A Case Study for the Development of Methods to Improve User Engagement with Digital Cultural Heritage Collections

Maristella Agosti, Giordana Mariani Canova,
Nicola Orio, and Chiara Ponchia

University of Padua, Italy
{maristella.agosti,giordana.mariani.canova,nicola.orio}@unipd.it,
chiara.ponchia.1@studenti.unipd.it

Abstract. The aim of this paper is to report the results of an ongoing project that deals with the exploitation of a digital archive of drawings and illustrations of historic documents for research and educational purposes. A prototype system, called IPSA (Imaginum Patavinae Scientiae Archivum), has been developed and is currently used as a case study to provide innovative tools for researchers and scholars active in the preservation and dissemination of cultural heritage. After describing the initial user requirements that motivated the development of IPSA, we focus on the research questions that can be addressed by new system functions and on its extension to additional user groups, including students, experts in other domains, and the general public.

1 Introduction

The ideas and concepts reported in this paper build upon our experience on the analysis of the user requirements, the design of a methodology, the development of a prototype system, and the analysis of the feedback from real users of a digital archive of historical material. The archive aims at the study and research on *illuminated manuscripts*, i.e. usually handwritten books which include illustrations and which in past centuries were manually and artistically decorated. Illuminated manuscripts are the subject of scientific research in different areas, namely the history of art and the history of science, and all disciplines related to the subject of the book – e.g. botany, astronomy, medicine [1]. Before the invention of photography, illuminated manuscripts played a central role in the dissemination of scientific culture, and to this end they bear witness to the heritage of different cultures, in Europe, Asia, and in the countries under the influence of Arab culture.

The digital archive of illuminated manuscripts that has been developed within our research activities is called *IPSA*, which stands for *Imaginum Patavinae Scientiae Archivum* (archive of images of the Paduan science) [2,3]. This is because the main focus of our initial project was to provide a tool for the analysis of the role played by the Paduan school during the Middle Ages and the Renaissance

in the spread of the new scientific method in difference sciences, from medicine to astronomy to botany.

IPSA can also be considered as a case study for our research on methodologies and tools for researchers and scholars working on the analysis, preservation and dissemination of cultural heritage. After a number of years of usage by scholars and students, we started a new phase of recollection of user requirements, with a focus on the *research questions* that can be addressed by an improved systems. To this end, our aim is to provide innovative services to improve user engagement with digital cultural heritage collections.

The final goal of this new phase of our work is to exploit the experience gained over the years both in the design and development of systems that manage digital cultural heritage metadata [4,5] and collections [6,7], and in the usage of multimedia digital archives in order to address new requirements that become evident only after the system has been extensively used as a research tool.

The paper is divided in two parts. The first part describes the initial requirements and development of the actual working system that meets them. The second part introduces our ongoing research on additional user requirements, in the form of research questions related to the disciplines involved in the study of illuminated manuscripts.

2   IPSA Motivations and Objectives

The development of models and tools for researchers and scholars in the area of illuminated manuscripts requires a careful analysis of user requirements [8]. As it turned out from the analyses, the requirements for carrying out scientific research will be more complex and articulated than requirements for final users. Final users access an image digital archive to acquire information in a given field, researchers access the archive to disclose knowledge and discover new relationships between digital objects.

Instead of limiting the analysis to a number of interviews, our approach was to create a research team, where computer scientists and scholars in history of art collaborate. Additional contributions from scholars in related disciplines, such as history of science, botany, astronomy, have been integrated as well and formalized in a draft proposal that has been presented and discussed with research users. A similar approach has been maintained during the development of the prototype system, because all the novel functionalities have been directly tested by members of the research team.

Main results of this initial study are summarized in the following sections. The interested reader can gain further and general information by accessing the Web site that has been developed to document both the projects which have made possible the design and development of IPSA and the managed digital cultural heritage collection[1].

---

[1] URL: http://www.ipsa-project.org/

2.1   Disclosure of Relations between Images and Manuscripts

Scholars in history of miniature are mostly interested in analyzing images, their style, their elements and possible relations with other images belonging to different manuscripts. In particular, it is of primary importance for researchers to discover whether illustrations have been copied from images of other manuscripts, merely inspired by previous works, or directly inspired by nature. A major requirement thus regards the possibility of enriching the digital archive by highlighting explicit relations that have been discovered by a researcher. The analysis of user requirements highlighted a number of issues that are of particular relevance.

- *Authorship*: The definition of a relation between two or more images depends on the scientific results of a researcher, who owns the intellectual rights of this additional knowledge.
- *Typology*: Since two images or two manuscripts can be related for a number of different reasons, the kind of relations should be explicitly expressed.
- *Paths*: Relations may form *historical paths* among images, because images in a manuscript can be copies of another one which in turn are copies themselves of previous illustrations.

These requirements suggested the use of annotations that allow the scholars to connect two manuscripts or two images. These annotations, which have been called *linking annotations*, have a type which describes the kind of relations between the two objects and provides a semantic to the link. We proposed a taxonomy for linking annotations [2] which is divided in two classes, including annotations that express either hierarchical or relatedness links. Annotations have been developed and integrated within the digital archive according to the formal model described in [9].

2.2   Personalization and Collaboration

Almost every digital archive dynamically changes over the years, because of new acquisitions that increase the number of documents and because of changes or redefinition of the descriptive records. In particular, the study of the digital archive content produces new knowledge that, apart from being disseminated through scientific publications, can be represented within the archive itself.

This novel information, which is due to original results, should be stored in the digital archive at a different level than the information based on a general consensus. To this end, both classical textual annotations and the proposed linking annotations can be a viable tool providing that a user is able to state which annotations can be shared with the community or his research group, and which ones have to remain private. Such a mechanism allows scholars to use the digital archive as an advanced research tool, which reflects their personal view of the collection of manuscripts, as well as protect their intellectual rights.

Scholars' annotations on the archive, besides being a means of personalization, can be exploited to foster collaboration. It has to be noted that illuminated manuscripts are of interest to both the historian of art and the historian of science, but at the same time, a herbal is of interest to the botanist because they represent plants and their possible variations through the centuries, a codex is useful for researchers on the evolution of civil and criminal laws, an astrological book may give insights to researchers in medicine on the way stars where perceived to influence the health of people and to astronomers on how constellations where seen and represented. Hence, the scientific research on illuminated manuscripts involves a number of persons with different expertise, which should be able to cooperate in order to share their different knowledge and background.

A digital archive of illuminated manuscripts has to provide a collaborative environment, such as in [10], where researchers should be able to interact and give different contributions to the definitions and redefinitions of objects in the manuscript. For this reason, different levels of users of the archive need to be considered. Apart from the administrators, the group of research users should be able to modify the records of the underlying database when new features of the stored objects are discovered.

3   Development of IPSA

A prototype implementing the proposed methodology has been developed. The close collaboration within a single team of researchers and scholars of all the disciplines involved allowed us to create a closed loop for evaluation, testing and refinement of the different functionalities. Once the underlying database structure had been designed and developed, the organization of the user interface and the development of the novel functionalities highlighted by the user requirements were done incrementally, with scholars in history of art starting to populate the archive and studying the collection of images during the refinement of the software tools.

Figure 1 shows the search page of the IPSA prototype, which is text-based using available metadata, because content-based search was not part of the user requirements. This means that metadata, especially in the form of annotations, should provide information about the visual similarity of the images as well, which – according to the interviewed scholars – can be stated only by an expert.

Figure 2 shows the selection of an image from the results. As usual, information and metadata about the image are reported. Moreover, the image can be explored and navigated. Only a small part of the complete image can be visualized at high resolution, and this is done through a proprietary Java application. Finally, the user is also presented with the annotations of the link the retrieved image to other images.

The prototype of IPSA has been used as a research tool by scholars in the history of art in our team. An effort of dissemination has been made to present it to other researchers in Italy and Europe, through a number workshops and presentations in the research area of illuminated manuscripts.
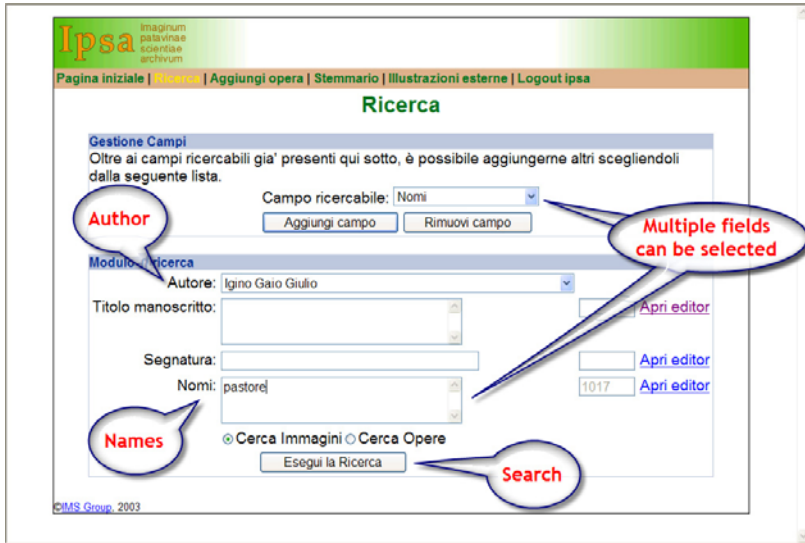
Fig. 1. Search functionalities in IPSA

4 Improving User Engagement

Building from the experience in using the actual version of the prototype, the next step in our ongoing project was to study how to extend its functions to develop it as an education and dissemination tool. At the same time, we wanted to elaborate on actual functionalities to address a number of *research questions*, which can be addressed by automatic tasks to help scholars to discover new knowledge. In this study process, IPSA can be considered as a sort of *case study* to be used to learn new ways of using and extracting information of interest from new categories of users. A further step will be to generalize the findings of this case study to similar digital cultural heritage collections and applications.

5 Research Questions

Using IPSA as a new starting point to develop tools for researchers in illuminated manuscripts, we begun a new analysis of requirement on the *research questions* that should be addressed by a digital archive. The analysis has been carried out on a focused group of scholars and professional users, including professors in the history of illumination, in the history of medieval art, and experts in digitized manuscripts.

The initial results of this ongoing study highlighted some priorities. The research questions described in Section 5.1 confirm the results of our initial analysis of requirements, introducing additional concepts to refine the existing tools.
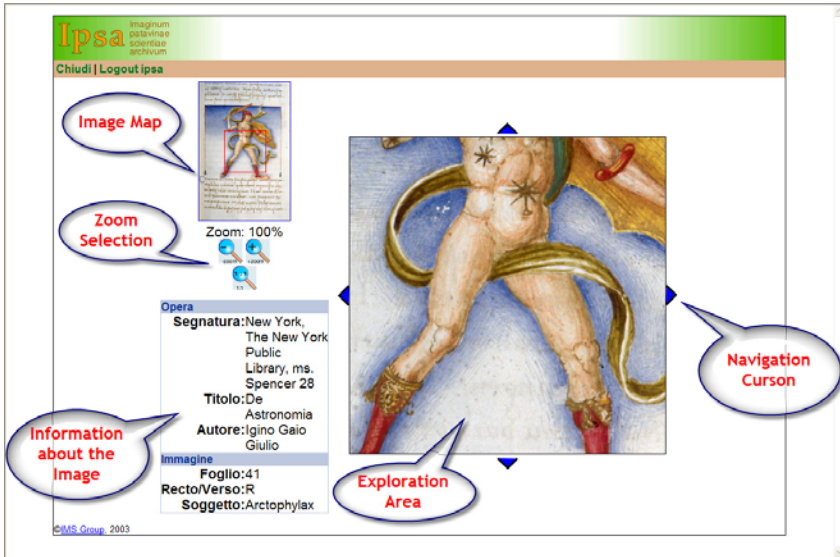
Fig. 2. User interface of IPSA for the analysis of images

The research questions that are described in the subsequent sections are novel, and we believe that the possibility of using IPSA as a research prototype helped the user group to highlight them more clearly.

5.1   Relations between Images

The user group underlined that images are the main subject of scientific research on illuminated manuscripts. Text surrounding the image is important as well, as described in Section 5.3, yet it has to be noted that in many cases the author of a manuscript copied the text of pre-existing manuscripts, while the illustrator added original drawings. These drawings can be copied, with some modifications, from previous images, or just be inspired by them.

Thus an image, besides being relevant because of its intrinsic artistic value, becomes of particular interest because of its relations with other images. The first research question regards the possibility of *following the development of illustrations of a specific text*, that is to track the evolution of iconography of the subjects described in a text, stating where changes have been applied by illustrators and which were their references while drawing images for a text.

At the same time, it is of interest the study on the representation of a specific subject, leading to a second research question that regards the possibility of *following the evolution of the images describing a specific subject*. Also in this case, it is important to state which were the references for illustrators and whether they copied or be inspired by previous drawings, which may be surrounded by different text.

The actual version of IPSA already supports an annotation mechanism that partially addresses these questions. In particular, linking annotations can be used to represent relations between images, while their type describes the kind of relations. Yet, a third research question regards the possibility of *expressing in detail the research results that motivate a relation*. This additional requirement regards the possibility of including textual annotations that describe the considerations for expressing a relation between images.

It is interesting to note that, although most of the research is carried out by analyzing images, also in this case the user requirements did not highlight the need of tools for the automatic computation of visual similarity. It has to be noted that scholars normally have a complete knowledge of the domain they are studying, and they are not used to count on automatic tools to discover relations between images. It is likely that non expert users could take advantage from automatic tools, although the extension to other user categories will be part of future work and it is not covered by this contribution.

5.2   Exploitation of External Resources

A second group of research questions regarded the relations between the content of the archive and external collections. Illustrators could be inspired by manuscripts that are part of other collections but also by other art forms of the same historical period. For example, a drawing can be derived from a painting, a fresco or from illuminated manuscripts with a different subject (for instance religious manuscripts are not included in the collection managed by IPSA). It is important to note that, in a period where traveling was difficult, illuminated manuscripts gave an important contribution to the spread visual representation styles across Europe and the Mediterranean area.

The main research question related to this point can be expressed in two main forms, regarding either the possibility of *finding relations with other digital archives* or the possibility of *querying the archive using external information*. From the analysis of requirements it seems that automatic tools that mine the content of online collections can be a valuable tool for researchers. In particular, scholars find particularly useful the automatic mining of metadata, including authorship, subject, iconography, and geographical area of production. At the same time, the scientific research on illuminated manuscripts can take advantage for any kind of documentation that can be related to the content of the manuscripts. The possibility of having this information available when studying an image is considered of great importance.

5.3   Relevance of Textual Information

As mentioned above, scholars in the history of miniature are mainly interested in images. For this reason, the user requirements for the development of the

actual version of IPSA did not highlight the need of including the text of the manuscripts in the archive. In many cases, the text was directly copied with only few variations mainly due to errors made by the copyist. The analysis of text is of interest for philologists, which were not included in the focused group.

Yet new requirements highlight the importance of the text surrounding the images, especially in relationship with the possibility of using external information to query the archive, as described in Section 5.2. In this case, the main research question regards the possibility of *segmenting the text of the manuscript and linking segments to images.* Information retrieval techniques can be applied to a focused crawling of external resources, integrating the results obtained with metadata. Textual information is considered by scholar a more important evidence of the relations between images than the ones obtained by visual features that can be extracted automatically from images.

The inclusion of textual information poses challenging problems, because text is hand written (and thus very difficult to parse automatically) and usually in Latin. Automatic translations, to English and other languages, should be used to find similar content outside the collection. Moreover, language processing techniques should be tailored to the particular application domain, where terms are not normalized and the writing style is not comparable to the one of modern languages.

5.4   Graphical Representation

A final set of research questions regards the visual organization of the archive content and of the results of a search. Scholars need to compare different images, that have to be presented on the screen at the same time, with high resolution and the possibility of zooming on details. The kind of relations between images have to be represented as well, with simple visual cues that represent it. At the same time, the presence of relations induce a hypertextual structure to the collection of images, which should be represented effectively. An example of the possible relations of a given image is depicted in Figure 3.

The research questions on the graphical representation regard both the possibility of *expressing graphically the kind of relation between two images* and the possibility of *changing the focus of the representation*, highlighting the image of interest inside the graphical representation. A number of possible representations has been designed and are currently taken into account by scholars. The hierarchical relations between images is of paramount importance for the scholars, who are interested in the spread of a given iconography across the century and are used to represent it with a *stemma codicum* (a tree-like representation of the hierarchical relations). To this end, an additional research question regards the possibility of *automatically building a stemma codicum for each image*, from an archetype to the image under analysis.
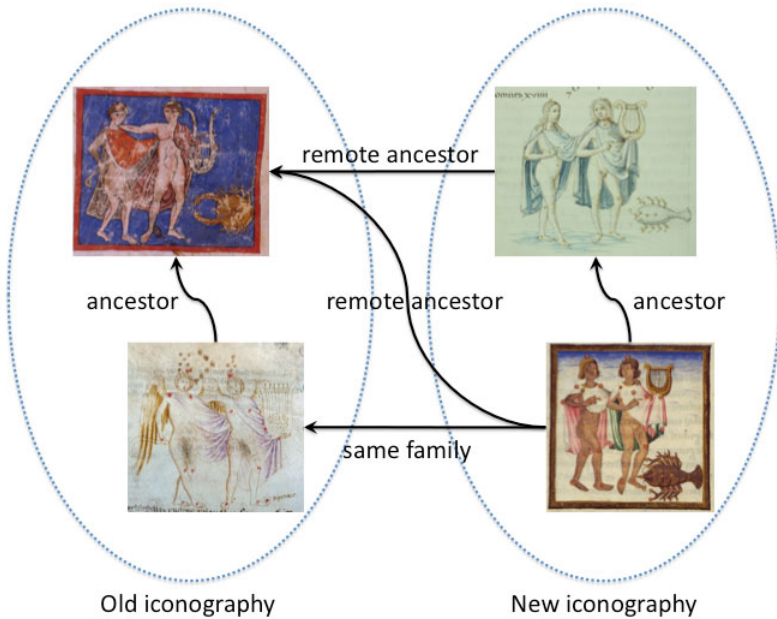
Fig. 3. Example of a personalized view on some linked images of the same subject, belonging to two different iconographies

References

1. Mariani Canova, G.: Hyginus De Astronomia. In: The Splendor of the Word. Medieval and Renaissance Illuminated Manuscripts at the New York Public Library, pp. 337–339. Harvey Miller, New York (2006)
2. Agosti, M., Ferro, N., Orio, N.: Annotating Illuminated Manuscripts: an Effective Tool for Research and Education. In: Marlino, M., Sumner, T., Shipman III, F.M. (eds.) Proc. 5th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2005), pp. 121–130. ACM Press, New York (2005)
3. Agosti, M., Ferro, N., Orio, N.: Graph-based Automatic Suggestion of Relationships among Images of Illuminated Manuscripts. In: Haddad, H. (ed.) SAC, pp. 1063–1067. ACM (2006)
4. Agosti, M., Masotti, M.: Design and functions of DUO: the first Italian academic OPAC. In: Berghel, H., Deaton, E., Hedrick, G., Roach, D., Wainwright, R. (eds.) SAC 1992: Proceedings of the 1992 ACM/SIGAPP Symposium on Applied Computing: Technological Challenges of the 1990's, pp. 308–313. ACM, New York (1992)

---

5. Agosti, M., Ferro, N., Silvello, G.: An Architecture for Sharing Metadata Among Geographically Distributed Archives. In: Thanos, C., Borri, F., Candela, L. (eds.) Digital Libraries: Research and Development. LNCS, vol. 4877, pp. 56–65. Springer, Heidelberg (2007)
6. Agosti, M., Bombi, F., Melucci, M., Mian, G.A.: Towards a digital library for the venetian music of the eighteenth century. In: Proc. of Third International Conference on Digital Resources in the Humanities (DRH 1998), pp. 75–77. The Humanities Advanced Technology and Information Institute, Glasgow (1998)
7. Agosti, M., Benfante, L., Orio, N.: IPSA: A Digital Archive of Herbals to Support Scientific Research. In: Sembok, T.M.T., Zaman, H.B., Chen, H., Urs, S.R., Myaeng, S.-H. (eds.) ICADL 2003. LNCS, vol. 2911, pp. 253–264. Springer, Heidelberg (2003)
8. Crane, G.: Cultural Heritage Digital Libraries: Needs and Components. In: Agosti, M., Thanos, C. (eds.) ECDL 2002. LNCS, vol. 2458, pp. 626–637. Springer, Heidelberg (2002)
9. Agosti, M., Ferro, N.: A Formal Model of Annotations of Digital Content. ACM Transactions on Information Systems (TOIS) 26, 3–57 (2008)
10. Thiel, U., Brocks, H., Frommholz, I., Dirsch-Weigand, A., Keiper, J., Stein, A., Neuhold, E.J.: COLLATE - A collaboratory supporting research on historic European films. International Journal on Digital Libraries 4, 8–12 (2004)