Abstract

Purpose – To provide an overview of the construction and features of an art image database for high-resolution images.

Design/methodology/approach – A description of technical information, standards and tools used for creating an art image database using open-source software products ZOPE, PLONE, and DIGILIB.

Findings – Provides information about the considerations for and implementation of the digitalisation process and the set-up of the information infrastructure. Recognises that the effort for the digitalisation project was higher than estimated.

Research limitations/implications – The focus is on specific issues of art image databases and the application of open-source software products.

Practical implications – A source of information for applied standards and software as well as digitalisation processes.

Originality/value – The paper presents practical information for institutions planning to set up an image database.

Keywords Cataloguing, Copyright law, Digital libraries, Photographs, Art libraries, Databases

Paper type Research paper

Introduction

For several years, the digitisation of images has been a major issue for photo libraries, especially when their collections are of interest to a broader community. With the internet and especially the world wide web offering new possibilities for cooperation and outreach, the creation of online image databases is an emerging trend. At the same time the current state of the art allows the building of a sustainable information infrastructure on the web. Following these trends, the German Institute for the History of Art in Florence (Kunsthistorisches Institut in Florenz, KHI) intends to improve scientific access to its photo collection and to preserve its materials for the future in a digital format. Within the last two years two terabytes of digital image data were produced by scanning the most important photo negatives. This is a major change in the photo library’s policy as for decades both its collection and the workflow were based on traditional photographic materials.

The paper commences with an introduction to the KHI and its photo collection and continues with a description of the challenges posed by digitisation and the decisions taken before and during the project. It includes information on the photographic material as well as on the applied technical standards and equipment. Moreover, the integration of image data and textual metadata are depicted. In accordance with
the policy of the Max Planck Society, the project made a deliberate decision for sustainable open-source software products and the intention to provide open access to knowledge in the sciences and humanities. Another important question was the long-term preservation concept for the digital image data, which had to be established. An open issue is how to integrate the image database and the library online catalogue.

The web site is available on the internet www.khi.fotothek.org. A translation into English and Italian is currently under construction and will be available in autumn 2005.

An overview of the institute
The KHI goes back more than 100 years. It was founded in 1897 by a group of art historians as a private research institute to increase the dissemination of knowledge about Italian art. Today, as a member of the Max Planck Society, the institute still pursues this objective by organising regular academic events such as talks, conferences, and specialized summer schools for young academics. The results of its various research projects are published on a regular basis. Together with institutes such as the Bibliotheca Berenson – Villa I Tatti (The Harvard University Center for Italian Renaissance Studies), the Dutch University Institute for Art History, the Bibliotheca degli Uffizi, the Biblioteca della Fondazione di Studi di Storia dell’Arte Roberto Longhi, the Biblioteca dell’Opificio delle Pietre Dure, the Biblioteca Istituto Nazionale di Studi sul Rinascimento, and others, the KHI is one of the major art historical research organisations in Florence.

The KHI’s library and photo library serves as a reference for art historians from all over the world. With its analogue photo collection of around 600,000 pictures relating to the art of upper and central Italy, the photo library of the KHI is an important study tool for art historians that expediently supplements the Institute’s special library holding some 220,000 volumes and some 1,360 periodicals. The special value for researchers consists both of the large-scale and full range of the collection and the extensive visual documentation of individual works of art. The latter is particularly valuable because it allows researchers to trace changes systematically over a long period of time – in some cases, back to the early days of photography.

Providing access to art photography
Until 1993, the photographic collection of the KHI was solely accessible through card catalogues. Since then, all new acquisitions and related holdings are catalogued using software called hierarchical data administrator (HiDA), which was developed for the scientific cataloguing of works of art. To date, art historical metadata for some 15 per cent of the KHI’s photo collection has been entered in HiDA based on the Marburg Information, Documentation, and Administration System description model (Marburger Informations-, Dokumentations- und Administrations-System, MIDAS, cf. Bove et al., 2001). These metadata contribute to the database run by the Digital Information System for Art and Social History (Digitales Informationssystem für Kunst und Sozialgeschichte, DISKUS), a nationwide consortium of German art museums, art historical institutes, and preservation institutes. In addition, these metadata could be searched on the Institute’s web site – although without the images they describe. That means, in order to take a look at the photographs, researchers still had to travel to Florence and gain access to the institute. Therefore, it was time to

Images of Italian art on the web
enhance the outreach of the photo library. With financial support from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), the photo library team started a digitalisation project in order to make the art photographs digitally accessible on the internet. The project was sponsored by the DFG, which endowed a two-year-grant for the digitalisation project and the set-up of the image database.

The digitalisation project
The DFG grant provided the financial means for a project team of three people and their computer workstations. The Max Planck Society financed additional information technology such as scanners, a two-Terabyte project server, and a printer. The objectives of the project, which started in September 2002, were:

- Defining the project goals.
- Choosing the technical specifications and the equipment for digitalisation.
- Selecting the appropriate material with regard to copyright issues and art historical importance.
- Creating digital images of the art photographs.
- Creating an organizational structure for data management.
- Combining the digital images with the art historical metadata.
- Planning the long-term preservation of digital images.
- Providing online access to the art photographs. For this purpose, a web presence and image database were created.

Defining the project goals
At the beginning of every project, there is a process of developing ideas and visions and the task of transforming these into concrete goals. In particular, two aspects were important for the whole range of the following considerations – the digitalization quality and the web presentation. Instead of scanning the photographs, we decided to take the negatives as the source for digitalization. The quality goal was to take as much information as possible and to create a digital facsimile of the analogue material. These high-resolution files are intended for external long-term data storage. As we were not satisfied with the search functions of the existing database, which also has reduced image viewer options, we decided to develop a more elaborate tool. As a consequence, we chose a more up-to-date content-management-system combined with an efficient image viewer to present our materials to the internet audience. To build up know-how in this increasingly important working field, the digitalization process was handled within the institute. To acquire the necessary background knowledge, the standard publications (Pfenninger, 2001; Gschwind et al., 2002) were analysed. In addition, Rudolf Gschwind, a digitalisation expert from the University of Basel was consulted.

Choosing the technical specifications and equipment for digitalisation
As mentioned above, to create digital images of the art photographs, the photo negatives had to be scanned. For conservation reasons, the collection consists almost exclusively of monochrome materials because analogous colour photographs fade relatively quickly in relation to the long-term perspective of a photo library.
The change in media gave rise to two considerations that are closely intertwined: The technical specifications for digitalisation and the selection of the technical equipment for the scan process.

As a starting point, the technical specifications for the digitalisation process were established. The goal was to create digital products that can be used both for reproduction and in the distant future. Therefore, a range of different aspects had to be considered. First of all, the kind of information and objects available were investigated. Second, we came up with the general set-up for the master scans and internal standards that support future use, such as options for publication and on-screen display for the working copies. Therefore, different scan resolutions for the different kinds of negative material (glass or film negatives, size, age, or rather physical condition) had to be established. The technical specifications for the digitalisation process are shown in Table I.

To implement these technical specifications, high-end technology was required for the entire workflow. The first link in the chain was two high-end scanners. The models chosen were a Flextight 2848 and a Flextight 848, both produced by Hasselblad Ilmcon. The Flextight 2848 can handle negative materials of formats ranging from 35 mm to 31 × 45 cm and of different densities such as film and glass. The advantage of the Flextight 2848 is that it has both a virtual drum and a flatbed. So the glass negatives did not have to be copied to film in order to be scanned. The Flextight 848 is a bit smaller and can handle negative materials ranging from 35 mm to 12 × 25 cm but no glass negatives. For more technical data on the scanners, see Table II.

The scanners are each connected to a Power Mac G5 with dual processors (64-bit power and high-bandwidth architecture, 2 GHz processor speed, and 8 GB main memory) for high-speed image processing. Each G5 is equipped with a LaCie electron22blueIV monitor. This 22-inch monitor is especially designed to display sharp images and accurate colours. These features are especially important for the next phase of our project, in which digital colour photography will become the tool of choice.

<table>
<thead>
<tr>
<th>Photo negative (cm)</th>
<th>Print media (dpi)</th>
<th>Enlargement</th>
<th>Size for output (cm)</th>
<th>Dpi for scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 × 6 cm</td>
<td>Art reproduction 600 dpi</td>
<td>× 4.5</td>
<td>27 × 27</td>
<td>2,700</td>
</tr>
<tr>
<td></td>
<td>Offset reproduction 300 dpi</td>
<td>× 9</td>
<td>54 × 54</td>
<td>2,700</td>
</tr>
<tr>
<td>13 × 18 cm</td>
<td>Art reproduction 600 dpi</td>
<td>× 2.5</td>
<td>32.5 × 45</td>
<td>1,500</td>
</tr>
<tr>
<td></td>
<td>Offset reproduction 300 dpi</td>
<td>× 5</td>
<td>65 × 90</td>
<td>1,500</td>
</tr>
</tbody>
</table>

Table I. Technical specifications for the digitalisation process

<table>
<thead>
<tr>
<th>Scanner features</th>
<th>Flextight 2848</th>
<th>Flextight 848</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optical resolution</td>
<td>From 900 to 6,800 dpi</td>
<td>From 80 to 8,000 dpi</td>
</tr>
<tr>
<td>Colour depth</td>
<td>16 bit</td>
<td>16 bit</td>
</tr>
<tr>
<td>Density</td>
<td>4.8 Dmax</td>
<td>4.8 Dmax</td>
</tr>
<tr>
<td>Scan speed</td>
<td>100 MB/min (16 bit)</td>
<td>100 MB/min (16/8 bit)</td>
</tr>
<tr>
<td>Scan principle</td>
<td>Virtual drum and flatbed</td>
<td>Virtual drum</td>
</tr>
<tr>
<td>Film formats</td>
<td>Transparencies: 35 mm to 29 × 42 cm (up to 12 × 17 cm virtual drum)</td>
<td>Transparencies: 35 mm to 12 × 25 cm</td>
</tr>
</tbody>
</table>

Table II. Technical data for the scanners
With digital photography, colour images enter the scene because conservation of colour photos is no longer a critical issue when the images are preserved digitally. But for digital photography, colour management is crucial to obtain results where the colour of the photo is – as far as possible – identical to the original. Therefore, it is essential to apply strict colour management from the start to the end of the process and use of high-quality monitors and printers to be able to appreciate the full quality of the materials.

Selecting the material

Before the digitalisation could start, two issues had to be resolved – the selection of the material and the copyright issues. The KHI photo library owns more than 580,000 photos but has the copyright for merely 30,000 images. Owing to legal restrictions, material could be digitised and made available online only in cases where the KHI holds the copyright. In compliance with the copyright ownership, several photographic archives were selected for digitalisation. The reasons for this selection were based either on the art historical importance of the collection – for example, the De Giovanni Archive, which consists of photographs of the Saint Francis church in Assisi – or on its historic importance – such as the Lotz Archive, which contain images depicting the urban development of Florence before the war-inflicted devastations of 1944 or the Bazzechi Archive, which depicts the urban development of Florence before the havoc caused by the great flood of the Arno river in November 1966. On this basis, a set of some 20,000 images was selected for digitalisation.

Creating the digital images

The digital images were created in a two-step process. First, a master copy of the photonegative is made with the intent to save as much information as possible. This master copy is used for producing further copies and for long-term preservation. Its size is approximately 100 MB. The technical details are as follows:

- TIFF format uncompressed.
- 16-Bit colour depth (65,536 colours) per channel. Due to the high colour depth, the file size is relatively large but this had to be accepted in order to gain some leeway for doing the digital work necessary to produce the copies (for example, the working copy).
- Dots per inch (dpi). The number of dots per inch is determined, on the one hand, by the informational content of the photonegative and, on the other hand, by the output device and size, respectively. In order to achieve the highest possible quality, the degree of dpi was identified separately for each output format and material (Table I).

A working copy is created from the master scan. This copy is digitally enhanced to allow viewing on screen and printing. Its size is approximately 20 MB. The technical details are as follows:

- TIFF format uncompressed.
- 8-Bit colour depth (65,536 colours) per channel.
- The short side is reduced to 4,000 pixels (approximately 33 cm on a 300 dpi printout).
- 300 dpi.
Developing an organizational structure for data management
Another important concern was the data administration process. To organize the storage and use of the image data during the project and later on in the image database, we had to develop a system for the file names and coordinate it with the existing numbering system of the photo library’s catalogue system. In addition, we had to create a set of directories to structure and access the data on the server. Finally, a sequence of work events was established to coordinate the working process of the photo library and digitalisation team. This included a final quality control of both the data entered in HiDA and the digitised images.

Combining digital images with the art historical metadata
The digital images have to be combined with the art historical metadata, which is captured in a database system called HiDA. This database system has been used by the KHI for more than ten years and is also used by the DISKUS consortium. Currently some 15 per cent of the KHI’s photo collection has been entered in HiDA and the cataloguing of all new acquisitions in HiDA is to be continued. The metadata in HiDA must be connected to the digital images to provide access for searching and information for interpreting the images. The link between the two data sources is the image file name, which is entered for each HiDA photo document. The HiDA data and the images are stored in the web database ZOPE, an open-source software that is used for web publishing.

Planning the long-term archiving of digital images
The digital images that are produced during the digitalisation process are not only used for display on the web and for printouts for academic publishing, but also have to be preserved for future generations as documents of works of art at a certain point in time. The long-term preservation concept of the Max Planck Society is currently under construction. The method of choice is the migration strategy. With this method, the data is moved from one medium to another and from one format to another as and when technical progress makes this appropriate. As a result, the data is always available in an up-to-date presentation format and can be processed using state-of-the-art technology. This is a significant benefit because users always have full access to this data. The drawback is that the digital objects might have to be changed each time they are migrated from one platform or format to another. As a result, the migration might cause changes to the digital objects and some of the original information might be lost during this process. As all the objects have to be handled with each migration step, this concept requires a lot of resources. Therefore, critics state that migration will consume too many resources to be practicable.

The image database
As the photo library team was working at full capacity on the digitalisation project, the technical implementation and set-up of the image database were outsourced to a commercial service provider. The prerequisite for the project was to use only open-source software to remain independent of proprietary software. Therefore, the web database ZOPE was chosen for web publishing (http://zope.org). The texts on the web site are written in Plone, the corresponding content management system (http://plone.org). The following sections describe the most important features of the image
database, created in collaboration with the Central Institute for the History of Art (Zentralinstitut für Kunstgeschichte) in Munich:

- Access to information and images.
- The image viewer digilib.
- The user interface.

**Access to information and images**
The access to information and images is provided through three indexes – the artist index, the location index, and the general index.

1. The artist index allows searches for artists’ names. In addition, it provides an alphabetical overview of all the artists’ names in the same form as they can be found in the German standard encyclopaedia Thieme-Becker/Allgemeines Künstlerlexikon.

2. The location index consists of an alphabetical list of names of places and locations in Italy. In addition, it allows a search for the different political regions of Italy. Each location link offers a list of all recorded buildings and monuments.

3. The general index offers a variety of different search options, for example, date, subject, location, and image content.

The results of a search are displayed in gallery mode. This presentation mode offers a brief description of the works, including information such as artist, date, material, location, and information about the images. The image information consists of two sections. The first section is a gallery view showing thumbnails of the digitised photographs including some basic textual information from HiDA. Each thumbnail is equipped with a link to a larger image (width: 600 pixels) and a link to digilib, image viewer software described in the next section. The second section lists all photographs that are available in the photo library but have not been digitised for different reasons.

**Digilib**
Digilib is an open-source image viewer software developed by the Max Planck Institute for the History of Science, Berlin, the University of Bern, and other partners. Originally designed for digital libraries, digilib lends itself well to image display because most of the functions that were designed for e-books can also be applied to digital images. Digilib offers numerous benefits:

1. It allows fast transmission of high-resolution digital images (the working copies) because they are broken up into tiles that are transmitted on demand. This allows a reasonable transmission rate for low bandwidth access.

2. It allows fast zooms on every detail of the image, so far that you can even see the granularity of the film material.

3. It protects high-resolution digital images (the working copies) against copyright violation because the image can be downloaded only in tiles and not as a whole.

4. It supports scientific collaboration by providing annotation and marking functions as well as a citation tool for exchange between scholars or students.

These features make digilib a valuable tool for scientific work and collaboration in art history (Plate 1).
The user interface

The user interface of the image database (www.khi.fotothek.org) will be translated into English and Italian as soon as the texts are stable (presumably in autumn 2006), whereas the content of HiDA must remain in German only. The reasons for this are as follows. First and foremost, HiDA offers no technical support for translation, and, in addition, the database contains too much information to translate without receiving proper funding for the present time and for the future. Nevertheless, the image database should be usable for scholars and researchers all over the world as the most important search fields, such as names of artists and locations, follow established conventions and should be understandable for users who do not speak German but have some knowledge of Italian art.

The advantages of online access to visual arts

For the first time in the 100-year history of the KHI's photo collection, the whole visual information conserved in the photonegatives is completely accessible. In the past, researchers consulted photographs of mostly 18 × 24 cm. Large reproductions existed only for a small part of the collection or were produced for special requirements – although the medium and large-format negatives would easily permit enlargement up to poster size. Such enlargements can be very helpful for art historians and restorers alike. For example, in 1997 an earthquake damaged and partly destroyed the ceiling frescos in the world-famous churches of St Francesco in Assisi (Umbria), which were decorated in the middle-aged by Cimabue and other artists. Fortunately, the German Art Institute had undertaken a detailed photo campaign with several hundred photos...
during an extensive restoration campaign in the 1970s and 1980s. These negatives provided valuable support for the restorers.

In the image database, interested people from all over the world can use the same images. Moreover, the zoom function of digilib allows users to discover and analyse every detail of such an image as the following examples show. One can really follow the brushstrokes of Giotto and other artists in frescos that are often too far away for details to be identified with the naked eye. One can now easily read inscriptions and epigraphs. Moreover, it is possible to discover signs of the working process on the stone sculptures of Michelangelo or the marks on Donatello’s bronze, which helps us to learn about the ways in which artists created their works.

As a conclusion, the digital environment can be considered an adequate instrument to show the sometimes overwhelming qualities of the material. With tools such as digilib, new value is added to the collection. Although photographs cannot replace the scholarly interaction with the original, they are sometimes the only remaining source of information because the original work of art might not exist anymore or has perhaps changed significantly as a consequence of damages. Other limitations can be that the original is part of a private collection and, therefore, hardly accessible or, as mentioned above, scrutinisation of minute detail may be difficult due to the location of the object.

Licensing of digital images
The licensing policy is currently under discussion. The intention is to allow scientific publication and usage free of costs. But there are still some open legal and technical questions which are mostly due to specific Italian copyright legislation (Legge Ronchey).

Open access policy
As a member of the Max Planck Society, the KHI promotes the open-access strategy that has been passed by the Max Planck Society’s Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities during the 2003 Conference on Open Access to Knowledge in the Sciences and Humanities.

Like the Budapest open-access initiative, the European Cultural Heritage Online (ECHO-Charta), or the earlier Budapest statement on open access publishing, before the German research associations emphasized the vision of a global and accessible representation of knowledge. Therefore, the future web must be sustainable, interactive, and transparent. Content and software tools must be openly accessible and compatible.

Initiatives like these are necessary with regard to the increasingly commercial interests also in the field of cultural heritage and the frequent disappearance of important content from the web because of financial problems on the part of the maintaining institutions.

Open access contributors must respect two conditions:

The author(s) and copyright holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to [the work], and a license to copy, use, distribute, transmit, and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (community standards will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now), as well as the right to
make small numbers of printed copies for personal use.

A complete version of the work and all supplementary materials, including a copy of the permission as stated above, in an appropriate standard electronic format, is deposited (and thus published) in at least one online repository using suitable technical standards (such as the open archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, inter operability, and long-term archiving (Berlin Declaration on Open Access, 2003).

We will try to take both points into account in our project. However, the legal situation in Italy is complicated because of new legislation that gives special rights to the owners of works of art. So first we have to establish how far we can follow the open access policy without violating Italian law.

Some considerations on a possible communication with the library catalogue

During the symposium of the Italian photo archives in Ravenna in 2004, large libraries (for example, the Biblioteca Nazionale di Firenze) presented their efforts to digitise and catalogue photos from their holdings. In this process, the existing library catalogues are generally adapted in order to make sure that they can adequately describe not only bibliographic units but also photographs. This provided a source of inspiration for us as our Institute faces a similar challenge. Two “autonomous” catalogue systems have coexisted for several years, each of them connected to different networks with their own structure and history. Nevertheless, we have to clarify whether and in what way both systems can expediently communicate with each other, without giving up specific requirements for a proper cataloguing of different materials. Basically, a closer connection between the different catalogue systems seems desirable: Combined access to the photo and book library catalogue would guarantee that users of our catalogues can obtain a more complete access to the Institute’s online facilities (Hansmann, 2005).

One way to achieve this goal would be to link the metadata. In the case of relatively few, yet fundamental, ways of searching, linking clearly-defined normative documents should be possible. In our image database, the search for location and artist (see above) are based on the Marburg geographical and biographical normative documents. Both types of normative data have their own uniform resource locators (URL) and would therefore also be accessible from other systems that are equipped with a crosswalk to the Marburg normative documents. As the photo library’s data – in comparison to the rather flat structure of the library data – is organized hierarchically, the most suitable solution would be to establish a meta-search engine, similar to a virtual library-catalogue. In this case, the viewing of search results could and should be shifted to each of the respective systems.

Conclusion

The efforts of the whole project were much higher than expected at the beginning. It took a considerable amount of time and effort to sort the archives of the photo library, to design and evaluate the workflow for digitalization, and to set up and manage the time-consuming processes of high-end digitalization for materials that were often sensitive and heterogeneous. Equally intensive was the creation of the new database
and the web presence. In particular, the information design required some thought – how to set up the structure of the web site and how to present the necessary content.

From the feedback that we have already received, especially from the public presentations at the conferences Kunsthistorikertag 2005 (annual German art historical conference) and at the EVA Berlin 2004 (Electronic Imaging and the Visual arts) conference, we dare to conclude that the result is worth the effort. With the digitalization process, we are creating a new virtual research platform for scholars around the world – and furthermore, we give access free of charge. We are well aware that a project like this one can only be realized within a non-commercial institution whose main objective is the promotion of scientific research. Relying on modern information technology, the KHI can, for the first time, show its important documentation of Italian art from the Middle Ages to the present day to interested users all over the world. We hope that new technical working tools such as our image database will add a new dimension to the importance of visual resources in the history of art.

References


