User Requirements for Effective Access to Digital Archives of Manuscripts

Maristella Agosti
Department of Information Engineering, University of Padua, Italy
Email: maristella.agosti@unipd.it

Nicola Orio
Department of Cultural Heritage, University of Padua, Italy
Email: nicola.orio@unipd.it

Abstract— **The availability of digital multimedia resources for cultural heritage poses challenging questions about the kind of interaction that is necessary to foresee to make them available to distinct categories of users. We report on a study on user requirements carried out on two groups of users: domain professional researchers, who need to carry out their research exploiting the functions of digital archives, and non-domain researchers, who need to enhance their experience of digital content and do not have a specific research interest for the digital content. A digital archive of illuminated manuscripts, called IPSA, has been used as a case study for recollecting user requirements. The results highlighted a number of additional requirements by specialized and by non-domain users. The relations between these outcomes provide relevant insights into the role of digital resources for the study and dissemination of cultural heritage.**

Index Terms— **Digital Libraries; Digital Archives; Presentation of Content in Multimedia Digital Archive Systems; Personalized, Human-centered Multimedia**

I. Introduction

The ideas and concepts reported in this paper build upon our experience on the analysis of user requirements [1], 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 digital archive aims at the study and scientific research on *illuminated manuscripts*, i.e. usually handwritten books that 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 manuscripts themselves – e.g. botany, astronomy, medicine [2]. Before the invention of photography, illuminated manuscripts played a central role in the dissemination of scientific culture because they allowed authors to enrich their text with iconographic material that could exemplify the content.

The interest towards historical manuscripts and their role in the dissemination of cultural heritage is witnessed by the number of projects involving their digitization and online availability. The collaboration between different groups of experts is of particular relevance in these kinds of projects, as for instance the DEBORA project [3] that involves end-users, librarians, historians, researchers in book history and computer scientists for the development of remote and collaborative access to digitized Renaissance books. The interest towards multimedia historical collections is witnessed also by the number of projects on the application to the cultural heritage domain of techniques of multimedia analysis. For instance, the Madonne project [4] combines automatic document analysis tools and image processing techniques to provide new tools for accessing manuscripts containing texts and images. Another relevant project focuses on illuminated manuscripts of the Renaissance, where the goal is to provide tools for the interactive exploration and annotation of artistic multimedia content [5].

The analysis of user requirements presented in this paper has been made possible mainly because specialized users had access to a prototype system that implements the main functionalities of searching, browsing and presentation of metadata and images of manuscripts. This allowed researchers to go beyond the basic requirements on the interaction with a digital archive. To this end, a digital archive can be considered as a sort of *case study* for learning new ways of using and extracting information of interest from specific categories of users. A further step is to generalize the findings of this case study to similar digital cultural heritage collections and applications and to contribute to the design of innovative adaptive and personalized services [6]. The need to personalize multimedia access and to tailor it to different categories of users has been described by [7] as a relevant challenge in human-centered multimedia.

The work described has been carried out within the CULTURA project[1], which aims at developing an interactive environment able to offer genuine user empowerment and unprecedented levels of engagement with digital cultural heritage collections and communities.

II. The Starting Point: the IPSA Digital Archive System

The digital archive of illuminated manuscripts developed within our research activities is called *IPSA*,

[1] http://www.cultura-strep.eu/

doi:10.4304/jmm.7.2.217-222

Figure 1. Examples of pages from manuscripts of the IPSA digital archive.

which stands for *Imaginum Patavinae Scientiae Archivum* (archive of images of the Paduan science) [8]. This is because the main focus of the projects that made possible the design and development of IPSA 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 different sciences, from medicine to astronomy to botany [9].

An illuminated manuscript can be a book or document written by hand rather than typed or printed. Each page of a manuscript can be constituted by different types of information, including: pure handwritten text, one or more images without text, or by a combination of one or more images with handwritten text as in Figure 1 where exemplary images of the IPSA digital archive are shown.

The IPSA digital archive is constituted by illuminated images taken from manuscripts, which were produced mainly in Padua and the Veneto region during the XIV and XV centuries. The archive includes 56 manuscripts belonging to some of the most important libraries in Europe and the world, including: the British Library, the Bodleian Library, the Pierpont Morgan Library, the Biblioteca Marciana, the Biblioteca Apostolica Vaticana, and the Biblioteca Medicea Laurenziana. All manuscripts included in IPSA are scientific books, of which 46 are astrological and 10 botanical codexes. The digital archive includes about 3400 images. The aim of the entire collection is to testify the shaping of a new scientific mentality in the University of Padua during the XIV century [10]. The collection has been digitized and marked up with key metadata information through the efforts of several research groups. The metadata include information on the content and the provenance of each digital image.

The current version of the IPSA digital archive already provides a number of functionalities for searching and accessing the manuscripts and their images. Searches can be carried out both as an exact match on user-defined fields and as a free text on relevant textual fields. Manuscripts are presented on-screen with all their images, which in turn can be selected and analyzed through a graphical interface that allows the user to zoom in on details with high resolutions as reported in Figure 2; the example shown also depicts (with a thumbnail on the right side) a similar image that a professional researcher has linked to the original one. This link can be used to directly access the related image using navigation as an alternative to direct search.

The IPSA digital archive system has been used as a starting point for a novel recollection of user requirements. The analyzed classes of users are expected to have special needs that can be the basis for a deeper engagement of interested members of the general public. In this work we present the results of this investigation as the starting point for the development of new types of systems for the dissemination of cultural heritage, taking advantage of the role of mediator played by experts in the domain.

III. Analysis of User Requirements

Two groups of users were considered:

- A focused group of four professional researchers, including professors in the history of illumination, and in the history of Medieval art.
- A group of five non-domain researchers, divided into two classes: experts in related research areas – i.e. painting of the same historical period – and master students in archival science, without a specific interest in the content of the digital archive.

Professional researchers were involved in the initial design of IPSA and were already using the digital archive for their studies. This new analysis aimed at identifying additional requirements that might have emerged while using the archive. For non-domain researchers, the analysis was carried out as a hands-on experience. Users were asked to interact freely with the digital archive, explore its functionalities and access its digital content. For both groups, requirements were gathered through interviews, which were carried out at the end of the interaction with IPSA.

A. Requirements of Professional Researchers

The results of the study highlighted some priorities in the development of new functionalities of IPSA, in order to be used more effectively by professional researchers as a research tool. The outcomes described in Section III-A.1

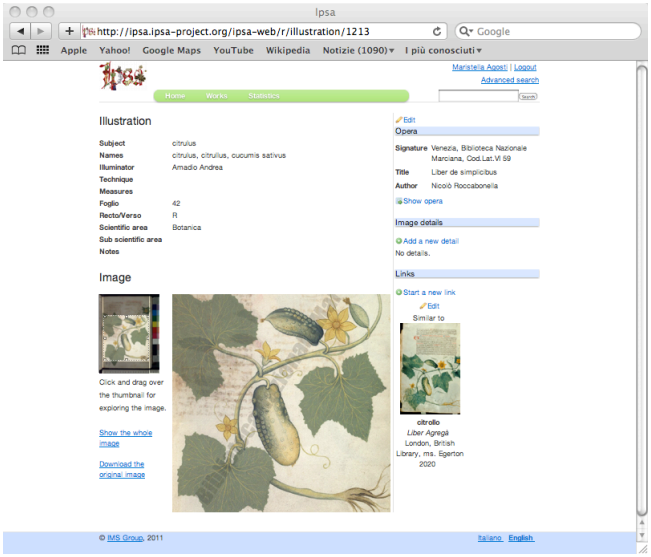

Figure 2. Example of image zooming in IPSA and of a link between the inspected image and a similar one.

confirm the results of our initial analyses, which were carried out at the time of envisaging and designing the IPSA digital archive. The requirements described in the subsequent sections did not emerge in the initial user requirements analysis, where the focus was on designing from scratch a new digital archive able to manage a collection of digital images of illuminated manuscripts. This may be due to the fact that professional researchers had been using the IPSA digital archive system for some time and they were becoming aware of additional needs for their research while using the system. This shows that user requirements can evolve over time while users are acquiring new knowledge on what can be obtained with information technology methods.

1) Relations Between Images: Professional researchers underlined once more that images are the main subject of their scientific research on illuminated manuscripts. Although text surrounding the image is important as well, as described in Section III-A.3, in many cases the authors simply copied the text from pre-existing manuscripts. In contrast, the illustrator usually added original drawings, which can be copied from or inspired by previous images, but which have usually some important modifications. Therefore an image, besides being relevant because of its intrinsic artistic value, becomes of particular interest because of its relations with other images belonging to other manuscripts.

The first requirement regards the possibility of *following the development of illustrations of a specific text*, i.e. to track the evolution of iconography of the subjects described in a text, stating where changes were applied by illustrators and what their references were while drawing a given image for a text.

At the same time, it is of interest to study the representation of a specific subject, thus leading to a second requirement about the possibility of *following the evolution of the images describing a given subject*. Also in this case, it is important to state what the references were for illustrators and whether they copied or were inspired by previous drawings, which may be surrounded by different text. Although most of their research is carried out by analyzing images, professional researchers did not express a need for tools for automatic image processing, for instance to extract visual descriptors aimed at computing image similarity. Scholars in history of art normally have a complete knowledge of the domain they are studying and they are not used to counting on automatic tools for accessing audiovisual content.

2) Exploitation of External Resources: A second group of requirements regarded the relations between the content of the IPSA digital archive and external collections available online. In fact, illustrators could be inspired by manuscripts that are part of other collections but also by other art forms of the same or preceding historical periods. For example, a drawing can be derived from a painting, from a fresco or from illuminated manuscripts with a different subject – e.g. religious manuscripts are not included in the collection managed by IPSA, but they may have been the source of inspiration for the drawing of images in astronomical or astrological manuscripts. In a period where traveling was difficult, illuminated manuscripts gave an important contribution to the spread of visual representation styles across Europe and the Mediterranean area.

The main requirement related to this point can be expressed in two main forms, regarding either the possibility of *finding relations with other digital archives and libraries* or the possibility of *querying the digital archive using external information*. Scientific research on illuminated manuscripts can take advantage of any kind of documentation that can be related to the content of the manuscripts.

3) Textual Information: As mentioned above, scholars in the history of illuminated manuscripts 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 for including the text of the manuscripts in the digital archive. The analysis of texts is surely of interest for philologists, which is an additional group of non-domain researchers that was not part of the study.

New requirements highlight the importance of the text surrounding the images, especially in relation with the possibility of using external information to query the digital archive, as described in Section III-A.2. In this case, the main requirement regards the possibility of *transcribing and segmenting the text of the manuscript and linking segments to images*. Professional researchers underlined that the text can be a valuable aid for identifying similarities between the subjects of illustrations.

Professional researchers consider that useful textual information also includes papers written about the manuscripts, thus *making available relevant literature on the collection*. Moreover, an effective use of the digital archive as a research tool requires the possibility of *adding researchers' comments on manuscripts, images and their relations*.

4) Graphical Representation: A final set of requirements of professional researchers regards the visual organization of the digital archive content and of the results of a search. Scholars need to compare different images, which therefore have to be presented on the screen at the same time, with high resolution and the possibility of zooming in on details. The kind of relations between images has to be represented as well, through the use of simple visual cues.

The requirements 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 interface. The hierarchical relations between images are of paramount importance for the scholars, who are interested in the spread of a given iconography across the centuries and are used to representing it with a *stemma codicum* (a tree-like representation of the hierarchical relations). To this end, an additional requirement regards the possibility of automatically *building a stemma codicum for each image*, from an archetype to the image under analysis.

B. Requirements of Non-domain Researchers

The IPSA digital archive was designed as a tool for researchers in history of art. However, it is important to verify if a system developed for one particular group of specialized users can also be of interest for scholars in different domains and for studying the overlaps between the requirements of users of related areas. This can be considered as a first step towards the opening of the IPSA digital archive to a wider audience, including students and the general public.

1) Effect of Computer Skills: The first outcome of the survey with non-domain researchers is that the competences on the use of automatic tools play an important role in the way users interact with the digital archive. In particular, specialists in related research areas show a *resistance to change* the way they interact with digital resources. The presence of multimedia content available for inspection and analysis was considered an additional difficulty in the interaction, because the non-domain researchers who took part in the analysis of requirements were accustomed to searching for bibliographic values rather than obtaining and studying images directly through the interface of the digital archive system. A requirement that is directly related to this resistance to change is *providing contextual aid to facilitate the interaction*.

In contrast, master students are more accustomed to interacting with multimedia systems, and did not find it problematic to access multimedia information directly from the digital archive. Their approach with the digital archive content was biased by their habit of interacting with large multimedia collections available on the Web. In this case, their requirements indicate a need for *providing search facilities that are typically available in multimedia online collections*. For instance the possibility to obtain recommendations based on users' tags, using a crowdsourcing approach as described in [11], has to be considered.

2) Navigation Tools: An additional set of requirements regarded the exploitation of links within and, especially, outside the digital archive. This finding is in line with the one reported in Section III-A.2 for professional researchers and summarized as *developing a linking mechanism towards other online collections and the Web*.

The possibility of exploiting links to navigate the digital archive content was considered of interest as well. In particular, non-domain users need different ways to access multimedia information. To this end, the possibility of expressing relations between images, as reported in Section III-A.1 for specialized users, may become a valuable tool for casual users to retrieve information of interest. A lack of knowledge on the domain of cultural heritage resources may prevent users finding information, while the presence of annotations made by experts can fill this gap.

3) Textual Information: A final outcome, which might also be extended to other collections in the cultural heritage domain, is that multimedia information – digital images in the case of IPSA – is not sufficient to raise interest when it is not paired with accompanying textual information. Both classes of non-domain researchers suggested directions in which the digital archive should be enriched by relevant textual information to be more attractive. In particular, they both were interested in information that keeps track of the research work carried out by professional researchers.

Research professionals in related domains asked to have *access to analytic descriptions of all the multimedia content of the digital archive*, to make them more aware of the characteristics of the images studied, to highlight relations with other research fields, and basically to follow and understand the research process of specialized users. Master students on archive and library science had a bias, as can be expected, towards bibliographic values. In

particular, they showed an interest towards *accessing the way bibliographic records were created*, their authorship, the motivations behind attributions and so on.

In both cases, the approach to IPSA content depended on the particular field of interest of non-domain users. A common outcome is that even if the content of the digital archive was perceived of important cultural value, the interest towards the digital collection needs to be stimulated by *enlarging the scope of the digital archive*.

IV. Discussion

As expected, the analysis of user requirements highlighted a number of differences between the two user groups, which are strictly related to the level of interest towards the content of the digital collection. On the one hand, professional researchers have clear research goals when interacting with the IPSA digital archive and in fact all their requirements are related to tools for carrying out scientific research more effectively. On the other hand, non-domain researchers have a marginal interest towards the digital content and thus their requirements are mainly related to enhancing the quality of interaction.

A. Requirements Shared by the Two User Groups

Although their aims were different, the two groups highlighted a number of common requirements regarding the availability of navigation tools, the interoperability with other collections and the need for additional textual information.

The development of navigation tools to access archives of cultural heritage material has been proposed in other application domains, such as accessing archeological sites described in [12], which describes the idea of including navigation maps as a third kind of data source together with digitized content and metadata. A similar approach seems to be interesting also for other cultural heritage collections. In the particular case of IPSA, the possibility of navigating inside the collection of illuminated pages of different books can help non specialists – from non-domain researchers to the general public – to enjoy the digital collection with limited effort spent in searching for interesting content. The current version of IPSA already supports a mechanism that partially addresses these requirements. In particular, linking annotations are used to represent relations between images, while their type describes the kind of relation [13].

Interoperability between digital archives can play a crucial role in their effectiveness as research tools for scholars. To this end, the methods and techniques used by multimedia search engines (i.e. methods of automatic near duplicates detection) can help in synchronizing different digital archives with overlapping content. In particular, it seems that automatic tools that mine the content of online digital collections (e.g. metadata on authorship, subject, iconography, area of production) can be particularly useful for professional researchers. At the same time, the need of non-domain researchers to access additional digital collections can be explained by the fact that they have only a generic interest towards the digital content of the archive. Providing a larger amount of information can enhance the interaction, even if this goes beyond the goals of the digital archive itself. In particular, the possibility to use the digital archive as a starting point to browse other related collections can improve the quality of interaction for non-domain users.

The inclusion of textual information appeared to be particularly relevant for non-domain researchers. The access to the IPSA digital collection required a shift in their interaction paradigm, from metadata to audiovisual content, which may result in a lack of interest towards the digital collection. Clearly, the addition of the original text poses challenging problems, because text is handwritten and usually in Latin. However, both professional and non-domain researchers expressed an interest towards a wider variety of textual information, including articles and essays written about the manuscripts and the images, which can be managed with digital libraries techniques. Textual information is considered important for describing and understanding the relations between images by both groups of users.

B. Possible Generalisation of User Requirements

An important outcome of the user requirements is that they implicitly underline the central role of domain experts in the dissemination of cultural heritage. On the one hand, professional researchers expressed the need for tracking the results of their research directly in the digital archive, on the other hand non-domain researchers expressed the need to access the digital content through the mediation of experts' point of view. The possibility for professional users to enrich the archive with comments, additional metadata, and links between multimedia content gives the additional positive effect of increasing the involvement of non-domain users. This result is probably related to the fact that the second user group that took part in the interviews is made up of researchers, although in different domains. Therefore they are likely to share a similar approach towards the analysis of cultural heritage content. Even if non-domain users do not have the tools to carry out their own analysis of the IPSA collection, they have the possibility of exploiting the results of scientific research to increase their enjoyment of the archive. An analysis of the interaction of non-domain users with the digital content showed that they were more interested in the research results rather than the digital content itself.

We argue that these results can be generalized to other categories of users as well, from students to the general public. A digital archive of cultural heritage material cannot compete in raising user interest with large multimedia collections available on the Web, containing millions of user-generated items such as Flickr[2] for photographs or YouTube[3] for videos. Cultural heritage collections are more focused and smaller in size, yet the possibility

[2] http://www.flickr.com/
[3] http://www.youtube.com/

of being guided inside the collection, understanding the value of the individual items and following the path of professional researchers can bridge this gap. We considered this of particular relevance in the re-design of the IPSA digital archive.

C. Exploitation of the User Requirements Outcomes

The results of the user requirements are under consideration for the re-design of the IPSA architecture at three levels.

The *external level*, which is based on Web browser interaction to give access to the data and functions to the user, will undergo development for the managing of articles and papers on the manuscripts and on the images. Navigation tools will be improved to provide non-domain researchers (and the general public as well) with additional ways to reach the digital content, stored both inside the IPSA collection and in external resources.

A *middle level* will be added to the IPSA architecture. The goal is to allow professional researchers to track the results of their studies. Comments, annotations and possible refinements of the metadata will be analyzed automatically to enrich the archive with this valuable content and to provide personalized access to the digital content. We consider that all these annotations should be an integral part of the IPSA digital archive and, providing that an information sharing mechanism is created, should be made available to the general public. Digital rights management functions will prevent from unpublished research being shared outside research groups. Moreover, a logging mechanism will track the interaction by professional researchers to improve personalized access, to create connections between related digital content and to provide final users with suggestions on interesting digital items.

Finally, the *internal level*, which is the one that manages and stores the permanent data assuring the permanence of the data, will undergo minor improvements with the aim of including as additional information the results of the analyses of professional users behavior.

V. Conclusions

Multimedia resources for cultural heritage play a double role. They (i) allow specialized users to carry out their research, keeping track of their research results, and they (ii) play a major role in dissemination to a wider public. To highlight the requirements that have to be met by a digital resource for cultural heritage, we carried out two sets of user studies using IPSA, a digital archive of illuminated manuscripts, as a case study.

Notwithstanding an obvious difference in the interest towards the content of the digital archive, the two groups showed a number of related issues that suggested possible directions for the development of digital resource for the study, preservation and dissemination of cultural heritage. To this end, an important outcome is the role of specialized users as mediators between the general public and the digital content. The same tools used by researchers to carry out their studies should be exploited to enhance the interaction of non-domain users, who can take advantage of the results of scientific research to improve their enjoyment of multimedia content.

Acknowledgment

The authors would like to thank Giordana Mariani Canova and Chiara Ponchia, of the Department of Cultural Heritage of the University of Padua, for the useful discussions on many aspects related to this investigation.

The work reported has been partially supported by the CULTURA project, as part of the Seventh Framework Programme of the European Commission, Area "Digital Libraries and Digital Preservation" (ICT-2009.4.1), grant agreement no. 269973.

References

[1] M. Agosti, "Information Access using the Guide of User Requirements," in *Access through Search Engines and Digital Libraries*. Springer-Verlag, 2008, pp. 1–12.

[2] G. Mariani Canova, *La miniatura veneta del Rinascimento: 1450-1500*. Alfieri, Venezia, Italy, 1969.

[3] F. Lebourgeois and H. Emptoz, "DEBORA: Digital AccEss to BOoks of the RenAissance," *Int. Jour. on Document Analysis and Recognition*, vol. 9, no. 2-4, pp. 193–221, 2007.

[4] J.-M. Ogier and K. Tombre, "Madonne: Document Image Analysis Techniques for Cultural Heritage Documents," in *Proc. of the Int. Conf. on Digital Cultural Heritage*, 2006, pp. 107–114.

[5] D. Borghesani, C. Grana, and R. Cucchiara, "Surfing on artistic documents with visually assisted tagging," in *Proc. of ACM Multimedia*, 2011, pp. 1343–1352.

[6] C. Mulwa, S. Lawless, M. Sharp, and V. Wade, "The Evaluation of Adaptive and User-Adaptive Systems: A Review," *Int. Jour. of Knowledge and Web Intelligence*, vol. 2, no. 2/3, pp. 138–156, 2011.

[7] A. Elgammal, "Human-centered multimedia: Representations and challenges," in *Proc. of ACM Workshop on Human-Centered Multimedia*, 2006, pp. 11–18.

[8] M. Agosti, L. Benfante, and N. Orio, "IPSA: A Digital Archive of Herbals to Support Scientific Research," in *Proc. of the Int. Conf. on Asian Digital Libraries*, 2003, pp. 253–264.

[9] G. Mariani Canova, "La cultura universitaria padovana e la nascita del realismo nell'immagine botanica," *Atti e memorie dell'Accademia di Storia della Farmacia*, vol. XX, no. 3, pp. 198–212, 2002.

[10] M. Agosti, G. Mariani Canova, N. Orio, and C. Ponchia, "Methods of personalising a collection of images using linking annotations," in *Proc. of the First Workshop on Personalised Multilingual Hypertext Retrieval*, 2011, pp. 10–17.

[11] O. Oomen and L. Aroyo, "Crowdsourcing in the cultural heritage domain: opportunities and challenges," in *Proc. of the Int. Conf. on Communities and Technologies*, 2011, pp. 138–149.

[12] A. Stenzer, C. Woller, and B. Freitag, "MonArch: Digital archives for cultural heritage," in *Proc. of ACM Int. Conf. on Information Integration and Web-based Applications & Services*, 2011, pp. 144–151.

[13] M. Agosti, N. Ferro, and N. Orio, "Annotating Illuminated Manuscripts: an Effective Tool for Research and Education," in *Proc. 5th ACM/IEEE Joint Conf. on Digital Libraries*, 2005, pp. 121–130.