

# From Parchment to the Network: Manuscripts Management and Cataloguing

Maria A. Alberti <sup>(\*)</sup>, Giliola Barbero and Cesare Pasini <sup>(#)</sup>

<sup>(\*)</sup> Dipartimento di Scienze dell'Informazione, Università degli Studi di Milano

E-mail: alberti@dsi.unimi.it

<sup>(#)</sup> Biblioteca Ambrosiana, Milano

E-mail: cpasini@ambrosiana.it, gbarbero@hotmail.it

## ABSTRACT

Conservation libraries, which hold manuscripts and are willing to provide catalogues, face the problem of their diversity: each manuscript, document or archive has its own features, due to its history. Information technology helps librarians both to record ancient documents descriptions, dealing with highly non-standard data, and to allow their dissemination. Biblioteca Ambrosiana collaborates with DSI of the Università degli Studi di Milano to find solutions to a variety of problems related to management, catalogue and description of these documents. In this paper we present some results: a topographical inventory and a filter to interface and make compatible archives of manuscript documents in *Manus*, a relational DB distributed in Italy, and XML descriptions, adopting the DTD for manuscripts defined by the EU Project MASTER.

**KEYWORDS:** Digital library, Manuscript catalogues, XML

## INTRODUCTION

Manuscripts collections preserve the entire knowledge of ancient western culture from Greek to Latin literature, philosophy and history. Also medieval vernacular culture and its products have survived, before the age of print, thanks to patient and often unknown copyists' work. We can access our cultural roots only by reading the manuscripts and the comments that have been written on them during the years, often in different temporal layers. Codices, closed in libraries, archives and private collections have been waiting for hundreds of years their liberators, as if they were in prisons. The word *prison* has been used for the first time by an Italian humanist, Poggio Bracciolini in 1416, to refer to manuscripts' confinement, and in a way it sounds right also nowadays. After visiting St. Gall's monastery, where he found a complete copy of Quintiliano's *Institutio oratoria*, Poggio wrote: *Erant enim non in bibliotheca libri illi, ut eorum dignitas postulabat, sed in teterrimo quodam et obscuro carcere*. Still at present many codices are preserved but left in a state of forgetfulness. Sometimes we don't know the exact contents of these treasures because they have not been studied; and even if they are, sometimes we have lost the information elaborated in the past by scholars who worked on them. Each manuscript copy of a work is indeed different and bears a unique witness of a particular stage of the tradition, due to the author himself, to the copyist or to some reader. Therefore philologists need to

take all copies into account: they need to compare and collate every existing manuscript copy to edit a critical text of a given work.

This is the first reason why it is very important to catalogue manuscripts and their contents. If we want to come in touch with manuscripts before moving to the institutions where they are preserved, we must read all ancient and modern printed catalogues describing libraries' collections. Even these catalogues are not easily found but in few institutions. That's why information about manuscripts is often available only to scholars who can travel and move to libraries reading rooms. Information technology (IT) can help librarians and scholars to change this situation.

Biblioteca Ambrosiana, which preserves since the beginning of XVII century about 15.000 manuscripts and 10.000 documents on parchment [8], at the end of a complete restoration asked the Dipartimento di Scienze dell'Informazione (DSI) of the Università degli Studi di Milano to introduce IT in the library, particularly in relation to the management and cataloguing of ancient funds. Some efforts have been necessary to establish the communication between specialists and library's employers, not yet involved in IT. This phase was conceived as a necessary training, in order to manage all the human resources involved in the process. The collaboration that followed provided a fruitful ground for further developments.

Indeed, from the co-operation has descended a research activity concerning manuscripts cataloguing, with particular attention to XML, developed in connection with different national and European institutions.

## MANAGEMENT OF ADMINISTRATIVE INFORMATION

Libraries already make use of IT to solve administrative tasks: such as loans management or library memberships. All the software products so far have a target of printed material more than manuscripts documents. These type of documents require different services. For instance, loans are not granted to individuals but to institutions for exhibitions; sometimes, for conservation concerns, digital or microfilmed copies are offered to the public in substitution of original documents; often, individuals are asked to justify their manuscripts requests and it may be useful to keep track of these users. These administrative tasks are simple in principle but traditionally require heavy duplication of procedures and back in 1997 no software system was known devoted to management of manuscript documents. Rarely, for instance, libraries have

topographical lists of the ancient documents they own, which would allow to manage, by their shelf-mark, any movement that occurs to the single item: loan for exhibition, restoration, reproduction on microfilm or digitization. To keep track of all these procedures a curator, traditionally, at each step has to fill a form, write a card or several sheets with all the data required.

### **A topographical list for Ambrosiana**

A topographical list of manuscripts has been implemented, with the aim to fulfill users and librarian's needs to save time and to ease library's management. It is the list of ancient documents preserved by Ambrosiana, where each manuscript is denominated by its shelf-mark, and shelf-marks are listed as they are *topographically* distributed in the library's rooms. After restoration all ancient funds have been checked, item-by-item. Cataloguers added information about custodial, restoration status and existing reproductions have been recorded. In the context of Italian conservation libraries such an electronic topographical list introduces an innovation.

The electronic topographical list is based on an object-oriented database, chosen because it allows producing an open archive, where many sorts of documents can be represented and related in different ways, by using the polymorphism and inheritance properties of object-oriented paradigm.

In this archive manuscripts, identified by their current shelf-mark, constitute objects holding data, such as: location in the library (room, floor, shelf number), reproduction (microfilm, CD-ROM), status of restoration (restored or not restored), custodial status (preserved in modern conservation box or not), availability (manuscripts are not available when they are sent to restoration or exhibition).

The database is also a reference archive for library's users. When they join the library, they receive an identification number, which they will use for any later consultation. Users have no more to fill forms, which usually implies writing two or three copies of the same data like name, surname and address. Now people digit only their identification number and the shelf-mark they are asking for. The query output tells if the manuscript is available, suggests to use microfilm reproduction if the codex is too much damaged, pretends a special justification for studying a particular manuscript, if it is a very precious one.

The system provides some useful features: it recognizes the most typical mistakes in the shelf-mark digitations, as omissions of some parts or of punctuation marks. For example, if the right shelf-mark is *A 1 inf.*, the system accepts input also in different formats, such as *A. 1. inf.* or *a. 1 Inf.* This is most important because in the handwritten copies of catalogues, known to the public, the format may not have been used consistently. Moreover, if an inexistent shelf-mark is being given in input, the system suggests similar shelf-marks present in the database<sup>1</sup> [6].

<sup>1</sup> The system has been recently further developed by the

### **MANUSCRIPT DESCRIPTION**

A second important task of IT is to support catalogue management and publishing. There were many efforts to standardize manuscript description even before introduction of IT. Several scholars tried to codify the information necessary to give a complete description of each codex, but in practice cataloguers can find new type of data at any new manuscript description. This happens because manuscripts are documents produced by human hand, not by machine, and their typologies depend directly upon each producer. In printed catalogues data difficult to classify can be easily expressed in prose and may be found only in few cases: any record contains, i.e. the physical support or the number of folios, but other information can be found only in few records. Just to give an example:

*Membr., ff. IV + 121 + IV'; mm.183x115 (124x80), 26 long lines.*

A line like the above holds typical information about physical description of the codex and almost any cataloguer places them in the same order in a record. More difficult is to give a codified description of a last page of a codex where, for example, a scholar has given a false date and a false attribution and some other scholars, deceived, apposed their final approval, like in the following excerpt:

*At f.130v the text ends with "Januarii X M543": the last three numbers have been added over the original ones. Three notes follow, written in the XVIth century by three different hands "explicit opus lucubratum et compositum per me Franciscum Arethinum Theanensem", "Ego dominus Ioannes Antonius de Gallutio laudo", "Hoc opus bene recteque esse compositum et quiddid in eo continetur laudabile esse confirmo. Ego umanitatis magister Lucas Capuanus manu propria".*

The first attempts to produce electronic catalogues were made using database technologies. These attempts questioned the efficacy of database technology for cataloguing manuscripts, whose descriptions are typically very complex, diverse and highly unstructured, as we see in the above example.

Moreover, the systems implemented are so many and diverse that, while moving in the direction of manuscript wide accessibility, they open a new problem: the incompatibility between different systems that makes impossible to move data across different platforms. Finally, some systems may be run only on some hardware or require some particular software platform.

### **XML: A STANDARD FOR MANUSCRIPT DESCRIPTION**

An alternative to database technology is the application of XML (eXtensible Markup Language), which is recently emerging as a standard also for manuscript description [1] [2]. The advantage of adopting the XML language consists in providing the possibility to tag technical and precisely defined information and to structure, in a less rigid scheme, also a prose description, frequently encountered in printed catalogues. In traditional database

company AISoftw@re based in Milan.

the verbose part of a handwritten catalogue either is lost or is inserted in generic *notes* fields, where the information lay unstructured and therefore un-searchable. Turning to the previous record, we exemplify how XML tags allow structuring precisely the information in the first line of the record:

```
<material>Membr.</material>, <extent>ff. IV
+ 121 + IV'</extent>; <dimensions>mm.183x115
(124x80)</dimensions>, <layout>26 long lines
</layout>
```

and how others tags, like the tag `<name>` can help to codify the story about the false attribution, as well, inside a generic prose tag `<p>`:

```
<p>
At f.130v the text ends with "Januarii X
M543": the last three numbers have been added
over the original ones. Three notes follow,
written in the XVI century by three different
hands "explicit opus lucubratum et compositum
per me <name>Franciscum Arethinum Theanense
</name>", "Ego dominus <name>Ioannes Antonius
de Gallutio</name> laudo", "Hoc opus bene
recteque esse compositum et quidquid in eo
continetur laudabile esse confirmo. Ego
umanitatis magister <name>Lucas Capuanus
</name> manu propria".
</p>
```

The definition of a standard allows making different archives compatible and therefore to exchange information, as well. Several institutions in USA and in Europe are strongly committed to elaborate a standard for electronic description of medieval manuscripts to allow communication between different catalogues. The main institutions are:

- Electronic Access to Medieval Manuscripts (EAMMS). The Andrew W. Mellon Foundation funds this project at the Hill Monastic Manuscript Library in Collegeville, Minnesota. It ends to determine the standards for on-line cataloguing of medieval and renaissance manuscripts and develops the tools for the two implementations of MARC and SGML/XML format.  
<http://www.hmml.org/eamms/index.html>
- Text Encoding Initiative (TEI). An international and interdisciplinary project, established in 1987, to develop guidelines for the encoding of textual material in electronic form for research purposes. TEI became a consortium in December 2000. Several projects use the TEI encoding scheme, also in archival and museum information field, but above all in publishing and managing electronic historical and literary texts. The University of Oxford, the Brown University, the University of Bergen and the University of Virginia host the TEI consortium. A sub-group is devoted to manuscript cataloguing.  
<http://www.tei-c.org/>
- Manuscript Access through Standards for Electronic Records (MASTER). A EU funded project, attempting to define and implement a general-purpose XML based standard for the

description of manuscript material. Master is collaborating closely with TEI.  
<http://www.cta.dmu.ac.uk/projects/master>

EAMMS has two working groups: the first one focusing on a standard definition in MARC (Machine-Readable Cataloging) [9], the second one studying a SGML/XML DTD (Document Type Definition) for manuscripts description. As MARC is a standard created for the exchange of bibliographic records, the EAMMS group is trying to find proper solutions also for manuscripts' information.

The TEI Description Group is elaborating a DTD for manuscript description, while MASTER is using its own, similar but not identical to the American one. The two organizations are now working to come to a definition of a common DTD; the result of the collaboration is a revised MASTER DTD, compatible with TEI DTD [4].

### Filter Manus/XML in Ambrosiana

We implemented a filter between the so-called *Manus* system and the MASTER DTD [2] [10]. *Manus* is a relational database, based on Microsoft Access, defined by the Italian Istituto Centrale per il Catalogo Unico (ICCU) and implemented by Finsiel/Telecom. Its design follows the traditional rules of manuscripts cataloguing, but being a database, as oppose to XML, it bridle cataloguers in a quite rigid scheme. For this reason we tackled the problem of providing compatibility between ICCU database and XML description of data. The implementation was carried over by student Stefania Smaldone as part of her thesis [10].

We analysed the correspondence between *Manus*' fields and the tags defined by the MASTER DTD and implemented an application to import into *Manus* archive descriptions originally created in XML. This is done transferring the XML files in the Ambrosiana's object-oriented database, in the form of a syntactic tree: the tree root represents the document, the nodes are the XML tags and the leaves are the tags' content. The tree structure allows to search tags and to manage their contents, in order to transfer them into *Manus*' fields. We can therefore produce manuscripts description in XML, conserve them inside the Ambrosiana archive, and still be compatible with ICCU project as well.

Another procedure allows transferring the manuscript descriptions created in *Manus* into XML format, once more using the object-oriented database, as interface between the two different data description.

These procedures, written in Java, guaranty the compatibility of ICCU application, started as a national tool, with any other in the world adopting the XML standard. This approach represents a good solution for all those manuscripts cataloguing projects, which do not use *Manus* for data input, but still want to be compatible with this archive<sup>2</sup>.

---

<sup>2</sup> The filter is in its prototypal phase. ICCU and Finsiel/Telecom are planning to extend Manus with the filter. Little steps like this one are most useful in the direction of standardization.

## CATALOGUES ON-LINE

There are few catalogues on-line, but they are sufficient to demonstrate how much they can help scholars. One of the best examples of catalogue on-line is *Digital Scriptorium*, a project started in 1996 with a grant of the Andrew W. Mellon Foundation to the Bancroft Library of the University of California at Berkeley, and the Rare Book and Manuscript Library of Columbia University, in New York. Digital Scriptorium has been created with an Access database compatible with the TEI DTD. As Consuelo W. Dutschke, Curator of Medieval and Renaissance Manuscripts at Columbia, says: *Digital Scriptorium is a web-based visual union catalogue of medieval and renaissance manuscripts: web-based, in that it was intended from the start as an expandable and correctable tool, in the way that the static media of print or even CD-Rom cannot be; union, because it meshes into one database the holdings of a number of institutions; and visual, because it includes at least one image of every manuscript described, thus allowing the user to test and eventually correct the catalogued information by means of his own specialized knowledge.* At present Digital Scriptorium contains about 2200 entries, and 8000 images [3].

The aim of disseminating manuscripts' information through the Web is also the concern of the British Library's Manuscripts Catalogue, a big archive, started in 1994, by retrospective conversion of previously printed catalogues (the catalogue is, therefore, not a first-hand product). Rachel Stockdale, head of manuscripts cataloguing, explains the rationale: "... [users] who now expect information to be available in their own homes and universities literally at the click of a button". To realize this project they scanned and parsed the data of about 60 volumes, and used them to populate a database, now accessible to the public [11].

References to other significant projects can be found in the EAMMS web page with links to major international web sites devoted to manuscripts and catalogues (<http://www.hmm1.org/eamms/links.html>). There are also some Italian projects [12] of electronic manuscript catalogues and we refer to some of them:

- *Progetto Galilei*, started by the Biblioteca Nazionale Centrale, Florence, in collaboration with the Max-Planck-Institut für Wissenschaftsgeschichte, Berlin. The archive contains one of the most important documentation about history of science in XVII and XVIII centuries. <http://www.bncf.firenze.sbn.it/>
- *Dante on-line*, started by Società Dantesca Italiana. It offers manuscripts' descriptions and images of codices of works by Dante. <http://www.danteonline.it/>
- *Donne e cultura scritta nel Medioevo*. It begun at the end of 2000 by the Universities of Cassino and "La Sapienza" of Rome, in order to gather documents written by women till the end of XV century. <http://edu.let.unicas.it/womediev/>;
- the *Microfilms' Catalogue* of the Istituto di Storia del Diritto Italiano of the Università degli Studi of

Milano is a catalogue of juridical manuscripts' reproductions owned by the institute <http://www.historia.unimi.it/istituto/microstat.ht>;

- the Biblioteca Universitaria Alessandrina, Rome has published an HTML version of the descriptions of manuscripts 236-450, with an historical introduction and some indexes <http://www.alessandrina.librari.beniculturali.it/>;
- the Biblioteca Pannizzi, Reggio Emilia has created an archive with some description of its medieval manuscripts. <http://panizzi.comune.re.it>.

All these experience pose the problem of distributed databases: it is obviously a strong point to be able to access an archive through the network, but even more would be to be able to send your same query to different archive and to be able to search different archives in a uniform way. To work in this direction standards are to be accepted by the community and an effort to open database and to provide filters has to be placed. The standard XML to describe data provides a solution to this and appears very promising but a consensus in the community is still needed.

## MAINTENANCE IN TIME

Dissemination is the main advantage of IT applied to manuscripts. Standardization is a chief pre-requisite, that allows fulfilling this aim. Another condition is considering the persistence in time of electronic supports and the management and maintenance of systems.

Libraries should calculate the costs of replacing software and hardware in time, following the technological evolution, which at times is really so fast to force a breath taken technological run. Any digital project must be kept alive and it is not done once for ever. The software has to be maintained, the platforms have to be updated, the hardware on which they run becomes obsolete and even the physical support of digital data may become obsolete. This experience differs from traditional experience of conservative libraries, where the supports of texts, like parchment and ancient paper, are still alive after hundreds and hundreds of years and we don't need to keep on changing them. Manuscripts can be read after centuries and sometimes they look stronger than our freshly printed bestsellers. What happened to the first electronic archives implemented in the recent years of the XX century, according to time scale of manuscripts, but in the prehistory according to the electronic evolution?

### An archive rescue in Ambrosiana

Another result of the collaboration between DSI and Ambrosiana allowed saving the content of an old tape, the last witness of an early and far-sighted database realized in the 70's, under the direction of Angelo Paredi, Director of the library at the time. He had imported in Stairs the content of an old handwritten catalogue on 10.000 documents on parchments, dated from the 10th to the 19th century, and composed by Alessandro Bianchi [5]. The electronic archive for the time was a very innovative one and has been very much consulted. Some organizational changes and the end of funding left the archive totally un-

supported, soon became unavailable and was forgotten<sup>3</sup>. Once the tape was found again, there was no machine to read it anymore and no one knew exactly what it was all about: only 20 years later. It required months of research, just to be able to find a machine still working to read the tape.

Finally, the data were exported in text format and patiently analyzed to find both their syntactical and layout features, which marked different kinds of information. For example: each description of a single document began with a number corresponding to a shelf-mark, while the second line represented the data of the document, and the third one the place where it was subscribed. The data were thereafter logically subdivided and were used to populate a new object-oriented database, compatible with the Ambrosiana archive [7].

### CONCLUSION

Since the advent of the World Wide Web and the effort of initiative like TEI, important advances in digital catalogues of manuscripts have been made. The promise of a universal accessibility of manuscripts may still be a goal but it is in broadening in ways never envisaged before. The vision by Poggio of freedom of manuscripts by a too long secret custody can now count on technological development, like XML, which appears to be quite appropriate. There still is a gap between what is technically feasible and what is practically realized, due to a number of constraints that vary from economical limitations to librarian awareness. We suffer from a loss of communication between specialists of contents, and that's why we have undertaken dissemination initiatives. Recently at DSI, in collaboration with the MASTER project, the Università Cattolica di Milano and the Biblioteca Ambrosiana, we organized a workshop for manuscripts curator to introduce them to the XML DTD, developed in the MASTER project, and to experience the possibility of uploading manuscript description to a server on the web for accessibility. For most of the librarians this was a very new and intriguing experience.

Even though a universal catalogue may never be feasible, what is strongly desirable is that several national initiatives can meet other experiences and agree on a standard.

### ACKNOWLEDGMENTS

The authors would like to acknowledge the initial impetus given by Gianfranco Prini of Università degli Studi di Milano to the collaboration between DSI and Biblioteca Ambrosiana; Mons. Pier Francesco Fumagalli, member of the Collegio dei Dottori of the library, for supporting the project and the MASTER initiative, and Peter Robinson, Director of the Centre for Technology and the Arts, De Montfort University UK, and leader of the MASTER project.

### REFERENCES

1. Barbero G., Manoscritti, basi di dati e le riflessioni di

Paul Canart, *Biblioteche Oggi*, 18/10, dicembre 2000, 32-35

2. Barbero G., Smaldone S., Il linguaggio SGML/XML e la descrizione di manoscritti, *Bollettino AIB*, 40/2, giugno 2000), 159-179
3. Barbero G., Digital Scriptorium, *Biblioteche Oggi*, 19/1, gennaio-febbraio 2001, 69-71
4. Burnard L. (ed), *Reference Manual for the MASTER Document Type Definition*, January 2001. <http://www.hcu.ox.ac.uk/TEI/Master/Reference/>
5. Bianchi A., *Inventario delle pergamene della Biblioteca Ambrosiana*, 12 ms volumes with shelf-mark K 201-212 suss
6. Fava E., *La Veneranda Biblioteca Ambrosiana: un sistema informatico per la gestione della movimentazione dei documenti e sua interazione con i cataloghi*, Tesi di laurea in Scienze dell'Informazione, Università degli Studi di Milano, AA 1996-1997
7. Ghezzi G., *Recupero di un regesto di pergamene e sua pubblicazione in Internet*, Tesi di laurea in Scienze dell'Informazione, Università degli Studi di Milano, AA 1998/1999
8. Pasini C., Catalogazione e conservazione nel fondo manoscritto dell'Ambrosiana, *Bollettino ABEI*, 8/1, 1999, 46-54
9. Pass G. A. (ed), *Descriptive cataloging of ancient, medieval, renaissance, and early-modern manuscripts*. Draft March 2000. <ftp://165.134.156.3/vatican>
10. Smaldone S., *Manus e Master: due differenti approcci alla descrizione dei manoscritti e conversione tra i rispettivi formati*, Tesi di laurea in Scienze dell'Informazione, Università degli Studi di Milano, AA 1998-1999
11. Stockdale R., Retrospective conversion of the British Library Manuscripts Catalogue: a description of the project, *Journal of the Society of Archivists*, 21, 2, 2000.
12. Zamponi S., Iniziative di catalogazione di manoscritti medievali, *Studi medievali*, s. III, 40, 1999, 369-393

---

<sup>3</sup> Most information about this archive have been found in a leaflet, originally printed for the use of library's students, which survived almost by chance, because found in the pages of the ms catalogue by Bianchi [5]

**Maria Alberta Alberti** is professor at DSI, Università degli Studi di Milano. Her research field is multimedia systems and user interfaces. She is interested to the applicative field of cultural heritage.  
DSI, Università degli Studi di Milano  
Via Comelico 39, 20135 Milano  
alberti@dsi.unimi.it

**Cesare Pasini** is Vice-Prefetto of Biblioteca Ambrosiana, in charge of the Library. His research focuses on Greek paleography and patristic.  
Biblioteca Ambrosiana  
Piazza Pio XI, 2  
20123 Milano  
cpasini@ambrosiana.it

**Giliola Barbero**, Ph.D. in Medieval and Humanistic Philology, since 1995 in Biblioteca Ambrosiana as manuscripts cataloguer. She collaborates with the MASTER Project and with DSI. She is editor of the column *Manoscritti & Computer* for the review *Biblioteche Oggi*.  
gbarbero@hotmail.com