Can a Digital Archive System Developed for Researchers Be Extended to Non-domain Users?

Maristella Agosti, Lucio Benfante, and Nicola Orio

Department of Information Engineering – University of Padua
Via Gradenigo, 6/a – 35131 Padua – Italy
{agosti,benfante,orio}@dei.unipd.it

Abstract. Digital resources are becoming an important tool for research in all the domains related to cultural heritage. Scholars have special requirements that need to be matched when developing digital library and digital archive systems that are to be used as tools to carry out scientific research. After having designed and developed a digital archive called IPSA as a system for researchers in illuminated manuscripts, we investigated how the digital archive can be evaluated by non-domain users. Our goal was to highlight the overlaps and the differences in the user requirements between specialists, who use the digital archive to fulfill their research goal, and non-domain users, who interact with the digital archive because of a general interest about its content. The results show a number of directions that need to be explored in order to extend the access to digital libraries and archives also to non specialists.

1 Introduction

The availability of digital multimedia resources of cultural heritage poses challenging questions about the kind of interaction that needs to be foreseen to make them available to distinct categories of users. An important question that we address in this contribution is how requirements gathered from a distinct category of users can be generalized to other groups. In particular, we focus on how requirements gathered from domain specialized users (professional researchers), who need to carry out their research exploiting the functionalities of a digital archive system, can be extended also to non-domain users, who are interested in enhancing their experience when accessing digital collections.

A digital archive system of illuminated manuscripts has been used as a case study for this investigation. The digital archive is called IPSA, which stands for *Imaginum Patavinae Scientiae Archivum* (archive of images of the Paduan science) [1]. IPSA was developed, from 2001 to 2005, as a scientific 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 and botany, through the study of illuminated manuscripts [5]. From the goal of the project, it is clear that IPSA aimed at being used by professional researchers, scholars in history of medieval art specialized in history of illumination. The aims and the results of IPSA are documented in

Fig. 1. London, British Library, Egerton 2020, f. 27v, Grapevine.

the website[1]. As an example of the iconographic content of IPSA, Figure 1 shows the digital representation of a page of the *Liber Agregà de Serapion* manuscript (London, British Library, ms. Egerton 2020) where a plant of grapevine is depicted. It is important to note that the text of an illuminated manuscript can be copied verbatim from older manuscripts, because the most relevant part of the illuminated manuscript is the iconographic part, so the text is accompanied by illustrations that can be copied from or inspired by older manuscripts, or taken directly from nature.

IPSA is a combination of digitised images of manuscripts and related metadata information, and its content can be of interest to a much larger group of users in respect to the one that has been the initial target of the work. In line with this consideration, IPSA has been selected, together with the collection named "1641 Depositions", which is held by the Trinity College Dublin, Ireland[2], to contribute to the design, development and evaluation of the innovative research environment that is under design and development in the context of CULTURA (Cultivating Understanding and Research through Adaptivity)[3], a EU funded STREP project which was launched in February 2011 and which is now under development [4].

CULTURA aims at personalisation and community-aware adaptivity for digital humanities through the implementation of innovative adaptive services in

---

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

[2] 1641 Depositions Website, URL: http://1641.tcd.ie/about.php

[3] CULTURA Project Website, URL: http://www.cultura-strep.eu/

an interactive environment. The intention is to offer genuine user empowerment and different levels of engagement with digital cultural heritage collections and communities. The objective of the CULTURA project is to pioneer the development of personalised information retrieval and presentation, contextual adaptivity and social analysis in a digital humanities context. This is motivated by the desire to provide a fundamental change in the way digital cultural heritage is experienced, analysed and contributed to by communities of interested individuals. These communities typically comprise a diverse mixture of professional researchers, apprentice researchers (e.g. students of history and art history), informed users (e.g. users belonging to relevant societies or interest groups, cultural or authorities) and interested members of the general public. One of the main tasks of CULTURA regards the characterisation of users and communities, in order to identify the variety of individual users and community types who will interact with the CULTURA environment. The groups that are taken into account include:

- Professional researchers: scholars, academics, tutors and historical curators;
- Non-domain professional researchers: same as above though in a different domain;
- Student community: post-doctoral, postgraduate and undergraduate students;
- Non-professional researchers: members of a historical/cultural society, authors, publishers, members of the public with a sustained interest in history or in arts;
- Interested members of the general public. This can include both adults and school children.

2   Requirements of Professional Users

As mentioned in the introduction, IPSA was developed as a tool for professional users, and, instead of limiting the requirements analysis to a number of interviews, the design approach was to create a research team where computer scientists and professional users (that is researchers in history of art specialized in history of illumination) collaborated together. Additional contributions from scholars in related disciplines, such as history of science, botany, astronomy, were integrated as well and formalized in a draft proposal that was presented and discussed with professional users. A similar approach has been maintained during the development of the prototype system, because all the novel functions have been directly tested by members of the research team.

The requirements for carrying out scientific research are in general more complex and articulated than the requirements of final users. Final users access an image digital archive to acquire information in a given field, researchers access the same archive to disclose knowledge and discover new relations between digital objects. IPSA was designed and developed taking into account the requirements

of professional users in history of illumination. This means that IPSA is the outcome of the effort of producing an original and innovative system for a specialised group of professional users.

To understand the effort that has been recently started, in the context of the CULTURA project, to re-design IPSA to add new functions to the original ones to face the characteristics of interest of the new user groups of interest, it is necessary to know the inspiring requirements that pervade the original IPSA system. For this reason, those relevant characteristics are briefly presented in the rest of this section.

2.1 Disclosure of Relations between Images

It is of primary importance for professional users in history of illumination to discover whether illustrations have been copied from images of other manuscripts, or they have been merely inspired by previous works, or if they are directly inspired by nature. A major IPSA function thus regards the possibility of enriching the digital archive by highlighting explicit relations that have been discovered by a domain professional user. In particular, he should be able to create *links* that connect one image to another that it is related to, in some way. The analysis of user requirements on link management highlighted a number of advisable features that needed to be implemented. These are:

- **Link authorship**: The creation of a link between two images depends on the scientific results of a professional user who owns the intellectual rights to the disclosure of a new relation between images; for this reason the author of each new link is recorded by the system.
- **Link typology**: Since two images can be related to a number of different reasons, the kind of relation has to be explicit. Different typologies of links are envisaged to express the possibility that an image is the progenitor of another image, of a set of other images, or that one image is a copy of another.
- **Paths**: Links 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; hence two images may not be directly linked, because there is no direct relation between them, but it could be possible to follow a path from one to the other by exploiting existing links.

These requirements suggested the use of typed annotations connecting two manuscripts, two images, or even two parts of different images. These annotations, which have been called *linking annotations*, have a type that describes the kind of relation between the two objects and provides a semantic to the link. For this reason, we proposed a taxonomy for linking annotations [3] 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 [2].

2.2 Dynamic Records and Intellectual Rights

Almost every digital archive dynamically changes over the years, mainly because of new acquisitions that increase the number of documents that are stored and managed by the archive system. This is also true for a digital archive of illuminated manuscripts, but there are other reasons that produce changes on the archive over time. These include the creation of records describing the documents and the images of an illuminated manuscript, that is part of the scientific research itself as for any collection of historical works. Some examples of changes to records are that new relations with other works have been discovered, or that the attribution to a given author became less certain.

Because creating a new record or modifying an existing one is part of the scientific work of researchers, the data management has to deal with intellectual rights. A researcher may prefer that some of the newly created records are not accessible by other users, at least until the results of his research have been checked and afterwards published. This situation implies that users may decide which information can be shared with other users and which can not.

This novel knowledge, which is due to original results, should be stored in the digital archive at a different level than the information that is based on a general consensus. To this end, the use of annotations, both classical textual annotations and the linking annotations, can be a viable tool providing that a user may state which annotations can be shared with the community or with his research group, and which ones has to remain private. Such a mechanism allows researchers for both using the digital archive as an advanced research tool and protecting their intellectual rights. Moreover, linking annotations add a hypertextual structure to the archive, which is different for each user and reflects his personal knowledge in the field. The relation between two images, that is used to make visually explicit that the image on the left is precedent to the one on the right is shown in Figure 2 where the time axis is added for this purpose, and that the system needs to be internally able to keep and actively manage.

2.3 Collaborative Environment

The study of illuminated manuscripts involves a number of researchers from different fields. In fact, illuminated manuscripts are of interest for both the historian of art and the historian of science, but at the same time, a herbal is of interest for the botanist because it represents plants and their possible variations through the centuries, a codex is useful for researchers on the evolution of civil and penal 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, who should be able to cooperate in order to share their different knowledge and background. A digital archive of illuminated manuscripts can provide a collaborative environment, such as in [6], where researchers can
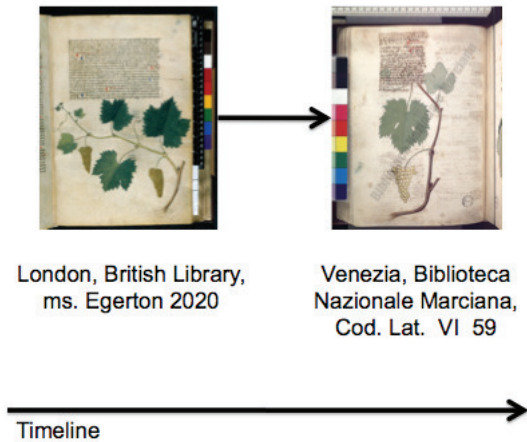
Fig. 2. Two representations of *vitis vinifera* (Common Grape Vine) that are present in two different manuscripts created in different years, connected by an explicit link.

interact and give different contributions on the definitions and redefinitions of objects in the manuscript.

In the original design of IPSA, such requirements were studied but not implemented in the system; this study becomes relevant now as it shows that it is possible to consider different types of professional users of the archive. Apart from the administrators, to implement the results of that study in the system, the group of research users has to be considered as made by different types of professional users that have to be able to modify the records of the underlying database when new features of the stored objects are discovered.

2.4 Presentation of Digital Images

A digital archive of illuminated manuscripts has the double role of preserving the cultural heritage and giving access to users in a networked environment. As it always happens in this situation, there is a trade-off between the high quality required for preservation and the small size needed for transfer over the network of the image files. Moreover, it has to be considered that research users should be able to perform comparisons among images belonging to different manuscripts that, in principle, may differ in their original size. According to professional users involved in the original design of IPSA, the number of images that should be presented on the computer screen varies from one to a maximum of six.

This last requirement implies that the image size, and hence its resolution, can dynamically vary depending on the context, because in principle a link can be created between any pair of images. The image files transfer load can be reduced through the use of thumbnails, at least for the first presentation of images. By

the way, thumbnails may also be a viable solution when the comparison between images is not part of the scientific research but it can be used for dissemination to students or, if future releases of IPSA will be available on the Web, to casual users without controlled access.

Image acquisition is another important issue, because researchers should be able to analyze even small details of images. At the same time, researchers need also to see the image of the complete page of a manuscript, because it gives the context in which a particular object is presented. Moreover, many manuscripts have more than a single image for each page, with images surrounding or overlapping with text. For these reasons, it is advisable to carry out multiple acquisitions of the same page, with different resolutions depending on the level of detail needed for the analysis by researchers.

3   IPSA Digital Archive System

The IPSA prototype implementing the requirements briefly recalled in Section 2 was developed. The tight 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 functions of the evolving prototype. Once the underlying database structure was designed and developed, the organization of the user interface and the development of the novel functions highlighted by the user requirements were done incrementally, with scholars in history of art starting to populate the archive with the initial collection of images while the refinement of the software tools was taking place.

The IPSA digital archive system was made available on its stabilised form in 2005 and from then on it has been used for research purposes by history of art researchers. Over the years the collection of manuscripts and images has been incremented.

Due to the launch of the CULTURA project, the IPSA digital archive system has been reconsidered for use by different categories of users, and a re-engineering of the system has taken place to bring the system up-to-date with the new technologies that in the meanwhile have been made available, while the underlying model of content management has been kept. Taking into consideration that the users mainly focus their attention on the graphical interface when interacting with a digital resource, the system interface has been re-designed to make it more in line with recent advancements.

The new IPSA user interface aimed at simplicity and easy user accessibility. The main layout is designed for optimal visualisation with a screen resolution of 1024x768 pixels and up, horizontally centered and filling the vertical space. The layout contains three zones: the top header, the main area and the bottom footer.

The main header is as thin as possible. It contains the main starting points to the IPSA functionalities: a small IPSA logo which links to the home page, the login/logout button, a structured multi-level menu and a form for searching the IPSA illustrations. When users are logged in, their name is shown in the header,
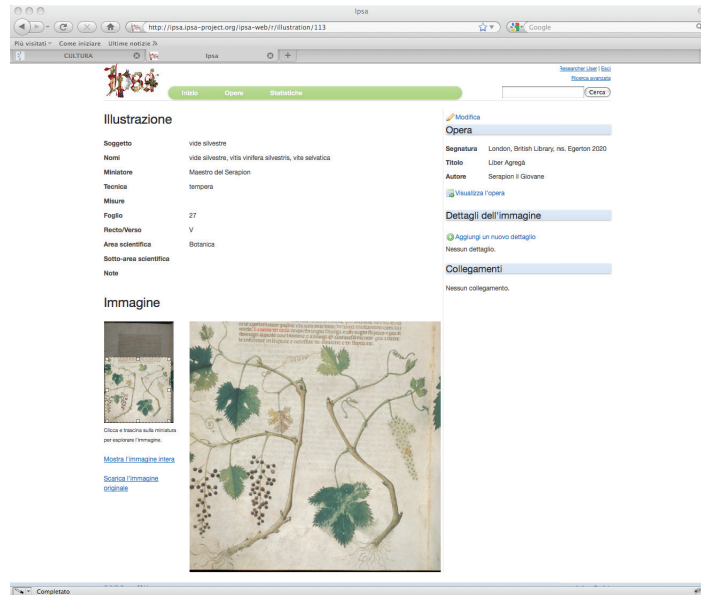
Fig. 3. Screenshot of the IPSA Web application where the grapevine is depicted.

linked to their profile for editing, if it becomes necessary. Near the search form there is a link to the advanced search function. The menu adapts itself following the user permissions, and it guides the user in the navigation, showing the most common functionalities in its first or at maximum its second level. The footer is designed for containing secondary menus and non critical information for the user. At present it contains the copyright information and the language selectors. The IPSA user interface is fully localized in Italian and English.

Most of the screen is occupied by the main area. The layout of this zone is strictly related to each functionality, and is designed and implemented following the user needs of usability. It is designed for showing the main information on the left, with a small sidebar on the right containing the links to the operations on the currently displayed object, and the related information.

A screenshot of the present Web interface presenting an image and related metadata of the IPSA collection is shown in Figure 3.

4   Initial Evaluation by Non-Domain Users

Since 2005, IPSA has been used by professional users to carry out their studies on illuminated manuscripts. Starting from 2011, the new challenge is to investigate whether IPSA features can be of interest to non-domain users as well. The goal of this paper is thus to use IPSA as a case study to compare the approach of different kind of users to the same digital content. We conducted two subsequent evaluations: the initial one with different perspective users belonging

to the class of non-domain professional research experts and to the class of the student community (in this case master students in archival science), the second one with the other groups that have to be taken into account in the context of the CULTURA environment and that were mentioned in Section 1.

The initial evaluation has taken place and been completed with the main goal of highlighting possible overlaps between the requirements of domain professional users and the two considered groups. For this reason we report in the rest of this section the findings of this effort and we report in Section 5 the re-engineering results on the IPSA system after having generalised the findings to give directions in re-designing some aspects of the user interaction with the system. The second evaluation is still under development.

4.1   Effect of Computer Skills

The first outcome of the survey with non-domain users is that competence in the use of automatic tools plays an important role in the way users interact with the system. In particular, specialists in other 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 users were accustomed to searching for bibliographic values rather than obtaining and studying images directly through the interface to the digital archive. This shift in the interaction paradigm may give rise to a lack of interest towards the collection. This result can be taken into consideration by specialists that design a new multimedia digital archive.

In contrast, master students are more accustomed to interact 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 were more related to the need for search facilities that are typically available in multimedia systems. For instance the possibility to obtain recommendations based on users' tags, although social tagging does not easily apply to the cultural heritage domain because of the kind of digital objects and the limited number of users.

4.2   Navigation Tools

An additional set of requirements regarded the exploitation of links within and, especially, outside the digital archive. This finding extends the requirements of specialized users, who probably already have other tools to access the relevant literature or other digital resources. In the case of non-domain users, the need to retrieve additional information can be explained by the fact that non-domain users have only a generic interest towards the digital content of the archive, which can be enhanced by providing a larger amount of information, even if it 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 collection was considered of interest.

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 retrieve multimedia information. To this end, the possibility of expressing relations between images, as reported in Section 2.1 for domain professional users may become a valuable tool for casual users to retrieve information of interest. A lack of knowledge on the domain of the cultural heritage resources managed by the archive system may prevent users from retrieving information, while the presence of annotations made by experts can fill this gap.

4.3   Need for Textual Information

A final outcome, which might 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 non-domain specialists and master students suggested directions in which the digital archive should be enriched to be more attractive.

Non-domain specialists asked for analytic descriptions of all the multimedia content of the digital archive, to make users aware of every characteristic of the images studied, to highlight relations with other research fields, and basically to follow and understand the research process of specialized users. The possibility of obtaining different graphical representations and to personalize the digital archive content, as described in Section 2.4, is of interest mainly if it is mediated by specialized users. Although this finding can be related by the effect of computer skills, it is important to note that non-domain researchers are more interested in how the digital archive can track the research process than in the content itself.

Master students on archival science had a bias, as can be expected, towards bibliographic values. In particular, they showed an interest in the way bibliographic records were created, their authorship, the motivations behind attributions and so on. Also in this case, the approach to IPSA content depended on the particular field of interest of non-domain users, where the research work of specialized users played a major role, notwithstanding the fact that the content of the digital archive was perceived of important cultural value.

5   Initial Extension of the Digital Archive System

In relation of what has emerged from the initial evaluation that has been completed, most of the IPSA system functions have been considered of interest, but some of the interaction functions needed to be re-engineered and the findings are at present used for the development of the new features of the system. In particular some new presentation features have been developed and some other are under way as reported in the reminder of this section. Since the focus of this paper is on the interaction of different kinds of users with the same digital archive, in the following we descibe the first extensions that were required to

carry out an effective evaluation with additional user groups. It is likely that the final outcomes of our evaluation will require a re-enginering of the system.

The search results are exposed as a paginated list of results in textual format, with the illustration subject, the signature mark of the work and the links to the illustration and to the general information on the manuscript. A new alternative visualisation of results is under development where the results will be presented as a "wall" of thumbnails of the illustrations, incrementally loaded in slots at user request, without generating a new page request.

The page showing a single work contains the textual metadata of the work and the thumbnails of the illustrations for each page of the work. Initially this wall of illustrations was designed for showing all thumbnails at a time. This organization proved very inefficient both for the user experience and for the server overloading of the operation, mainly because there are works with hundreds pages. Now it has been reorganized to incrementally load the images at user request in slots of thirty thumbnails.

A different problem emerged in the page showing a single illustration. The page shows the textual metadata of the illustration and allows the user to explore the image. Dragging the mouse over the thumbnail, the user can explore and zoom a magnified version of the image. The magnified image was rendered at the loading of the page, scaling on the server-side the full version of the image. This allowed a personalized rendering, customized according to the screen resolution and the domain professional user preferences. Unfortunately this process was too slow for an effective user experience and too heavily loaded the server in the presence of many concurrent requests of non-domain users. So we decided to pre-render the magnified image, using a predeterminate fixed screen resolution. This now allows fluid navigation with very fast page loading and much improved user satisfaction.

Other improvements are planned in the near future for the feature of creating links between images. For this feature users need more space on the screen for viewing more than a single image at the maximum possible size, and for maintaining a clear context during these multi-step operations. So we are planning for a design that could use all the horizontal space available in the page.

Last, a contextual online help system is going to be introduced for the use of non-domain users. The help on single elements of the page will be shown in tooltips. The help of a feature, usually attached to a page, will be shown in a sidebar that the user will activate or hire as he needs. All the help text will be multi-lingual and managed through an internal content management system.

6 Conclusions and Future Developments

Multimedia resources for cultural heritage play a double role. They allow specialized users to carry out their research, keeping track of their results, and they play a major role in the dissemination of cultural heritage to a wider public. To highlight the requirements a digital resource for cultural heritage has to meet, we carried out an initial evaluation with two groups of non-domain users of the

IPSA digital archive system, initially developed as a research tool for domain professional users. Through the initial evaluation, non-domain users, who have a generic interest towards the digital images of the archive, highlighted a number of functionalities that might enhance their experience with multimedia content, and those findings are used at present to complete an initial extension of the IPSA system. Another possibility that will be taken into account regards the development of different user intefaces depending on the user groups.

A second evaluation with the other groups that have to be taken into account in the context of the CULTURA environment is at present under way to collect new information on the way non-domain users are interested in interacting with digital culture heritage resources. We imagine that the findings of this second evaluation are going to produce other insights to be taken into consideration in the design of innovative adaptive services.

Acknowledgements

References

1. 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. Lecture Notes in Computer Science, vol. 2911, pp. 253–264. Springer (2003)
2. Agosti, M., Ferro, N.: A Formal Model of Annotations of Digital Content. ACM Transactions on Information Systems (TOIS) 26(1), 3–57 (2008)
3. 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, USA (2005)
4. Agosti, M., Orio, N.: The cultura project: Cultivating understanding and research through adaptivity. In: Agosti, M., Esposito, F., Meghini, C., Orio, N. (eds.) Digital Libraries and Archives - 7th Italian Research Conference, IRCDL 2011. Revised Papers. Communications in Computer and Information Science, vol. 249, pp. 111–114. Springer (2011)
5. Mariani Canova, G.: La cultura universitaria padovana e la nascita del realismo nell'immagine botanica. Atti e memorie dell'Accademia di Storia della Farmacia XX(3), 198–212 (2002)
6. 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(1), 8–12 (2004)