Biblissima's Prototype on Medieval Manuscript Illuminations and their Context

Stefanie Gehrke, Eduard Frunzeanu, Pauline Charbonnier, and Marie Muffat

Equipex Biblissima, Campus Condorcet, Paris, France
{stefanie.gehrke,eduard.frunzeanu,pauline.charbonnier,marie.muffat}@biblissima-condorcet.fr
http://www.biblissima-condorcet.fr

Abstract. Biblissima is an online digital library, which provides easy and coordinated access to a huge and complex mass of documentation on manuscripts and early printed books, the texts contained therein, their circulation and their readers, from the 8th to 18th centuries. This workshop presentation will give an overview of the steps followed and decisions made along the way to releasing a first prototype of the Biblissima portal: from mapping data to a common ontology, via the establishment of a thesaurus, to the technical development of a single interface and a common triple store for data deriving from different iconographic databases on medieval manuscripts.

Keywords: cidoc crm · frbroo · medieval manuscript · illumination · interoperability · descriptors · thesaurus · historical place names · semantic web · linked data · library · Middle Ages · Humanism · Renaissance

1 Objectives of the Biblissima Observatory

Biblissima - Bibliotheca bibliothecarum novissima - is an observatory for the written cultural heritage of the Middle Ages and the Renaissance, developed through the French government programme *Équipements d'excellence*, part of the *Investissements d'avenir* [1]. The observatory focusses on documents written in the main languages of culture in Medieval and Renaissance Europe (Arabic, French, Greek, Hebrew, Latin, etc.) and contributes to a better understanding of the circulation of texts, the evolution of libraries and the transmission of knowledge in Europe from the 8th to the 18th century.

In addition to its contributions to research, Biblissima plays an important role in disseminating knowledge about the written cultural heritage of the Middle Ages and the Renaissance to the widest possible audience.

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).

The two main components of the observatory are a cluster of the project's data on manuscripts and early printed books currently found in as many as 40 databases in different formats and with different research interests (including illuminated manuscripts, history of the transmission of texts and history of collections) and a digital image repository. The databases will be interconnected using semantic web technologies and linked to a platform for digital editions and to the project's digital image repository.

1.1 Semantic Web Solutions for Historical Data

In order to handle the heterogeneity of the database formats (MySQL, EAD, TEI P5, UNIMARC, etc.) and the variety of Biblissima's data (manuscript cataloguing databases, textual editions, iconographic databases) we have chosen to use the *CIDOC Conceptual Reference Model* (Comité International pour la Documentation Conceptual Reference Model [2]) and *FRBRoo* (Functional Requirements for Bibliographic Records object oriented [3]) as framework for a project-specific extension of those ontologies that facilitates the internal mapping to a single common model and allows the partners to expose their data in RDF compliant to a globally established standard.

CIDOC CRM is an accepted ISO standard (ISO 21127). As an event-centric ontology it covers different phenomena in space and time like provenance, copying of texts, creation of works and expressions, as well as the production of information carriers and attribute assignments. As CIDOC CRM and FRBRoo (which combines the CIDOC CRM approach with the common vocabulary for the transmission of works (WEMI) that is provided by the FRBR model) are generic models for the museum and library domains, it was decided to define a few more specific classes and properties related to manuscripts, early printed books and illuminations. For example, within the scope of the Biblissima project a medieval manuscript is an instance of the class bibma:Manuscript, which is a subclass of `frbroo:F4_Manifestation_Singleton` ("This class comprises physical objects that each carry an instance of F2 Expression, and that were produced as unique objects, with no siblings intended in the course of their production"). An instance of a `bibma:Manuscript` might be composed of several parts (`bibma:Component`) and might carry both text and illustrations.

As regards the illustrations, there are several possible modelling solutions in CIDOC CRM, such as E38 Image ("This class comprises distributions of form, tone and colour that may be found on surfaces such as photos, paintings, prints and sculptures or directly on electronic media") or its subclass E36 Visual Item ("This class comprises the intellectual or conceptual aspects of recognisable marks and images"). These solutions have been adopted both for book illustrations by the "*Illustrations of Goethe's Faust*" project [4] and for maps by the "*Carte de la nouvelle frontire Turco-Grecque*" project [5]. In order to model the illumination genre, we decided instead to define an illumination as an instance of a class called `bibma:Illumination`, which is a subclass of E26 Physical Feature

('This class comprises identifiable features that are physically attached in an integral way to particular physical objects"). The following RDF triple expresses this relationship.

```
:c a bibma:Component .
   :i a bibma:Illumination .
   :c crm:P56_bears_feature :i .
```

This is a shortcut for the fully developed path:

```
:folio a crm:E53_Place .
   :c a bibma:Component ;
   crm:P59_has_section :folio .
   :i a bibma:Illumination ;
   crm:P53_has_former_or_current_location :folio .
```

Instances of E53 Place are a folio or a particular zone on a folio, for example.

The ontology interacts with a thesaurus of technical terms used in medieval studies (codicology, palaeography, iconography etc.) and descriptors used for indexing medieval illuminations in the project's databases. The data is structured in a thesaurus compliant with the international standard for thesauri and interoperability with other vocabularies (ISO 25964). The different lexical and semantic relationships that can be defined between the descriptors will have an intrinsic (semantic) role, in that they will help to show the relationships of hyponymy or synonymy, as well as an extrinsic (technical) function for the search engine. In addition, the project's data on people, corporate bodies, places and titles are aligned with existing authority files and linked data repositories, such as Rameau, VIAF, and GeoNames.

1.2 The Historical Dimension of Biblissima's Data

The majority of Biblissima's databases contain descriptive and structural metadata for medieval manuscripts and early printed books, issued from the cataloguing of these documents or scientific research, but the project also includes digital editions in TEI P5 of library inventories and texts and records on illuminations. Metadata like the date of creation and place of origin of a manuscript and its illuminations, the identification of the scribe, translator or commentator of the copied text, former owners (people and corporate bodies) of a manuscript throughout the centuries and lists of books kept in libraries at a given moment in time can be used to study the history of the texts and manuscripts, as well as reading and collecting practices.

In order to develop the portal step by step we have chosen to begin by creating a unified access point to two iconographic databases.

2 Objectives of the Biblissima Prototype

Using Semantic Web technologies, the Biblissima prototype aims to demonstrate the potential of the available metadata produced by the Biblissima project. It was developed using open source solutions and all the data is publicly available under an open licence in order to facilitate reuse.

The prototype is built on a subset of two iconographic databases: *Mandragore* [6], the database of the Department of Manuscripts of the National Library of France (BnF) and *Initiale* [7], the database of the IRHT.

It provides federated access to a subset of data present in the two databases, such as illumination related data: caption, descriptor, folio carrying the illumination, illumination record, digital surrogate of the illumination, artist, context of the illumination (author and title of the textual work per artistic unit), date of origin and place of origin. The data set also contains manuscript related data such as shelfmark, common name, grouping, repository, digital surrogate, manuscript record.

The subset is limited, in the case of both databases, to records on illuminations indexed with at least one geographical descriptor, which equates to almost 5 000 descriptors for approximately 20 000 illuminations.

A SPARQL endpoint and a federated search engine make it possible to search all the data in the cluster. Users can search by descriptor, artist, date of origin, place of origin, author or work title, and can refine their search with a series of facet filters. The results are displayed in a user-friendly manner by grouping them in lists. Pages about manuscripts, their units and illuminations include frames that display the corresponding digital surrogate using *IIIF manifests* [8], relating text and images. Other visualisation features are available to the user, such as timelines and maps. The data from *Initiale* and *Mandragore* are augmented with data on the digital surrogates of illuminations in their context, extracted from other manuscript catalogues (*Medium* [9] - the IRHT manuscript repository, Gallica - the digital library of the BnF, and BnF archives et manuscrits [10] - the catalogue of the Department of Manuscripts of the National Library of France). Each illumination record in the prototype links back to the original record from one of the two databases as well as to the full digitisation of the manuscript when available. The search capabilities currently do not include manuscript genealogies. This can be achieved by including more databases and classes like `bibma:Type_of_Use_Manuscript` and `bibma:Source` in the future, when implementing solutions deriving from lessons learned about texts and their transmission.

Both databases have been used to index manuscript illuminations for the last 25 years and different systems were chosen for structuring the descriptors. A polyhierarchical classification of Biblissima's thesaurus may make it possible to retain the original descriptor classifications while reordering them in a new systematisation. However, these classifications do not reflect the medieval practices of organising knowledge. The identification of the iconographic elements is

sometimes based on internal information such as heading titles, chapter titles, inscriptions or notes present in the manuscript; in the absence of this kind of information, identification is dependent on the scholar's culture, especially in the case of living things and artefacts. This means that when using these descriptors to study medieval illuminations, we must keep in mind that their identification has a disparate chronological and cultural origin. Another feature of the indexing practices specific to these databases is that the data does not provide co-textual information, and the contextual information is not always available. This makes quite difficult to trace the diachronic evolution of the meaning of a word and of its iconographic representation.

The technical solutions adopted by Biblissima open new avenues and yield new ways of searching through data that could contribute to the analysis of iconographic representations. On the basis of the geographic descriptors, one could attempt to answer several questions regarding the status of cities in artistic imagery and define the notion of the city through iconographic choices: what are the criteria which confer an urban identity to a community and what makes the difference between an urban and a rural environment? From what point in time do cities begin to be represented and what cities are the most represented over the centuries? How could one explain the cases of single occurrence? Could one analyse the anachronistic representations of places, be they cities or battlefields? What are the most common descriptors associated with toponyms?

3 Conclusion

The semantic web solutions that Biblissima has chosen could be adapted in order to provide answers to other kinds of research topics. As such, Biblissima's poly-hierarchical thesaurus makes it possible to establish new classifications of the descriptors that already exist in the databases and to recreate a medieval taxonomy of living species as it was conceived by an encyclopedist or a physician, for example. One might also connect the thesaurus to the digital edition of exegetical texts such as the biblical Glossa [11], one of Biblissima's partner projects, and try to study the semantic relations between the four senses of the Scripture (historical, allegorical, tropological and anagogical) and the iconographic representation of the biblical words and scenes, for example.

By adopting common standards for the ontology and for the thesaurus, Biblissima's data might also be aggregated with and used by other semantic web projects in the future.

References

1. Investissements d'avenir (CGI), ANR-11-EQPX-0007 `http://investissement-avenir.gouvernement.fr`
2. CIDOC CRM 6.0, `http://cidoc-crm.org/docs/cidoc_crm_version_6.0.pdf`

3. FRBRoo 2.2, http://www.ifla.org/files/assets/cataloguing/frbr/frbroo_v2.2.pdf
4. Abrami, G., Freiberg, M., Warner, P.: Managing and Annotating Historical Multimodal Corpora with the eHumanities Desktop - An outline of the current state of the LOEWE project Illustrations of Goethe's Faust. In: Historical Corpora, pp. 353 – 363. Narr Francke Attempto, Tübingen (2015)
5. Gkadolou, E., Stefanakis, E.: A formal ontology for historical maps, http://galaxy.hua.gr/~heraclitus/images/gkadolou/3_Gkadolou_ICA.pdf (2013)
6. Mandragore, http://mandragore.bnf.fr
7. Initiale, http://initiale.irht.cnrs.fr/accueil/index.php
8. IIIF, http://iiif.io/
9. Medium, http://medium.irht.cnrs.fr/
10. BnF, Archives et Manuscrits, http://archivesetmanuscrits.bnf.fr
11. Biblical Glossa, http://www.glossae.net/