Towards the Project of an Open Catalogue of Manuscripts

Antonio Cartelli
University of Cassino, Italy

cartan@officine.it

Marco Palma
University of Cassino, Italy

mpalma@unicas.it

Abstract

After an introduction and a short description of the research methods usually adopted in scientific disciplines and particularly in palaeography, some hypotheses on the influence of new technologies on human learning and some examples of the use of Web technologies in manuscript cataloguing are proposed. The reasons for the project of an information system adopting Web technologies for manuscripts cataloguing are then reported and the meaning of an open catalogue strictly related to the above information system is presented. The project is applied to a concrete example to show one of the possible ways the information system can be used by the palaeographers' community and, more in general, by scientific community. The paper ends with the proposal of long term publishing hypotheses for materials collected within the database of the open catalogue.

Keywords: manuscript, palaeography, catalogue, Web, database, information system.

Introduction

If we look at the diffusion of the Internet all over the world the birth of DARPAnet in 1969, when only four nodes were connected by an experimental network, seems much more distant in time than it really is. It is also well known the Internet has gained a relevant role among communication media and its use has changed the way knowledge can be acquired, built and shared. Nevertheless it has to be noted that not in all disciplines and not in the same way within a discipline the Internet and its services had the same influence.

In Palaeography many experiences involving the use of the Internet have been carried out till now, but in the authors' opinion a further impulse to the use of the Net in Palaeographic studies can come from the adoption of specifically designed Information systems. The above proposal has its roots in the hypothesis of revision of Palaeography statute as a science, in the links the discipline will settle with other scientific disciplines and in a new role for the users of the information the discipline will produce. Before arriving at the proposals for the introduction of new elements in the discipline statute the authors will shortly describe the main traits of today Palaeography, its history and the ways the Internet changed human knowledge construction.

Is There the Need for a Revision of Palaeographical Epistemology?

It is well known the scientific statute of palaeography, like the ones of all scientific disciplines, is based on peculiar theoretical principles, some of which are: a well defined field of investigation, a specific language, one ore more methods of investigation and knowledge acquisition, a social purpose, a hermeneutic logic. The revision of the foundations of palaeography as a science or the

critical analysis of the scientific basis of the discipline are beyond the purposes of this paper. Main aims of this work are the analysis of the implications of Web technologies introduction on discipline knowledge construction and the study of the consequences the use of new instruments can have on the way new scientific knowledge is created.

The starting point for the remarks the authors propose will be the knowledge methods used by scholars in their studies. It is well known in fact that various methods are used at different extents when researchers are involved in manuscripts' studies: qualitative or quantitative methods and global or analytic methods are perhaps the main ways one can adopt to obtain new knowledge. As an example of qualitative method is reported here the comparative method by means of which manuscripts being differently aged or coming from different places are observed and analysed in their entirety while ignoring specific aspects for each of them. On another side quantitative methods, like the ones coming from physics, chemistry and all the disciplines that mathematically represent the reality, appear to be very important in the analysis of a specific manuscript; moreover statistics proves essential in gathering elements useful for comparing each manuscript with the others. These disciplines make easier the individuation of peculiar features of a single manuscript or groups of them and put the bases for their classification; i.e. the analysis of the ink composition can be used to find elements pertaining to specific ages, the non-destructive analysis of parchment can help to determine the physical-chemical processes it was submitted to, the measure of the parchment thickness in different manuscripts can help classify them and so on.

Together with the knowledge methods used by researchers it is important to remember here also two specific features of their work. The former one is mostly due to the way manuscripts have to be analysed; every researcher has in fact to study personally manuscripts in archives or libraries and compare the ideas emerging from collected data with other scholars' hypotheses in different times. The latter feature is strictly related to the material until now adopted by scholars to share the results of their work with the scientific community; the catalogue a researcher can publish is in fact a printed matter that acquires its definite form and scientific value only when it is published.

In these last years the spreading of networking communication techniques and particularly of the Internet produced relevant effects on the sharing of knowledge and information also within the community of the scholars involved in manuscripts' analysis. First of all the Web was used to publish the images reproducing manuscripts pages, but there were also experiences concerning on line historical magazines, bibliographies or databases (Cartelli, 2001).

It is important for what follows to remember here the results of some recent studies on the influence of the Internet on knowledge and learning.

First of all Rheingold (1994) proposed the concept of *virtual community* to define a set of individuals systematically using networks and the Net for the interpersonal communication and the sharing of knowledge. More recently Veltman (1999) affirmed that the excess in the use of new technologies can lead people to lose the sense of the distinction between reality and illusion and induce a passivity feeling. He also asserted the Internet can give rise to great opportunities for knowledge development because people can systematically collaborate on line and access greater and greater amount of resources.

Other proposals concerning the effects of the Net on human learning came from Lévy (Lévy, 1996), who proposed the terms *collective intelligence* to describe the increase of the interpersonal communication speed and the increase of the amount of information people can access with the Net. He also proposed the terms *connective intelligence* for the process connecting different documents, points of view and knowledge.

At last Calvani (1999) asserted the use of the Net introduced a deep change in the meaning of scientific knowledge. Knowledge is no longer a continuum with the main aim of limiting ignorance (on the basis of the Enlightenment principle of progress), it has now multi-perspective features and not well defined lim-

its. Knowledge isn't only linear, sequential, closed, hierarchic, it is also hypertextual and multimedial. Calvani stated also that the classical models of rationality using the analytic-deductive principle have to be revised and based on the dialogic-hermeneutic model.

In the authors' opinion the results of the above studies represent a chance for revising the investigation methods and the knowledge building processes of every scientific discipline and particularly of Palaeography and manuscripts' analysis. The authors are also persuaded that the process leading to the above epistemological changes is not automatic and lays on the use people can do of the Internet communication instruments. This *good use* of technologies can be helped by the creation of suitable instruments based on the Internet services that have to be specific for each scientific community (in the special case of the field the authors are interested in they have to preserve the individualistic feature of manuscript study while introducing the instruments and opportunities for collaborative working). Particular attention has also to be given to "social effects" of CMC (Computer Mediated Communication); as Paccagnella reports in his work (Paccagnella, 2000), in fact, a naive use of the Internet can have bad effects on virtual communities and can determine their collapse. A good evidence for the damage Internet services can cause to the survival of virtual communities is given by flames and spamming (both phenomena lying on the excess and the distortion in the use of e-mail); other examples of possible trouble elements for those communities can be found in the indiscriminate open Chats, where people with the use of a nickname can mask their identity and assert what they want without any control.

The above considerations led the authors to the proposal of the information system described in the following sections and persuaded them of the importance of finding good ways of interaction between Palaeography and Informing Science. As Cohen (1999) reports "the fields that comprise the discipline of Informing Science provide their clientele with information in a form, format and schedule that maximizes its effectiveness" and "The Informing Science Framework has three components: the informing environment, the delivery system and the task-completion system".

If the authors' proposals will create the premises for a change in the scientific statute of palaeography is too early to say because it will be only the agreement the scientific community will manifest towards these proposals to guarantee their acceptance and reception among the methods of the discipline.

Catalogues of Manuscripts and the Proposal of an Open Catalogue

The history of manuscript cataloguing began between 17th and 18th century, when some scholars (Peter Lambeck in Vienna, Bernard de Montfaucon in Paris) prepared the first printed catalogues of manuscript collections. Their example was followed in the last two centuries by hundreds of cataloguers, who described thousands of manuscripts written in ancient and modern languages stored in many big and small libraries throughout Europe and the rest of the Western world. Yet nowadays an enormous amount of manuscripts remains still uncatalogued, if not unknown. The main trouble with old, handwritten books is the enormous difficulty of getting in touch with them, as you simply don't know that these manuscripts exist in a given place. The case of manuscripts is therefore really different from the one of texts, where bibliographies and Internet resources make relatively easy to find a printed book, or at least a bibliographical description of its contents (several copies of the same book are kept in many libraries: so you can know where to find the book you look for, what is its shelfmark, at what conditions you are allowed to loan it).

A manuscript is a unique book: there are no duplicates. Even if the text is the same, as in the case of a model and its copy, the two books are different in script, physical appearance and even text, because a handwritten copy always contains some changes in comparison with its source. So there is the need for an accurate description of the external and internal features of every single book to identify and qualify it beyond its sheer shelfmark, i. e. what for a human being is her or his name.

There is no limit to the thoroughness of a description. As for the text, you can give a complete list of the texts and their authors (with reference to printed editions or other works for unpublished texts); as for the physical appearance the simple material (i. e. parchment or paper) and the number of folios can be held as the lowest level, but dimensions, quire organisation, binding appear indispensable to give an idea of the book as it is. A cataloguer may object that he can't escape a definition of script and illumination, particularly if there is no photo to publish. The history of scientific studies about that single manuscript may be reduced to the list of the main works about it or include every citation of that precious book.

So a catalogue of manuscripts is a really difficult task, which requires a deep competence in a lot of fields, from philology to history, from the history of art to palaeography, from literature to library science. A catalogue of medieval (or modern) manuscripts can be defined as a scientific product by connoisseurs who devote an enormous amount of time and work to describe some tens or hundreds of these peculiar objects which lie hidden in the shelves of libraries, normally accessible only to the happy few scholars who are admitted to the direct study of them.

It is so easily understandable that in the last decades many librarians or researchers preferred to write books about the theory of description than prepare catalogues of manuscripts. And their theories often found their way into the rules that library authorities or research institutions decided, the observance of which was a necessary condition to get the catalogues printed at public expense.

In the authors' opinion time has come for a radical change in catalogue publishing. With the help of the Internet, as shown in the above section, it is possible in fact to give scholars new and more powerful tools to improve their work and create an *open catalogue*, i. e. a catalogue in progress, constantly changing with the continuous contribution of new information.

The Information System

The project of the Information System to be used for the *open catalogue* had to take into the right consideration the communication medium it is based on. The Internet and particularly the Web looked very soon the best solution for an on line community within which people could meet other people or share their researches and ideas; the same system looked reliable enough both on the server side and on the client side for the publication of the data concerning manuscripts and their continuous updating. In the first case the reliability laid on the server technical features (its fault tolerance, backup procedures etc.) and on the Net steadiness (nowadays only seldom happen breakdowns that prevent people from using its services for more than a few minutes or some hours), in the second one it is due to the clients connecting features that guarantee everyone a good and easy access to the Internet resources; in both cases the property rights of the information one can put on the Net are today reasonably guaranteed. A Web server interfaced with a database containing all the data concerning manuscripts, the notes and the messages researchers could collect and exchange among themselves looked to the authors enough for building up the Information System that had to make real the *open catalogue*; on the client side in fact every user needed only the hardware and the software one can adopt to connect to the Internet.

While planning the information system implementing the open catalogue the authors decided also to keep apart the following environments: a former one where researchers and scholars had the scientific liability of the materials to be published and a latter one people could use occasionally or systematically to obtain information (and not to produce them).

The above separation affected the way users could access the system and led to three different access levels:

a. system and scientific administrator level; only a person having the scientific responsibility of the catalogue can enable researchers and scholars to enter into the system and gives them the access rights to

manage the materials to be published; at this level it is also possible to manage all the data produced by everyone

b. scholars and researchers level; allowed people can insert, modify, delete and display the materials to be published (manuscripts' descriptions, notes, images etc.), at least when they decide their materials have an ultimate structure and can be publicly accessed

c. user level, where only the answers to specific queries can be obtained (i. e. the descriptions and images of manuscripts, notes etc.).

It is quite obvious the first two types of users need well defined authentication procedures to access the system (i.e. all user of the above first two types need an ID number unique within the system and a password).

People interested in accessing the system for managing information have to send an application form (accessible from the home page) to the system administrator via a Web Form containing some unavoidable fields (i. e. people who don't write out data like first name, last name, job/activity, institution/company, e-mail, phone, desired password etc. will not be allowed to access the system).

The system administrator, on the basis of the data sent by the above users, will assign them a unique ID number so that they can access the system and work within it (every operation they will do in the system will be stored in the database and identified with their ID).

The following list reports the operations allowed people can do:

1. they can input data concerning manuscripts

2. they can input observations on manuscripts already in the database

3. they can modify, delete and display all the above data until they decide these materials haven't an ultimate structure

4. they can enter a Web-Chat with different access levels (until now two levels have been identified: a) a generic one where everyone can read all messages, b) a specific one where only messages concerning
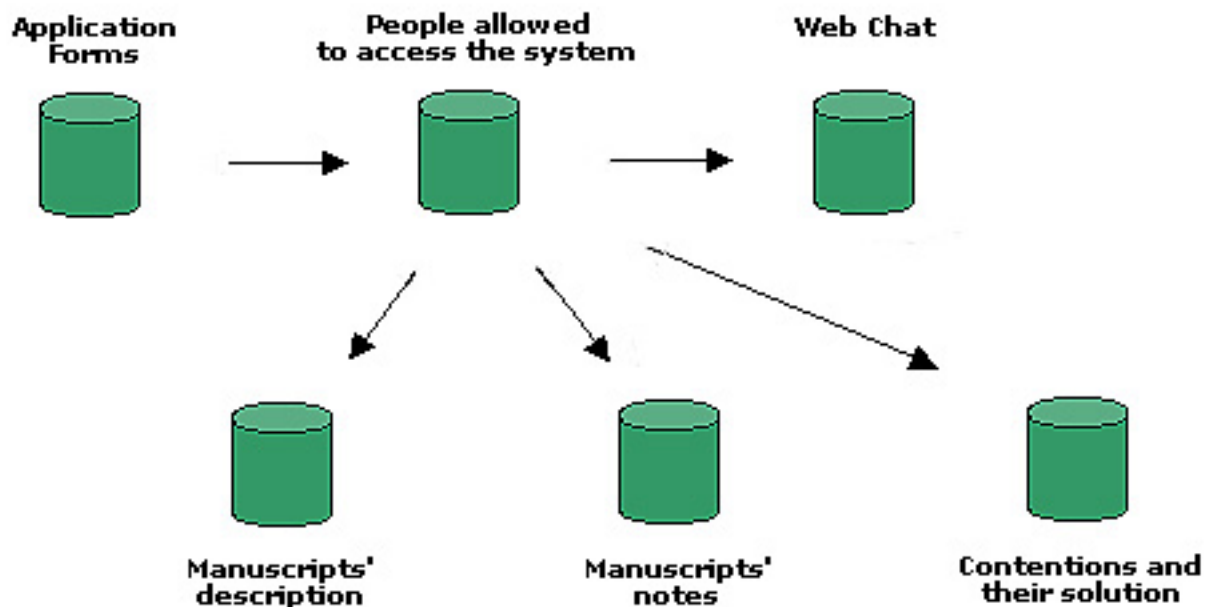

Fig.1

a specific manuscript can be read by everyone)

5. the system administrator can convert a message in the Web-Chat concerning a specific manuscript into a note to the same manuscript (so that it can be read in the public section of the site)

To prevent misunderstandings caused by the text of the notes concerning a specific manuscript a specific discussion area has been hypothesized. This area can be accessed only by allowed users of the above a. and b. categories.

The system administrator can also collect, if needed, all materials under contention (together with everything concerning it within the database) and put them in a specific area only accessible by allowed people. The results of the debate following the publication of those materials will determine their attribution and destination.

Two figures showing the structure and functioning of the system follow: Fig. 1 reports the database structure (i. e. the tables of the RDB it lays on), Fig. 2 shows the data flow within it.

At last a few words have to be said on the hardware and software solutions authors adopted for building the Information System. For the hardware a good PC with a dual Pentium III processor, 128 Mb of Ram and an HD with 10 Gb of data can be used. A similar box will be equipped with the Open Software Linux as operating system together with the Apache server Web, the Postgres-SQL RDBMS and the PHP-3/4 language.

Particular solutions based on XML open packages like Cocoon (Apache project) or JSP Tomcat server (Jakarta project within Apache project) can be adopted for the presentation of manuscripts and will be analysed when and if the project becomes real.
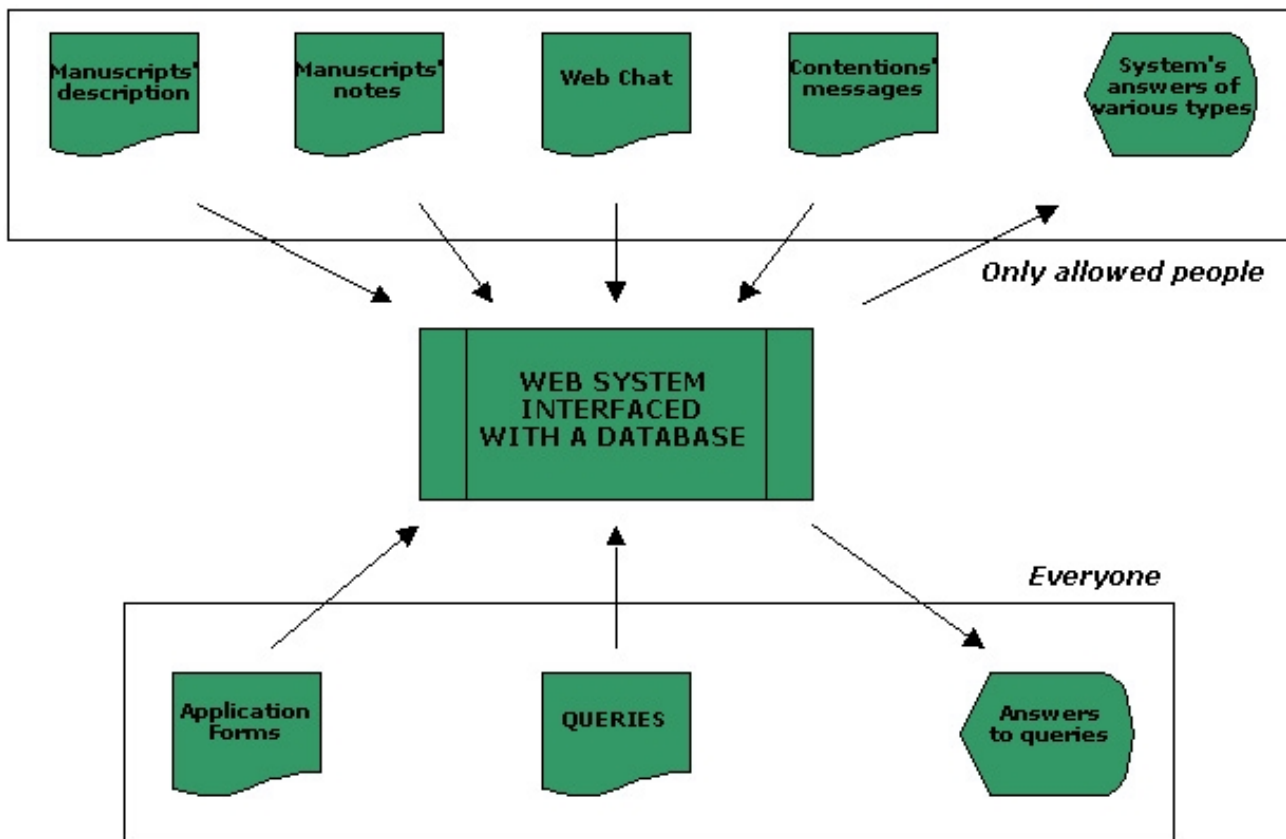

Fig. 2

A Concrete Example: The Manuscripts of the Biblioteca Malatestiana (Cesena, Italy)

In a recent paper by the Committee of the Deutsche Forschungsgemeinschaft (CDF, 2001) (the publicly funded German foundation for research) responsible for the cataloguing of manuscripts in that country a clear statement is made about the uselessness of the publication of 'stationary' catalogues (i. e. in the form of printed books or CD-ROMs). After sponsoring more than 200 printed catalogues throughout Germany in the last 40 years they are now convinced that the Web is the central medium to present and disseminate the results of research. Information about manuscripts should be free and open to collaboration and integration. We are convinced that such a radical change of mind deserves a close scrutiny and an attempt to put it into practice even in countries where such organising criteria appear to be extraneous or indigestible.

A good example in Italian country may be the historical library of Cesena, conceived in 1452 by Malatesta Novello, the town lord, for public utility and still run by the city council. It preserves more than 300 extremely interesting medieval manuscripts, an important group of which were ordered by the founder himself. They have been described in several catalogues and particular studies, but never in an organic and coherent form (Manfron, 1998). The usual process in such a case would require a general catalogue of the whole collection or at least the description of the manuscripts that still lack it. Our proposal is a completely different one, based on the publication on the Web of all useful and available materials, apart from the fact that they have already been or are going to be published in print.

This sort of open catalogue should be composed of five sections. In the first one users should find every general information, already published at the beginning and eventually fresh later, about the collection owned by the Library. In the second section should appear the individual bibliography of each manuscript, continuously updated. The third one should contain the description of the manuscripts, taken from previous catalogues or, if lacking, gradually made. We think that different criteria of description are a small price to pay if you are offered on line information about some hundreds of manuscripts. The fourth section should be devoted to reproductions, the number and quality of which depend on the available means. The last section plays a leading role in the project, as it is planned to host papers, opinions, queries by the public. Everybody interested in the Library's manuscripts would find here a space for intervention, debate and even on line publication of articles.

It has to be noted that the above sections don't conflict with the access levels reported in section 3. The first four sections, in fact, can be accessed only by the users of the Information System while the last section can be accessed by scholars and researchers.

Such a project needs to be monitored daily by scientific personnel, typically librarians or curators who should take the responsibility to run the site in an active way, i. e. by promoting research on their manuscripts and linking scholars interested in them.

And printing?

Our project does not exclude at all the publication of research results in printed books or articles, as we are convinced that Net and print can live together well. A printed catalogue is for ever, in the strict sense of the word: it can never change for centuries ahead. An open catalogue should always change, it is planned to be never the same. If a moment comes when the library can't afford to keep it alive, a publication in print is worthwhile, even in a few copies, even on demand, as low cost printing technology allows nowadays and predictably more so in the future.

References

Calvani A., (1999). *Comunicazione e apprendimento in Internet*, Erickson, Trento (Italy).

Cartelli A., Miglio L., Palma M., (2001). New Technologies and New Paradigms in Historical Research, in *Informing Science*, Vol 4, 2, pp. 61-66, available on the Web at http://www.inform.nu/Articles/Vol4/v4n2p061-066.pdf

CDF - Committee of the Deutsche Forschungsgemeinschaft (2001). *Neue Konzepte der Handschriftenerschliessung. Informationsysteme zur Erforschung des Mittelalters und der Frühen Neuzeit*, available on the Web at http://www.dfg.de/foerder/biblio/download/handschriften.pdf.

Cohen E., (1999). Reconceptualizing Information Systems as a Field of The Transdiscipline Informing Science: From Ugly Ducking to Swan, *Journal of Computing and Information Technology*. 7 (3), 213-219

Lévy P., (1996). *L'intelligenza collettiva. Per un'antropologia del Cyberspazio*, Feltrinelli, Milan (Italy).

Manfron A., (1998). *La biblioteca di un medico del Quattrocento. I codici di Giovanni di Marco da Rimini nella Biblioteca Malatestiana*, Allemandi, Torino (Italy).

Overgaauw E. A., (2001). Neue Konzepte der Hanschriftenerschliessung in Deutschland, in *Gazette du livre médiéval*, 38, pp. 49-54.

Paccagnella L., (2000). *La comunicazione al computer*, il Mulino, Bologna (Italy).

Rheingold H., (1994). *Comunità virtuali. Parlare, incontrarsi, vivere nel ciberspazio*, Springer & Kupfer, Milan (Italy).

Veltman K., (1999). New Media and Transformations in Knowledge, in *Heinz Nixdorf Museums Forum*, Paderborn, available on the Web at http://www.sumscorp.com/articles/art52.htm

Biographies

Antonio Cartelli is a teacher of CS (i.e. European I.T.) in an Italian Technical High School and has a yearly contract for teaching Educational Technologies at the Faculty of Humanities in the University of Cassino - Italy. A. Cartelli is also administering the Network Services in the same Faculty and has authored more than 15 papers on the subjects of Web technologies and conceptualisation, misconceptions and mental schemes in different disciplines.

Marco Palma is a faculty member of the University of Cassino - Italy, where he teaches Latin Palaeography at the Faculty of Humanities. Currently he works on the material aspects of manuscripts as well the writing of books and charters in the Middle Ages.