by the Getty Center, Europeana, the Netherlands eScience Center or Stanford University, in addition to international projects like Sharing Ancient Wisdoms (SAWS) and Standards for Networking Ancient Prosopographies (SNAP)[12].

[11] Such is the case with prosopographical data: even if some resources have collected information of a prosopographical nature (by studying academic, genealogical or intellectual relationships), it is not encoded in the same way. Moreover, it is not possible to transfer this information to the rest of the databases as one can do in other cases, for example by adding information about a person's sex wherever it is missing. For these reasons, it will not be possible to make queries of a prosopographical nature in the Biblissima portal.

[12] Bibliothèque nationale de France (http://data.bnf.fr); Biblioteca nacional de España

In order to put into practice the semantic web standards, we have carried out three main operations: compare and map similar data types and the relationships between them, and from there develop an ontology; cluster, align and standardise the different graphical forms of the data; implement and document technologies that are compatible with IIIF specifications.

1.1. Data Model and Authority Files

As to the data categories, the Biblissima team has developed two approaches: on the one hand, we have created an ontology that is compliant with CIDOC-CRM and FRBRoo[13]. This facilitates the internal mapping of all the databases' structures to a single common model and will make Biblissima's metadata interoperable in the future with other projects using the same data model. On the other hand, we have created a common XML model that reflects the equivalent relationships between similar data types in order to facilitate the collection and the management of data from the project's partner institutions. To this end, a global category named 'participant' includes all the people and corporate bodies that are associated with a cultural heritage object and its creation, modification, annotation, sale, transmission or destruction. A particular role is used to specify each of these actions[14] in the same way libraries do with relator terms to indicate the relationship between a name and a documentary resource.

> *Author/Protagonist (author)* are mapped in the XML Biblissima model to Participant role 70 which corresponds in FRBRoo to the class *E21_Person*

(http://datos.bne.es); The Getty (http://getty.edu); Europeana (http://www.europeana.eu); Netherlands eScience Center, espeacilly the projects DIVE https://www.esciencecenter.nl/project/dive and BiographyNet http://www.biographynet.nl; Stanford University for the interesting portal, Kindred Britain http://kindred.stanford.edu; SAWS http://www.ancientwisdoms.ac.uk; SNAP https://snapdrgn.net. Many of these projects provide rich documentation on how to query their data using a specific query language (SPARQL), which is very helpful for users: The Getty Vocabularies http://vocab.getty.edu/queries and Europeana http://europeana.ontotext.com/sparql/queries are two such examples.

[13] For these two international data modelling standards see the documentation on their respective websites: CIDOC-CRM http://cidoc-crm.org, FRBRoo http://www.ifla.org/node/10171

[14] The roles are based on the list used by the BnF in its cataloging process: http://data.bnf.fr/vocabulary/roles

Concerning the metadata, the Biblissima team decided not to make the distinction that some databases make between an entity as it exists in reality and its conceptual status: this would have required that the entity be indexed both in an authority file and in a concept thesaurus, thus generating two different IDs for the same thing. In addition, we decided to build a thesaurus for the concepts indexed in the databases (especially for scientific terminology and iconographic descriptors) and an XML-TEI authority file for all non-concept data.

The various graphical forms used to record a same entity in the Biblissima databases are a very real obstacle to achieving interoperability. Even if some databases use a controlled vocabulary for indexing their resources, it is always a bespoke product that has not been shared with other projects. Yet, it is vital to identify all graphical expressions of a same item in order to retrieve all relevant results for that item. With the development of Linked Open Data (LOD) repositories, it is now possible to use national and international identifiers provided by various libraries and projects, such as the Bibliothèque nationale de France authority file, VIAF, DBpedia and GeoNames, and align our different data types (people, corporate bodies, places, etc.) with these resources. There are many advantages to linking entities in this manner: one can identify identical items whose names have different graphical forms in our metadata and cluster them around a single identifier, one becomes interoperable with other data sets that are using these same international identifiers, and one can extract other information from these repositories in order to complete and enrich one's own metadata.

Software applications like OpenRefine[15] offer many functions (key collision methods, nearest neighbour methods) that are suitable for identifying the values that might represent a same item. It is also very helpful for reconciling one's own metadata with LOD repositories and for parsing data from websites.

For example, one may have the following names in a dataset:

Louis, duc d'Orléans (1372-1407)
Louis, duque de Orléans (1372-1407)
Louis, hertog van Orléans (1372-1407)
Louis, Herzog von Orléans (1372-1407)
Louis, duc d'Orléans (1372-1407) (?)
Louis, Duke of Orléans (1372-1407)
Louis, Duke of Orléans (1372-1407) (?)

[15] More information can be found at http://openrefine.org

Using the nearest neighbour method in OpenRefine with the PPM (Prediction by Partial Matching) function, one finds that all of these forms probably represent the same person. When reconciling these names with VIAF, at least one of the forms may have a chance of being automatically detected among the alternative forms stored in this repository and it is possible to send its ID (http://viaf.org/viaf/72829747) to the other forms of the name in our data cluster. This will provide a numerical string that can be used as a control key to identify all relevant graphical forms. One could also choose to extract from VIAF the URI attributed to this person by the BnF (http://data.bnf.fr/ark:/12148/cb119426642) and thereafter extract the name heading (prefLabel) proposed by the BnF (Louis Orléans (duc d', 1372-1407)) as well as other kinds of information, such as the URIs of other LOD repositories (ISNI, Wikipedia, Idref etc.) that have already been retrieved by the BnF.

It sometimes occurs that none of our name forms are detectable by automatic reconciliation because the graphical forms are too different from those available in the LOD repositories, as is true of the following name series:

Georges I d'Amboise, aartsbisschop van Rouen (1460-1510)
Georges I d'Amboise, Archbishop of Rouen (1460-1510)
Georges I. d'Amboise, Erzbischof von Rouen (1460-1510)
Georges I d'Amboise, arquebisbe de Rouen (1460-1510)
Georges I d'Amboise, arzobispo de Rouen (1460-1510)
Georges Ier d'Amboise, archevêque de Rouen (1460-1510)

In this case, one may try to make some changes to the character strings, for instance by eliminating the descriptive text between the comma and the opening parenthesis, and trying to reconcile again. Sometimes, despite all the tricks one might imagine, there will be no results. Depending on the time one has available, one could manually search for an alignment to see what should be changed in a string in order to enhance the automatic reconciliation.

It may also happen that our metadata are not at all referenced in existing repositories. This is the case for many of Biblissima's databases because the historical people or places they are studying are not yet in the libraries' field of interest, nor in that of Wikipedia or other similar LOD datasets. The Biblissima team considered two possible solutions to this problem: create a record for the non-referenced metadata in the BnF authority file, thus obtaining a persistent URI that could be used in future by other projects; build an XML-TEI database to manage our authority file in its entirety, which allows us to define our own model for structuring the information, provides an XML identifier for each entity and makes it possible to manage them in a more flexible manner. This second solution still needs to be improved upon by creating a system of persistent URIs that can be reliably cited by other projects.

1.2. Organising the Descriptors in a Thesaurus

The situation is somewhat more complicated in the case of metadata used as descriptors in the codicological or iconographic databases that must be included in the Biblissima thesaurus. On the one hand, they are stored in more or less hierarchical thesauri with different organisation principles, while on the other there are descriptors which are also relevant to the authority file given their double status, that of real entity and that of subject of a document. To take just one example, that of geographical data, the Biblissima databases store them not only as descriptors (geographical places—be they historical, disappeared, fictional, or non-identified—that have been indexed in manuscript illuminations), but also as places of origin (city or abbey where an item was copied, edited or painted) and as holding institutions (former or currently existing archives and libraries). Furthermore, each of these three types of data are managed differently: the geographical descriptors are stored in hierarchical thesauri or flat lists with all other kind of descriptors; the places of origin are organised in a hierarchical structure whose levels include geographical areas, countries and provinces; the locations of holding institutions are related to a country, city and repository.

Combining two or more sets of geographical metadata implied the creation of a new host structure based on the organisational principles of the original databases. A first attempt at fusion was made with the geographical descriptors of two iconographic databases, Mandragore (illuminated manuscripts held at the BnF, http://mandragore.bnf.fr) and Initiale (illuminated manuscripts held in other French libraries, http://initiale.irht.cnrs.fr). Both databases have created a "Geography" category in their subject index. The hierarchy of Mandragore's thesaurus is based on the Dewey Decimal Classification (DDC), which is used to subdivide the "Geography" category into large geographical areas corresponding to the British Isles (DDC 941), to France and Monaco (DDC 944), and so on. Within these geographical areas, each geographical descriptor is provided with details of an historical nature [Ophir (unlocated country)], of physical geography [Loire (river in France)] or of administrative geography [Orléans (France, Loiret)]. In the Initiale thesaurus, the hierarchy is not as deep as in the Mandragore database: under the concept heading "Geographical Subject", one finds a semantic division (continent, geographical place, morphological geography, spatial directions); under "geographical place", there are countries and historical cities like Constantinople; at the subsequent level one finds currently existing cities. One can see that the same toponym is differently indexed by the two databases.

To create a new classification system, the Biblissima team decided to combine existing organisational principles with the classification features used by Geonames, a LOD geographical database[16]. To that end, we have aligned the descriptors of our two databases with Geonames and extracted the feature codes used by Geonames to classify different geographical entities (approximately ninety codes were relevant for our descriptor types)[17]. Seven category headings were used to structure our thesaurus: general notions, political geography, physical geography, human constructions, fictional places, non-identified places and disappeared places. For political geography, we have retained Mandragore's organisational choices based on the Dewey classification, in order to provide access to the data by geographical area. The other clusters have been organised using the Geonames feature codes: as such, the geographical entities having the codes PCLS, PCLIX, PCLI, PCLD, PCL have been classified in the cluster 'Countries'. For all the places that are not referenced by Geonames (especially former historical provinces or cities), we have manually applied a method of labelling similar to that used by Geonames in order to maintain a coherent classification system. In addition, the metadata have been enriched with geographical coordinates extracted mainly from Geonames, and have been checked and manually corrected or complemented through automatic alignment with other LOD repositories[18]. The relationships between the descriptors have been defined with the following SKOS properties: prefLabel, altLabel, broader, narrower, exactMatch, closeMatch, relatedMatch. The resulting geographical thesaurus has been integrated into Biblissima's recent prototype[19], which contains metadata from the Mandragore and Initiale databases, and is structured as follows[20]:

[16] http://www.geonames.org

[17] http://www.geonames.org/export/codes.html

[18] The BnF (http://data.bnf.fr, for Map Department metadata & Rameau), Wikipedia, and specialised repositories such as Pleiades and Trismegistos for historical places (http://pleiades.stoa.org; http://trismegistos.org) are the most representative LOD repositories.

[19] The prototype can be accessed at this address: http://demos.biblissima-condorcet.fr/prototype

[20] The prototype's geographical thesaurus is temporarily available at this address: http://nossl.demo.logilab.fr/biblissima/ConceptGroup

I. General notions

II. Political geography (based on feature codes of Geonames)

  A. Geographical areas (= Dewey classification)

  A.1. Countries

  A.1.1. Counties

  A.1.1.1. Cities

  A.2. Ancient cities and provinces

III. Physical geography (based on feature codes of Geonames)

  A. Continents

  B. Islands & Peninsulas

  C. Deserts & Oasis

  D. Rivers, Lakes, Seas

  E. Mountains & Volcanos

  F. Forests & Parks

IV. Human constructions (based on feature codes of Geonames)

  A. Monasteries

  B. Castles & palaces

  C. Religious sites

  D. Bridges

  E. Towers & fortresses

V. Fictional places

VI. Non-identified places

VII. Disappeared places

This geographical thesaurus will be included in a greater one, which will contain all the descriptors, specialised terminology and other kinds of concepts that can be found in the partner resources. It will be used in the search engine for the Biblissima portal and will form the basis for navigation by facets.

Each term in both the authority files (including participants, works, repositories, etc.) and the thesaurus will be assigned a persistent URI. Once they become publicly available, they can be reused by other projects and cited in online publications. The persistent URIs will also make it possible to directly query the SPARQL endpoint of the Biblissima portal.

2. Image Interoperability and the Contribution of Image Annotation for Prosopographical Purposes

Besides achieving interoperability for metadata from various specialised databases, another challenge we face in the Biblissima project is that of providing a single access point for a large number of digital facsimiles from three different image repositories: Gallica, the BVMM ("Virtual Library of Medieval Manuscripts"), and the BVH ("Virtual Humanist Libraries")[21].

Our approach to bring together the images from these digital libraries is based on emerging interoperable technology called the "International Image Interoperability Framework" (IIIF)[22]. It is becoming a de facto standard for image delivery and access on the Web. This technology defines a set of APIs based on an underlying data model called Shared Canvas[23].

Mirador[24] is one of the leading image viewing applications that implements the IIIF protocols. It provides a workspace to zoom, compare, annotate and share image-based resources. It is open-source and developed mainly at Stanford and Harvard Universities. One of its most attractive features is that it allows comparison of images from repositories dispersed around the world in a multi-window workspace. It also comes with a set of tools to annotate images in accordance with the W3C Open Annotation standard[25]. In these respects, Mirador is more than a simple image viewer and may be seen as a major component of a feature-rich scholarly workspace.

A tool like Mirador has great potential to enhance digital practices for research purposes and could fulfill a wide range of scholarly use cases. Some of its features might improve the content of prosopographical metadata by associating the record of a person with visual elements that are likely to help identify him. This is a valuable contribution as it is known that the identity of an individual or of a corporate body was often linked with visual signs (seals, coats of arms, ex-libris, etc.), especially during the Middle Ages and the Renaissance. For instance, Mirador's annotation feature can be used to study autograph handwriting in a set of digitised primary sources in order to:

[21] Gallica, bibliothèque numérique de la Bibliothèque nationale de France: http://gallica.bnf.fr

BVMM, Bibliothèque Virtuelle des Manuscrits Médiévaux (Institut de Recherche et d'Histoire des Textes, CNRS): http://bvmm.irht.cnrs.fr

BVH, Bibliothèques Virtuelles Humanistes (Centre d'Etudes Supérieures de la Renaissance, CNRS-Université François-Rabelais de Tours): http://cesr.cnrs.fr

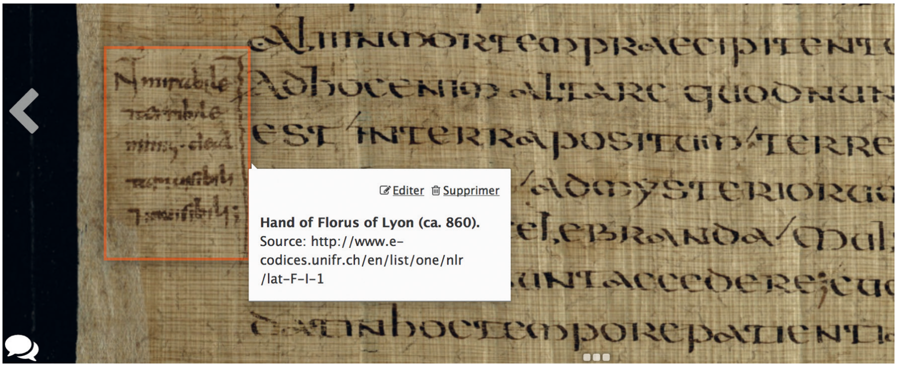
[22] International Image Interoperability Framework (IIIF): http://iiif.io

[23] *"The SharedCanvas data model specifies a linked data based approach for describing digital facsimiles of physical objects in a collaborative fashion. It is intended for use in the cultural heritage domain, although may be useful in other areas, and is designed around requirements derived from digitized text-bearing objects such as medieval manuscripts."* (http://iiif.io/model/shared-canvas/1.0/)

[24] Mirador : *"Open-source, web based, multi-window image viewing platform with the ability to zoom, display, compare and annotate images from around the world"* (http://projectmirador.org)

[25] Open Annotation Data Model (http://www.openannotation.org/spec/core/)

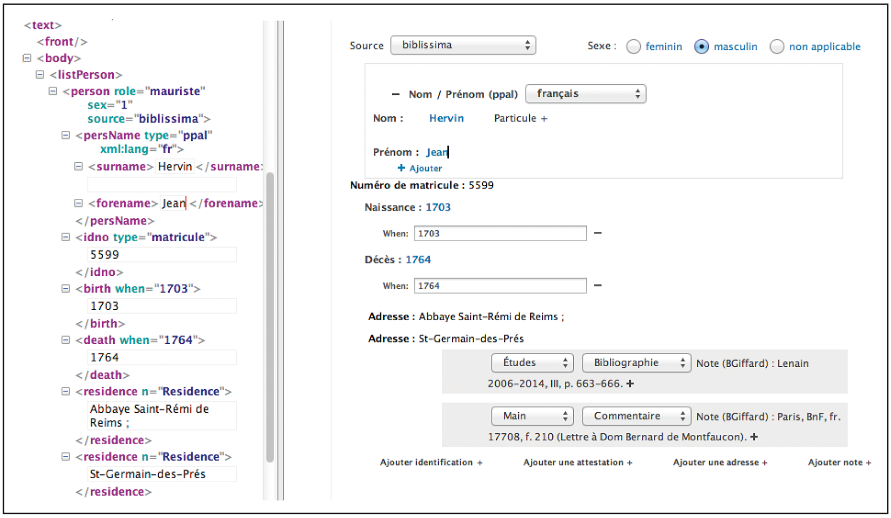
- trace, identify and index an author's personal annotations (marginalia), as has been highlighted by an experimental study of Florus of Lyon's annotations[26]



Hand of Florus of Lyon (ca. 860).
Source: http://www.e-codices.unifr.ch/en/list/one/nlr/lat-F-I-1

Fig. 1. Marginalia by Florus of Lyon in the manuscript St Petersburg, National Library of Russia, L.F.papyr. L.1, b

- create a database of autograph writings or ex-libris in order to help identify scribes, writers or owners



```
<text>
  <front/>
  <body>
    <listPerson>
      <person role="mauriste"
          sex="1"
          source="biblissima">
        <persName type="ppal"
            xml:lang="fr">
          <surname> Hervin </surname>

          <forename> Jean </forename>
        </persName>
        <idno type="matricule">
          5599
        </idno>
        <birth when="1703">
          1703
        </birth>
        <death when="1764">
          1764
        </death>
        <residence n="Residence">
          Abbaye Saint-Rémi de
          Reims ;
        </residence>
        <residence n="Residence">
          St-Germain-des-Prés
        </residence>
```

Source: biblissima

Sexe : ○ feminin ● masculin ○ non applicable

– Nom / Prénom (ppal) français

Nom : Hervin Particule +

Prénom : Jean

+ Ajouter

Numéro de matricule : 5599

Naissance : 1703

When: 1703 –

Décès : 1764

When: 1764 –

Adresse : Abbaye Saint-Rémi de Reims ;

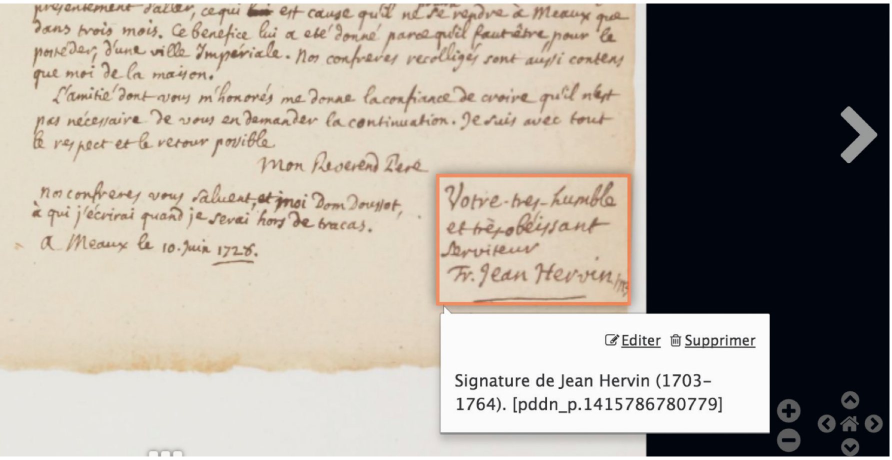
Adresse : St-Germain-des-Prés

Études · Bibliographie · Note (BGiffard) : Lenain 2006-2014, III, p. 663-666. +

Main · Commentaire · Note (BGiffard) : Paris, BnF, fr. 17708, f. 210 (Lettre à Dom Bernard de Montfaucon). +

Fig. 2.1. Note about an autograph letter by Jean Hervin in the manuscript Paris, BnF Français 17708, folio 210r (user interface of XMLMind XML Editor used by researchers to edit old catalogues and inventories).

[26] More details on this topic can be found on the Biblissima Demos website: http://demos.biblissima-condorcet.fr/florus



présentement d'aller, ce qui ~~[illegible]~~ est cause qu'il ne se rendre à Meaux que dans trois mois. Ce benefice lui a eté donné parce qu'il faut être pour le posseder, d'une ville Imperiale. Nos confreres recolliges sont aussi contens que moi de la maison.

L'amitié dont vous m'honorés me donne la confiance de croire qu'il n'est pas nécessaire de vous en demander la continuation. Je suis avec tout le respect et le retour possible

Mon Reverend Pere

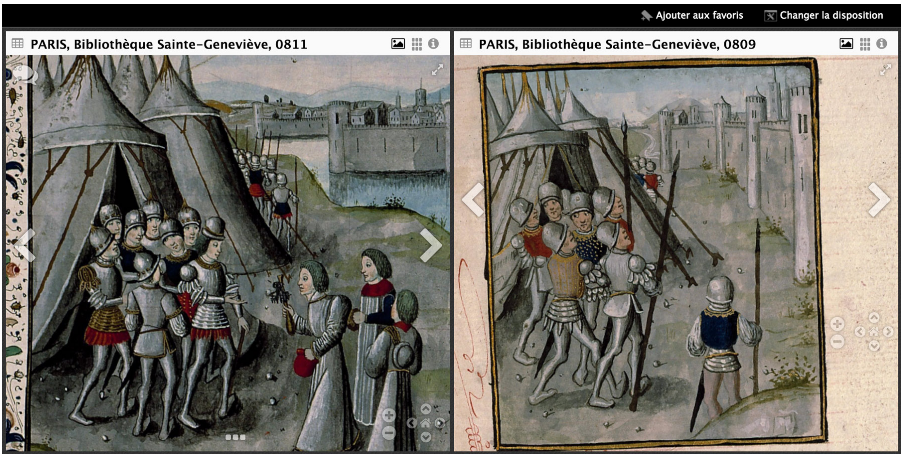
Votre tres humble
et tres obeissant
serviteur
Fr. Jean Hervin

Nos confreres vous saluent, ~~et~~ moi Dom Doussot, à qui j'écrirai quand je serai hors de tracas.
a Meaux le 10. Juin 1728.

Signature de Jean Hervin (1703–1764). [pddn_p.1415786780779]

Fig. 2.2. Image of folio 210r in the manuscript BnF Français 17708, annotated in Mirador: the annotation identifies the hand of the letter and links it to the entry "Jean Hervin" in the prosopographical index hosted by the University of Caen, France.

This annotation feature makes it possible for the user to save the selected image detail together with its context, both visual and textual, unlike other existing tools, which allow one to select, cut out and export a detail from an image without retaining anything of the original context.



Ajouter aux favoris
Changer la disposition
PARIS, Bibliothèque Sainte-Geneviève, 0811
PARIS, Bibliothèque Sainte-Geneviève, 0809

Fig. 3. Two images of miniatures displayed side by side in the Mirador workspace. On the left: Paris, Bibliothèque Sainte-Geneviève, ms. 811, f. 005 : Reddition de Valenciennes à Herman, comte de Mons. On the right: Paris, Bibliothèque Sainte-Geneviève, ms. 809, f. 317 : Siège de Mayence par les Romains. The two miniatures are attributed to the circle of Willem Vrelant.

Furthermore, Mirador's multi-window system and deep zoom capability make it easier to compare stylistic features, in illuminated manuscripts for example. The ability to inspect an image in detail and view multiple images side by side (potentially coming from different repositories) could be highly useful in comparing the iconographic representation of different objects, studying anachronistic representations, and could help to better identify artists or workshops and to evaluate attributions made by other scholars.

The methodology, modelling solutions and technologies that were briefly presented in this article are suitable for achieving interoperability between Biblissima's partner resources for both metadata and images, provided that some changes are made to the metadata to standardise and enrich them, and that technical standards for image delivery are implemented. In order to create a federated access point for all these resources it was necessary to determine what kinds of metadata can be made interoperable, to consider the queries one would like to make, to design a new data model, and to define a new kind of relationship between metadata and images that is often absent in the original resources. The textual transcription and description of a document stored in a database can henceforth be verified by directly checking the digital reproduction of the document. Technical features such as image annotation may be linked to metadata in order to better document the identity of a person or a descriptor. In this way, commentary about visual elements may be linked to the image itself and point precisely to the specific detail in question. The data model will allow more complex queries and give access to various visual representations of the results. This entire process of building interoperability is meant to accompany scholars, students and curators in their research by facilitating access to information that has been stored for decades in separate scientific databases, and has the varied goals of proposing new ways to query data, raising doubts about data and calling scientific methods into question, bringing new research topics to light, and arousing new interest in cultural heritage.

Webography

Biblissima Demos, virtual reconstitution of the manuscript Châtearoux, BM 5: http://demos.biblissima-condorcet.fr/chateauroux

Biblissima Demos, Florus of Lyon's manuscripts: http://demos.biblissima-condorcet.fr/florus

Biblissima Demos, the Mandragore-Initiale prototype: http://demos.biblissima-condorcet.fr/prototype

Biblissima's geographical thesaurus: http://nossl.demo.logilab.fr/biblissima/ConceptGroup

Biblissima resources: http://www.biblissima-condorcet.fr/en/resources/biblissima-resources

Biblissima website: http://www.biblissima-condorcet.fr

Biblioteca Nacional de España, semantic portal: http://datos.bne.es

Bibliothèque nationale de France (BnF), Archives et Manuscrits (catalogue): http://archivesetmanuscrits.bnf.fr

Bibliothèque nationale de France (BnF), Gallica (digital library): http://gallica.bnf.fr

Bibliothèque nationale de France (BnF), Mandragore: http://mandragore.bnf.fr

Bibliothèque nationale de France (BnF), Reliures: http://reliures.bnf.fr

Bibliothèque nationale de France (BnF), semantic web portal: http://data.bnf.fr

Bibliothèque nationale de France (BnF), semantic web vocabularies: http://data.bnf.fr/vocabulary/roles

Centre d'Études Supérieures de la Renaissance (CESR, Tours), Bibliothèques Virtuelles Humanistes: http://www.bvh.univ-tours.fr

CIDOC-CRM: http://cidoc-crm.org

CIHAM - Histoire, archéologie, littératures des mondes chrétiens et musulmans médiévaux (Lyon), Sermones: http://www.sermones.net

Ecole des chartes (ENC, Paris), Esprit des livres: http://elec.enc.sorbonne.fr/cataloguevente

Europeana: http://www.europeana.eu

Europeana, SPARQL user guide: http://europeana.ontotext.com/sparql/queries

Geonames: http://www.geonames.org

Geonames, list of feature codes: http://www.geonames.org/export/codes.html

Getty Center: http://getty.edu

Getty Center, SPARQL user guide: http://vocab.getty.edu/queries

IFLA, FRBRoo: http://www.ifla.org/node/10171

IIIF, International Image Interoperability Framework: http://iiif.io

Institut de recherche et d'histoire des textes (IRHT, Orléans-Paris), Bibale: http://bibale.irht.cnrs.fr

Institut de recherche et d'histoire des textes (IRHT, Orléans-Paris), BUDE: http://bude.irht.cnrs.fr

Institut de recherche et d'histoire des textes (IRHT, Orléans-Paris), BVMM - digital library: http://bvmm.irht.cnrs.fr

Institut de recherche et d'histoire des textes (IRHT, Orléans-Paris), Codicologia: http://codicologia.irht.cnrs.fr

Institut de recherche et d'histoire des textes (IRHT, Orléans-Paris), Initiale: http://initiale.irht.cnrs.fr

Institut de recherche et d'histoire des textes (IRHT, Orléans-Paris), Pinakes: http://pinakes.irht.cnrs.fr

Mirador, image viewing platform: http://projectmirador.org

Netherlands eScience Center, BiographyNet: http://www.biographynet.nl

Netherlands eScience Center, DIVE+: https://www.esciencecenter.nl/project/dive

Open Annotation Data Model: http://www.openannotation.org/spec/core

OpenRefine: http://openrefine.org

Pleiades: http://pleiades.stoa.org

Shared Canvas: http://iiif.io/model/shared-canvas/1.0

Sharing Ancient Wisdoms (SAWS): http://www.ancientwisdoms.ac.uk

Standards for Networking Ancient Prosopographies (SNAP): https://snap-drgn.net

Stanford University, Kindred Britain: http://kindred.stanford.edu

Trismegistos: http://trismegistos.org