



IST-2001-32706

Toledo Workshop Report

June 23rd-25th 2002

DOCUMENT NAME: Toledo Workshop Report, 22nd-25th June 2002
DATE: 1.8.02
WORK PACKAGE: WP 5
TASK NUMBER: 5.1
EDITOR: Monica Greenan
AUTHOR TITLE: British Content Editor
CONTACT DETAILS: Tel: +44 141 330 4568
Fax: +44 141 330 3788
Email: British.Editor@erpanet.org

VERSION: Draft Final



Toledo Workshop Report

Table of Contents

Table of Contents	2
Executive Summary	3
Introduction	4
Aims and Objectives of the Workshop	5
Session 1: Digitisation of Audio and Visual Materials	6
Session 2: Digitisation of Images and Manuscripts.....	12
Session 3: Quality Assurance	18
Session 4: Migration.....	21
Recommendations and Preservation Strategies.....	27
Conclusions (Role of ERPANET).....	31
Annexe 1: List of Participants.....	32

Executive Summary

ERPANET held its first experts workshop in June in Toledo. The event offered experts the opportunity to take part in an investigation of key topics in digital preservation. Participants examined the interplay of digitisation, conservation, and preservation. This report is a result of two and a half days of presentations and discussion, and it recommends improvements practitioners, their institutions, and funding agencies can bring to their conservation and preservation practice.

Digital objects continue to acquire an increasingly pivotal role in all sectors of society. This workshop brought into sharp focus concerns surrounding not only issues of digitisation, but also analogue originals themselves.

Introductory speakers came from all over Europe and from diverse fields. They introduced key topics to the expert audience and this report is a blend of the talks and discussions that took place. The report recommends to institutions, projects and bodies with a stake in digitisation activities that:

- long-term planning and management structures should be in place at the outset of any digitisation activity to ensure that digital assets can be secured and fully exploited ;
- a comprehensive understanding of the nature and needs of the original object and its relationship to all digitised representations is in place;
- quality assurance processes are utilised;
- standards to protect the objects throughout their entire lifecycle are employed;
- storage and delivery issues are resolved at the project design stage;
- migration policies are in place to meet the uncertainties ahead; and
- the costs of these activities are understood and appreciated at the outset.

Introduction

On 23rd-25th of July 2002, ERPANET, the Biblioteca Nacional España, and the Caja Castilla la Mancha jointly presented an experts' workshop to tackle issues of *Digitisation, Conservation, and Preservation* in the elegant surroundings of the Palacio Benecazon, in Toledo, Spain.¹ This event brought together communities from Austria, Bulgaria, France, Italy, the Netherlands, Spain, and the United Kingdom to investigate key topics in digital preservation.

The workshop, the first hosted by EPRANET, is part of a suite of six expert meetings that are enabling ERPANET to address the lack of awareness and fragmentation of knowledge and skills amongst the stakeholder communities about how to handle existing digital preservation problems, and how to plan effectively for the future.

These workshops offer an opportunity for practitioners and experts involved in digital preservation activities to collaborate in discussion and in the production of guidelines and recommendations. The Toledo Workshop was designed to promote and develop understanding and new approaches to digitisation, conservation, and preservation.

This report may benefit information and preservation professionals, including commercial and private sector librarians, conservators, digital imaging and audio professionals, content owners, and managers and funders.

The report is organised into five sections. The first four summarise the individual sessions of the workshop:

- Digitisation of Audio and Visual Material
- Digitisation of Images and Manuscripts
- Quality Assurance
- Migration

The final section lays out recommendations and agreements on practice agreed by the participants.

¹ ERPANET is immensely grateful to the Caja Castilla La Mancha for providing the workshop with access to the beautiful Palacio and for their exceptional hospitality.

Aims and Objectives of the Workshop

Over the course of three days, participants examined the issues and challenges arising from the interplay of digitisation, conservation, and preservation. The principle aim of the workshop was to recommend improvements in practices and guidelines, which will enable practitioners, their institutions, and funding agencies to improve conservation of analogue sources and preservation of their digital representations. This report does not examine issues related to born digital materials because the workshop on which it was based did not focus on this topic. Naturally there is some degree of overlap, but issues related to born digital materials will be the subject of future workshops.

The Workshop was organised into four sessions, each dealing with a particular aspect of the topic. Each discussion session began with a provocative scene-setting paper by an expert building mainly from case studies that outlined current practices, obstacles, and emerging opportunities. The participants used these case examples as a basis for broader discussion. Discussion sessions lasted a few hours and involved the comparison of views and opinions as well as the raising of concerns and problems with past, current, and future practices. The sessions aimed to find areas of common understanding of best practices for many different types of materials. The range of institutions and personal experiences that each of the participants brought to the discussion highlighted the range of existing practices and the need for a more common approach.

Session 1: Digitisation of Audio and Visual Materials

Dietrich Schüller, the Director of the Phonogrammarchiv at the Austrian Academy of Sciences, introduced the first working session. His experiences at the world's oldest sound archive provided the participants with a rich overview of the methods of digitisation for preservation, and the relationship of digitisation to conservation and preservation.

The Archive

The Phonogrammarchiv, founded in 1899, is the Austrian research sound archive. Its objective is to produce, collect, accession, preserve, and make available sound recordings as acoustical sources for all research activities. Recently, it has expanded its collecting strategies to include video research footage. The Phonogrammarchiv holds 58,000 recorded objects. Over the last century it has acquired extensive experience in the preservation of audiovisual documents. It is actively involved in the development of long-term strategies for safeguarding audiovisual heritage, and it actively shares its experience with the international community of audiovisual archives.

The Problem

Audiovisual documents are under immediate threat of loss due to the instability of carriers and the obsolescence of format-related equipment and software. The diversity of carriers in archives worldwide – such as records, discs, tapes, CDs, film, and video – itself poses substantial preservation obstacles. The experience of the Phonogrammarchiv demonstrates that the long-term future of audio and video carriers is bleak. Formats are becoming increasingly sophisticated, and their active life-spans shorter and shorter. This rapidity of change has rendered traditional preservation strategies and methods ineffectual.

The 1990's witnessed a shift of paradigm in audio preservation. It became apparent that adhering to the classical concept of preserving the objects placed in the archives' care did not provide a sustainable mechanism. Dietrich Schüller argued that preservation efforts focus now on content rather than the carrier. This content must be preserved by subsequent migration from one carrier to the next. In the audio-visual domain the Phonogrammarchiv has concluded that preservation of this all important content is only viable in the digital domain, therefore all analogue originals must be transferred to the digital domain. Despite this paradigmatic leap, basic principles and procedures of traditional carrier preservation remain intact for the

moment as transfer of the enormous analogue holdings worldwide will take years, if not decades.

This modern preservation strategy embraces a number of basic requirements and steps. Firstly, adequate resolution, the right target format, and an appropriate new carrier must be chosen. Secondly, the content (signal) must be fully extracted from the original object and transferred in accordance with established principles. Dietrich Schüller stressed the essential role played by policies and procedures, and these apply to all preservation strategies, not just the audio example. These must be in place in any digitisation of materials for preservation purposes.

Selection and appraisal form an important part of such policies and procedures. For example, they should facilitate the coordination of selection and appraisal activities with other audio collections to avoid duplication of effort and resources. Priority should be given to objects which are at immediate risk, not likely to be the focus of commercial digitisation, reflect cultural diversity, reflect cultural uniqueness, or are in regular demand. In all cases, the best copy available should be selected for transfer. What constitutes the 'best copy' is open to interpretation; it may be the best sounding copy, the most complete copy, or indeed the copy that will be easiest to transfer. Once the best copy has been identified the equipment for transfer must be selected. Replay equipment must comply with all format specific parameters of the original. In order to produce the best signal, modern equipment should be used in the process to minimize distortions. This may also involve correction of errors created by misaligned recording equipment, such as addressing azimuth errors of magnetic tape recordings.

Finally, a target carrier storage format must be selected for archiving. The carrier should be computer tapes rather than CD-R's (CD Recordable), as the instability of the latter is widely recognised. One possible solution to the carrier problem receiving much attention at the moment is the use of Digital Mass Storage Systems (DMSS). These provide new dimensions of remote access and are likewise ideal installations to manage data checking, regeneration, and future migration with a minimum of human effort. However, there are some overheads to using DMSS. Transfers of content from single carriers to DMSS must reflect realistic time factors in order to enable those undertaking the transfer to create and maintain metadata. The use of DMSS must also be governed by a strategic transfer plan.

Established archival transfer principles need to be employed to guarantee the authenticity and reliability of the digital copy once the selection has taken place and has been successful. The audio document should include both the intended signal from the original object as well as the unintended, including distortions or noise. Both must be transferred accurately.² If after the transfer any signal restoration needs to be undertaken it must only be on a copy of the archival master, not the archival master itself. The central message was that the restoration of analogue materials should take place only in so far as it is necessary to enable the signal to be extracted from the media. For example, crinkled tapes are ironed to allow them to play back. Similarly data reduction, more commonly called 'compression', must only be carried out on copies intended for access or browsing.

The question still remains as to what format to use for the digital master. All digitisation activity should employ recommended standards and practices to allow for adequate quality transfer from analogue to digital, and this is the case for audio too. A key to ensuring long term access is that standards should have widespread market penetration, especially in high content industries. Luckily, the recording industry is better placed than some other fields to provide this. The WAV file (WAVEform sound format) has long been the recommended format for encoding digital files and has now become the *de facto* standard in the audio environment.

Archiving the Content

Once the transfer has been successfully completed³, there still remain several digital archiving principles that must be adhered to and understood at the outset of the project: all digital target carriers must be free of uncorrectable errors and should contain the lowest rate of correctable errors: the error status of each carrier should be recorded and checked regularly; and, refreshment must be carried out when the error rate rises significantly, and certainly before the signal becomes irretrievable. In addition, the contents must be migrated before the system becomes obsolete. Of particular note in the audio field is the extremely time-consuming effort of transfer. With well-preserved originals, it takes approximately three times the object's playing length to transfer to a digital platform. This must be taken into account at the beginning of any digitisation activity.

² Sampling rates and amplitudes resolutions are all important considerations here. Presently, forty-eight kHz / twenty-four bit is widely employed in radio archives, while memory institutions are generally digitising at ninety-six kHz / twenty-four bit.

Again, Digital Mass Storage Systems (DMSS) seem to provide a solution to these principles. They include automated data integrity checking, automated refreshing, and migration facilities among their tools. Mass Storage Systems are very efficient at offering new dimensions of access to sound archives while simultaneously solving the problems of checking the data integrity of the digitally stored contents. However, they can be expensive and there are alternative solutions available if funds and size of collection are not conducive to their use. Manual checking, refreshing and migration is always an option for smaller collections, but can be labour-intensive, and more costly than originally envisaged.

For increasing the chances of complete preservation metadata must of course be kept. It can either be integrated into the audio files themselves as in the case of BWF files (European Broadcasting Union Broadcast Wave Format), or maintained separately. It is necessary, however, to carefully assess and permanently maintain technical data related to originals, transfer, and the new storage environment.

Theoretically, audio principles are also applicable to video, but they are not yet blessed with such a *de facto* standard as the audio field. Additionally, one hour of uncompressed video requires 122GB of data storage. Even with the rate at which storage rates are declining, storage requirements on this scale are financially prohibitive. It is recommended that preservers should wait as long as possible before making any immediate changes, as the soon-to-arrive MXF format (Media Exchange Format) will provide new technical opportunities. It is expected that this will bring with it a decrease in storage costs.

Digitisation of film for preservation is not yet feasible; the storage capacity and bandwidth required are outside the financial constraints of public sector institutions. There are digitisation options for access, however, including DVD (Digital Video Disc) and MPEG2 (Motion Pictures Expert Group) format compression at various data rates which can be employed according to the expected quality.

Digitisation is already widely employed for accessing audiovisual objects. It is now recognised as an indispensable tool for the long-term preservation of audio and video. Without transferring these analogue materials into digital form there will be no way to access them in the years to come. Film preservation will have to rely on conventional strategies to preserve the carriers in the short term, but even here technical developments will, in the longer term, enable similar strategies to be applied to film.

Discussion

The discussion following Dietrich Schüller's presentation revolved around the relationship between digitisation, conservation, and preservation. Schüller demonstrated that in the audio world preservation and digitisation are not distinguished as separate activities. The digital copy is treated as the new original, and the analogue original was frequently destroyed in the process. Understandably this provoked a strong response from other experts, some of whom felt that some distinctions between the treatment of different types of materials had to be elucidated.

Some of the main reasons for digitising in different fields were discussed, and this became a thread which was revisited frequently during the course of the workshop. The increase of access to the object, better exploitation of the object at no risk to the original, and virtual restoration (where technology can be employed to allow the digital object to be used in a manner not feasible with the original format) were all cited as reasons. However, the presentation by Schüller forced an exploration of digitisation as a preservation tool. There is no doubt in the audiovisual environment that without digitisation there would be little hope of most of Europe's aural and visual heritage surviving and remaining accessible. One key element that stuck with the experts was the degree to which funding and time spent on digitisation had become a bargaining tool in the argument that digital copies should be used to address preservation purposes as well.

A second strand of the discussion looked at the practical care of the original analogue object. Conservation activity in recent times has fallen out of favour with funding bodies; they tend to put the greatest emphasis on activities that promote access such as digitisation. Funding for digitisation has become a detriment to funding for conservation. The Toledo experts agreed:

- that much more has to be done in order to increase the resources directed at conservation activities, and that the most practical way of doing this is to include conservation as part of any digitisation project proposal;
- that curatorial responsibilities needed to be reasserted by those responsible for our visual and audio heritage;
- that digital copies do not replace the analogue original and they do not absolve curators of their responsibilities in relation to the original.

At the end of this opening session, participants were given some time to explore the surroundings of the Palacio, and to talk informally.

Session 2: Digitisation of Images and Manuscripts

The size of collections subject to digitisation changes the scope of the work and poses unique challenges. Session Two looked at large quantity of materials and the techniques used to digitise not only efficiently but also to attain a high level of quality. Dr. Hans Petschar, the Director of the *Porträtsammlung, Bildarchiv und Fideikommissbibliothek* at the National Library of Austria introduced this session by describing his experiences digitising over two million photographs for online access. His presentation examined issues related to automation and quality levels.

Key to the success of the *Porträtsammlung, Bildarchiv und Fideikommissbibliothek* project, was the six month planning period at the outset. This time was spent analysing various sources and information before any digitisation work commenced. The result of this planning was an understanding of the complexity of the task, a clear definition of the main goals and objectives, and a realistic *modus operandi* to achieve them. In the Library's case, distinctions were made between different types of users as well as drawing a clear distinction between commercial and educational requirements.

One of the major problems facing any photographic digitisation project is the limited number of experts. In order to fill this void in expertise, it is essential that staff from sections with a vested interest are brought together (conservators, librarians, archivists, information technology staff, and curators). If this collaboration and integration is not in place, contextual information is usually lost. Most of the relevant metadata can be found on, near, or beside the object, but in uncoordinated digitisation work these resources are often overlooked. Petschar's timely example of their digitisation of card catalogues displayed how this type of collaboration ensured that an important historical contextual source was utilised.

The Process

Software tools developed by the National Library and its partners enabled the library to scan 60,000 catalogue cards per day during the project. It had been found in the planning stages that no existing software company offered a product that met the requirements, and so it was developed in conjunction with a private company. The digitisation of the cards took a total of two weeks to complete. Content was converted into a Dublin Core compliant XML (Extensible Mark-up Language) database.

The metadata collected in this process related to the 40,000 digitised photographs. The aim of the digitisation of the photographs was not to create archival quality copies, but rather to produce digital versions for dissemination, study and exploitation. This decision obviously had a great impact on the approach the project took to scanning and the throughput the project was able to achieve. The project adopted an approach to digitisation that was novel. For the process, an external contractor was used, and they scanned between 2000 and 2500 objects per day, producing three versions or more of both the front and the back of each photograph. For the first time it was possible to produce a 'perfect working copy' of an image for 0.5 cents (€). Under existing standards for preservation and conservation this technique falls short. The National Library made a conscious choice to focus on access rather than preservation. Perhaps this solution is not ideal, but the Library did not find itself in an ideal situation to create archival quality copies. Of importance to the Library was the throughput they achieved, rather than definitive quality: they generated 40,000 digital images in one month rather than one year.

The National Library's situation emphasises that the process of digitisation must be redefined to reflect the goals of the initiative. If the goal is to create archival quality images that comply with existing standards, then the project would have needed to be done by hand.

Hans Petschar felt that automation provided the only solution to the problem of digitising large amounts of material. For him time and money necessitate that digitisation by hand will never be a viable option. Automation will enable institutions to digitise extensive holdings at a reasonable cost and within a viable time frame. The automatic process also cuts down on the likelihood of damage and error. Any handling or human involvement, however minor, increases the rate of degradation. Put simply, Petschar believes that machines can do a better job than humans in handling sensitive materials. To highlight this, he pointed to the medical and industrial world, where machines handle objects in a manner that humans are not capable of doing.

This example from the National Library highlights the issues raised by digitising a large body of material. The route that the Library chose does not address the concerns of archives standards and quality. However, digitisation is being used here as a way of providing access to a large collection, and in automation, the National Library of Austria have found a solution to the huge costs and long timescales that delay other projects. Crucial conclusions for other projects include:

- careful planning is essential. Projects can no longer justify the approach of buying a scanner and scanning. They must develop a detailed strategy and tailor their processes to the goals that the project aims to achieve
- the need to reduce the manual handling of materials to reduce risk to the originals and to improve throughout
- that the experience and knowledge held by current staff who work with the collections being digitised should be tapped.

Other paper

Julia Craig-McFeely's presentation on the *Digital Image Archive of Medieval Music (DIAMM)* developed the issues related to access further. She offered an insight into the digitisation of extremely fragile manuscripts for access via the Internet. In complete contrast with Hans Petschar, the primary concern was on generating as high a quality image as possible, although this was not specifically for archival purposes, but rather to enable a virtual restoration of manuscripts through manipulation of the digital image, and to enable scholars to use new computer based technologies to discover information in the manuscripts that has in the past gone unnoticed.

DIAMM is concerned with the re-discovery of lost fragments of musical annotation. Throughout history it was commonplace to scrape text off manuscripts and reuse the vellum parchment for another purpose. So for example, a manuscript that had music written on it might be scraped clean and then re-purposed (for accounting, a journal entry, etc). Usually the scraping and cleaning did not remove all evidence of the earlier text from the page, although the original text was often difficult to read in any thing more than a fragmentary way. These palimpsests often contain texts and musical notations that is otherwise lost. In establishing this digital collection, DIAMM is aiming to preserve a group of fragmentary sources that are in danger of loss, through not only physical deterioration but also through ignorance of their contents.

The main aim, once manuscripts have been identified, is to produce a printed copy of the original manuscript at the highest possible quality and at real size. This determined the minimum dots per inch (dpi) for capture, resulting in an extremely large digital copy.⁴ Because storage issues are not at present an issue for DIAMM,

⁴ Each image is about 250MB in size.

they have not needed to compromise the quality of the images. All images are captured and stored in uncompressed TIFF (Tagged Image File Format) format, with capture metadata inserted into the image tag. Fuller metadata that includes bibliographical and research information about the manuscript is stored in a database, archived with the images, and linked to the original images via unique filenames and recorded drive-paths for all versions (including the archive, printable resolution, and web-browser sizes).

Originally, accessibility to the archived images was not a primary concern, but as the Internet has developed as a research resource during the four years of the project's existence, it has become clear that accessibility is a highly desirable additional service. Libraries, as well as granting DIAMM permission to create high resolution copies for research purposes, also grant the project a non-exclusive licence allowing DIAMM to place reproductions at low resolution of items in the collection on the DIAMM website. The images on the website are print- and copy-disabled in PDF format (Portable Document Format), but have some support for zoom.

The project uses these very high-resolution images to achieve a form of virtual restoration, enabling the recovery of lost information and the discovery of hitherto unavailable information. DIAMM found that only the highest resolution images provided sufficient colour information and resolution to enable researchers to conduct useful restoration work. The original raw image is preserved for future researchers so that they can test the conclusions drawn by contemporary scholars.

The Process

DIAMM brings its mobile high-resolution digitisation studio to the documents rather than requiring them to be brought to a fixed-site studio. Following imaging, DIAMM gives the contributing institution a digital copy of the high resolution image. The project adopted the mobile imaging studio approach because it found that this is the only way it could achieve consistency in image quality and guarantee that the digital images were created from the original manuscript themselves.

Images are brought back to Oxford and archived on the Oxford University Computing Services (OUCS) Hierarchical File Server. Files are kept active by regular downloads and checks - the entire archive is on a rolling download system, so each file should be downloaded and opened to check for file compatibility and readability at least once every year. If the file passes these tests it is returned to storage without changes having been made. Should an update to the format become necessary, this

would be managed through the rolling check system. At present this system is not automated, but it is hoped that an automation procedure will be put in place that will work through the images and only inform the administrator if it encounters a problem. At the end of the project, all data will be handed over to the Arts and Humanities Data Service (AHDS) for long term preservation.

These two introductions stimulated discussion about quality, authentic copies, storage formats, the role of restoration, practical costs, and techniques.

The relationship between the original and the digital copy was identified as a complex issue. There was an attempt to better understand the relationship between the two and improve the treatment of both the copy and original as well as aiding the interpretation through use of the digital representation. The relationship between the original and the digital interpretation is one that must be maintained, but there should be a degree of separation between the originals and the digital representations. It is important to remember that the digital version is a representation of the analogue object at a particular point in time. It is not a copy of the original, but a representation of the original on the day the digitisation was made. The digital representation takes on its own lifecycle and becomes a separate object as changes are made through additional uses or migrations. The digital object is not an object in the traditional understanding, but remains a series of reconstructed bits. There is, therefore, a range of interpretations accorded to the digital interpretation. The experts argued that there should be a clearer understanding of the nature of both the original and the representation.

The costs of digitisation and preservation are an important issue for all stakeholders and play a major role in determining what digitisation work can be done. Often a compromise must be reached between what is desirable and what is economically feasible. Digitisation will not bring down costs, but it will open up access to the materials. It is still difficult to put figures on digitisation and preservation costs because of the range of options currently available for preservers, but all our evidence shows that the cost of digitisation is overestimated and the cost of preservation is underestimated. The focus lies on the technology aspects of the costs instead of considering the costs of the intellectual environment itself. This problem is exacerbated by the fact that institutions tend to choose the cheapest option. There needs to be a greater understanding of where the costs of digitisation and preservation arise. The most expensive aspect is not equipment, but human resources. Human resource issues become especially important in considering the

value of quality processes. Costs of transferring digitised objects to a repository are substantial, and this tends to be overlooked in the initial estimate or grant application; a distinction should be made between initial transfer costs and preservation costs.

Automation is a way of keeping down costs. It is faster, for example, to digitise images and preserve them at one resolution rather than at two, where one is used for preservation purposes and the other for access. The less human intervention, the less there is in cost. Many of the transfer and long term maintenance costs of digital entities can be reduced if processes are automated.

The discussion moved onto a consideration of another practical problem, storage formats. As Julia Craig-McFeely explained, decisions on file formats and file size impact preservation strategies. There is a great deal of misunderstanding and confusion around the adoption and acceptance of available formats. The Toledo experts expressed concerns that preservers are held to ransom by file formats, and have less control than is desirable over the development of preservation storage formats. Instead of dictating their preservation needs and demands, many feel that they are dictated to by the market (e.g. market dominance of TIFF over PNG⁵). Issues of object resolution become problematic when storage is considered. Preservers need to look beyond CD-R as a storage medium as the risks associated with it (e.g. variability in quality, unpredictable life expectancy, portability) are too extensive. Without metrics for measuring the preservation quality of CD-Rs they shouldn't be adopted for archival storage. DLT Tape (Digital Linear Tape) is still one of the most secure formats. Storage requirements should be built into the archival system at the beginning of any digitisation project, and not left for future organisations or institutions. The storage system should be part of the overall management system in place in order to guarantee the long-term preservation of authentic and reliable digital objects. It is important to remember that all storage technologies are at best interim measures.

⁵ Portable Network Graphic

Session 3: Quality Assurance

This session focused on the roles of quality processes and standards in digitisation and preservation. Victoria Fenner from *Images of England* examined quality assurance methods employed to maintain high standards throughout this project's digitisation, presentation, and preservation processes.

Images of England (IoE) has received funding of £4.4 million (€6.7 million) from English Heritage, Heritage Lottery Fund, and the Royal Photographic Society. Its main purpose is to create a digital image library of 370,000 images to be accessible via the Internet, free of charge. This will be a 'point in time' record preserved for future generations of the listed buildings and structures in England, and will make heritage information more widely available. This project provides an example of how planning and quality controls can assure the production of digitised resources of lasting value that can be guaranteed both authentic and reliable.

There are 370,000 listed buildings in England designated for architectural or historic interest, historical associations, group value, and age and rarity. It was felt that this register of buildings would be a more useable and interesting resource if it were supplemented by images. As of June 2002 over 1100 volunteers are taking photographs for the project using 35mm colour negative film. The IoE survey team provides the volunteers with film, a list of targets, and supporting documents. All volunteers are briefed on the photographic standards. The volunteers are invaluable for both their photographic skills and local knowledge.

Standards

The quality of Images of England images and their digitisation rests on standards, guidelines, and audits. A set of guidelines was developed by the project team to ensure that each image meets established criteria. These guidelines standardise the 'defining image' that the volunteers are to take; this image should show architectural character, indicate function, show something of a building's context, and be truthful. The project maintains the relationship between the number of the listed building record, the film number and the frame number throughout the lifecycle of the image. This is essential in a project that needs to track 370,000 targets, 15,000 films, and over 1100 volunteers. As the processing, printing and scanning work is contracted out, IoE have implemented a number of safety and quality assurance procedures. For example, IoE has developed contractual specifications covering such factors as

resolution, colour depth, dynamic range, and metadata to ensure that the scans and associated data are of a high enough quality.

A number of digital images are created from the scans. A master image file stored as an uncompressed TIFF provides the archive copy. In addition, thumbnail, low, medium and high resolution JPEGs are created. Only the thumbnail and low-resolution image are available freely on the website. Images are checked for scan quality, photographic quality, and to establish that the right building or structure has been photographed. A numerical score is awarded to each image. Photographic quality relates to how well an image meets the defining image criteria (see above). A grade is determined by the presence and severity of such defects as obscured and cropped structures, framing problems, poor composition, and lighting difficulties.

This project aims to be self-sufficient when funding runs out. A subset of the images (about 10,000) will be available for sale, and there is a possibility that IoE will also offer other derived products. The next generation of the website will incorporate an e-commerce facility. This idea has been part of the original planning of IoE. The project team recognised that popular images make money and that the viability of the project depended on income from sales.

Victoria Fenner's introduction raised several key issues for discussion, and revisited others. Quality assurance as a concept has little to do with the actual contents of a digital entity. Fenner's approach is designed to guarantee quality of the content. There are many problems associated with defining quality, especially in the heritage and cultural environments where there are many different types of objects and users. In addition, any quality criteria are likely to change over time. Participants discussed whether or not there was a basic assumption of quality that could be applied to all formats and forms of preservation, and also be relevant for all types of users. One of the only reasonable ways to address this issue is through the use of standards, as these provide a way to achieve consistency of quality. The development and implementation of standards must not occur in isolation. Communication across projects and digitisation activities to understand choices and decisions made and share experiences provides an essential framework.

Increasingly, there are more and more standards in existence because of the increasing range in types of documents and objects. In choosing standards to apply, the goals of the project must be examined as well as the different nature of the various standards available (e.g. *de facto* or proprietary). Standards, when used,

provide an agreed upon level of quality. Industry standards are currently inadequate for most forms of digital preservation. There are a few instances where an industry standard has become a *de facto* standard because of the significant amount of use it receives. The WAV format is almost exclusively used for audio files that are targeted for digital preservation, as is TIFF for image formats. The fact that there are now millions of hours of audio material in the WAV format only adds to its potential longevity. Other industry standards are not so stable however. Even a relatively common standard like PDF has problems. For example, current PDF software cannot read the PDF files produced with first generation PDF writers. There are also many forms of JPEG in existence. The high usage of Windows software has made this an almost *de facto* standard as well, but there is little reassurance for preservation concerns due to its proprietary nature. In recent years there has been a shift towards open source standards and a positive move away from proprietary systems.

Discussion turned to the position and influence that the cultural and heritage sectors have in bringing about a change in digital preservation standards and practices. Experts agreed that industry standards do not meet preservation requirements. The participants considered the likelihood that the heritage and cultural sector would never have enough clout to set standards. When compared to other professional communities such as the medical and astronomy communities, the heritage and cultural sector falls a long way short of the influence that can be wielded by certain groups over manufacturers and markets. These groups have more power to set standards and demand equipment and software to meet their needs. In addition, large organisations can deal directly with suppliers, but this course of action is not open to cultural or heritage institutions or even collaborative projects. Vendors and producers need to be made more aware of demands for products that have preservation functionality. The participants pointed out that there needed to be agreement first in the sector as to preservation requirements before demands could be put to industry. There was also the suggestion that bodies like IFLA (International Federation of Library Associations) or the ICA (International Council on Archives) should take this issue on, and use their substantial influence to raise awareness of the issue both inside the sector and externally.

Session 4: Migration

The final workshop session dealt with the issue of migration, and addressed some of the long-term digital preservation concerns raised by this approach. Julián Bescós, Manager of the Archives Group at *Informática El Corte Inglés* introduced this topic with an overview of the relationship between migration and digitisation. His presentation outlined the necessary components of a digital preservation system. His experience and unique case study examples provided the participants with an excellent basis for wider discussion of technology, best practices, tools, and techniques.

The advantages of transferring analogue materials to digital form include a relief from the traditional conflict between preservation and access, the ability to operate on and manipulate images of the original object, remote access, and effective additional methods of access such as extended search capabilities. However, there are also drawbacks to digitisation, including the need to ensure long term access to the digital entities created by a project.

The need for migration as a preservation strategy arose from information becoming inaccessible as hardware, software, and storage media continually evolved. As a result of technology obsolescence, new functionalities of software, storage concerns, and the fact that the technology world is still relatively young and immature, the need, when it arises, for migration of digital information is acute.

The preservation of digital assets is greatly influenced by decisions taken at the time the digital entities are created. Digitisation equipment, storage facilities, access and consultation facilities and diffusion activities make up this digital information system. Julián Bescós described a project that used either flatbed scanners or zenithal scanners for digitisation. The storage requirements were met with magnetic discs for online image service, indexing and search facilities for navigation and access (in own media and/or in a relational database), and massive optical media (CD-R and DVD) for off line access and backup. Access and consultation requires the functions for navigating, searching, and display of images (such as zoom and scroll), as well as image processing functions (such as stain removal and contrast). Diffusion required Internet facilities in addition to CD-ROM (Read Only CDs) and DVD.

Best Practices for a Digitisation Project

In general, equipment should be based on PCs and mainstream commercial equipment. Hardware should be provided by prominent IT companies. The system

should be modular to allow for progressive installation and the easy update of system elements. The system should be built with widely used operating systems, with applications including database management tools. Official or *de facto* standard file formats should be employed such as TIFF, JPEG, and XML.

The digitisation process should be undertaken with the understanding that the digital images produced should allow for a broad range of future uses. Where compression formats are necessary, standard methods should be adopted (e.g. LZW⁶). The highest quality image technically possible and economically feasible for large-scale production should be captured. While both the informational content and physical appearance can be captured, the capture of the physical appearance can involve a higher cost. Selection of documents or objects to be digitised should be based on an assessment of value, condition, expected use and the availability of access aids.

The storage media should be one in wide use and available at a low cost. Magnetic disks should be used for online image services. For backup and recovery, industry standard procedures should be used. There should be periodic backup to magnetic tape (e.g. DLT) with one copy held onsite for rapid recovery, and one copy held offsite for disaster recovery. There is little experience at present of storing massive quantities of culturally valuable materials and these are the best practices currently available.

A key contribution that Julián Bescós and his team have made is in the area of migration. Migration is a term that is used to encompass a variety of different processes. The process can involve migration of media, migration of the technology platform, and migration of the digitising and access software. Migration of media might include refreshing digital information by copying it from one type of media to another, or converting files from their current formats into another to be interpreted by new programmes, as well as decompression and compression. Migration of the technology platform involves changing servers and PCs, peripheral devices as well as the operating system and database manager. Migration of the digitising and access software involves the maintenance of the software in the new platform as well as development of new migration software. These changes may require further migration of formats.

Planning is a key component of any successful project, but poses problems in the case of migration. Planning is difficult due to the limited experience in this area and

⁶ Lempel-Ziv-Welch, a lossless compression method

lack of ability to predict accurately when media, software and hardware will become obsolete. In addition, no single strategy applies to all forms of digital information. There is a great variety in the different application environments, the different forms of digital information, and preserving different degrees of computation, display and retrieval. Preservers need new solutions for each new format and process they encounter. Currently these factors make automatic conversion only partly possible.

It is important to be aware of the changes occurring, and to treat these with some urgency.

- New releases of software and databases can be expected every two to three years, with minor updates more frequently.
- Migration from one storage media to another can be expected every four to five years.
- Migration to new hardware and software occurs less frequently, but can be expected every five to ten years.

Archivo General des Indias, Sevilla

The practical experiences of the *Archivo General des Indias, Sevilla* provide a very useful example of digital preservation and migration techniques. This digital archive system has been in service since 1992, and has undergone one migration of hardware platform and two migrations of media in the time since. The main aim of the project was to create a computerised system for the *Archivo General des Indias* comprised of 86 million pages of original manuscripts relating to the Spanish Administration in America (the fifteenth to the nineteenth centuries), with an additional 25,000 pages of inventories and catalogues. The information system that was created comprised a textual database containing all the descriptive information, a digital image archive of eleven million pages, and a user management module that would control access and use statistics. By 1995 eleven million digital images had been created along with 400,000 descriptive entries. The digital image archive contained images scanned at a resolution of 100dpi and sixteen grey levels. These images underwent Differential Pulse Code Modulation compression, and each image was stored in files 300-350Kbytes each. These images were stored on *Plasmon* optical disks and backed up on DAT magnetic tapes and CD-R. When accessed by the user at one of the Archive's forty workstations these images could be manipulated and 'restored' by applying algorithms to remove blotches, compensation for ink bleeding, and enhance the contrast to make the images more legible.

The system was migrated in 1999. The operating system was shifted from OS/2 to Windows NT and Oracle, and the application services were moved to ArchiDOC and ArchiGÉS, developed by *Informática El Corte Inglés*. Then, the images stored on 10,600 optical disks were migrated to CD-R and then onto magnetic disks in 2000. 6000 CD-Rs were used with backups copies. Images were decompressed from the DPCM format and then compressed into JPEG format before verification and copying onto CD-R. This process took three operators four months. The migration from CD-R to magnetic disk in 2000 involved only copying, as the images were already in the desired format (JPEG). They used 10 RAID (Redundant Array of Inexpensive Disks) cabinets of 350Gbytes each. This process took one operator a month. At present, the cost of the magnetic disk is lower than in 2000, and so 10 million images with a size of 3,520 Gbytes would now cost a total of €20,000.

With this experience, Bescos outlined an example for a digitisation system, including costs and time resources required. Based on a project involving description, scanning and providing access to 10,000 files and 500,000 images, the costs and equipment are outlined below.

COSTS	2 YEARS WORK	MAINTENANCE
Personnel		
2 Archive Experts	€120,000	
2 Scanner Operators	€ 66,000	
Computer Operator (part time)	€ 12,000	
Hardware		
Server with storage for images (160GB redundant magnetic storage)	€ 12,000	
Backup Elements	€ 1,800	€ 600
1 Flatbed scanner	€ 3,000	€ 300
1 Zenithal scanner	€ 24,000	€ 1,800
4 Workstations	€ 7,200	
Basic Software		
Windows	€ 1,800	
Oracle 8i	€ 6,000	€ 1,200
Application software		
Software, installation training, Assistance, management	€ 40,000	€12,000
TOTAL	€293,000	€15,900

© Informática El Corte Inglés

A digital preservation strategy should address:

- the use of mainstream commercial equipment;
- use of standard formats,
- storage on magnetic disk with redundancy, backup policies, and maintenance; and
- support periodic update policy covering hardware, media, operating system, and application software.

The experts discussed the meaning of the term migration, and concluded that it could be understood on three levels. The first or most basic level of migration was the change of media used to store the digital objects. The second level of migration involved a modification or some change of the system and verification that the object remained accessible within the new environment. The third level of migration entailed a change in the format of the object, and would commonly also include one of the other two steps as well.

As migration has been relatively rare in the heritage and cultural sectors, especially for the type and volume of images that Julián Bescós was dealing with, as well as the frequent nature of change in technologies, it becomes very difficult to strategically plan for this activity. There is no clear and definitive method or timetable for migration. Migration has to be different for different media; for example there are different speeds of change and deterioration for media types. The participants did agree, however, that it is best to keep the rate of migration as low as possible, but at the same time ensure that nothing is put at risk and digital objects remain accessible at all times.

The participants also discussed the importance of including some sort of migration expectation into all digitisation initiatives at the outset, so that creators of digital objects understood the lengthy commitment necessary to retain access to these resources for the indefinite future. All organisations should develop a migration policy which should involve a close watch on technology and format changes and developments. Creators and organisations should be prepared to migrate every five to ten years in the way that consumers have become prepared (or conditioned) to change cars or fridges and other products.

Recommendations and Preservation Strategies

Underlying the discussion throughout the workshop was the desire to elucidate some useful digitisation strategies that would enable practitioners, funders, and preservation organisations to appreciate the range of concerns and issues that need to be taken into account when undertaking any digitisation activity.

Digitisation and Conservation

The main reasons for digitisation, as agreed by the Workshop participants are:

- better access;
- exploitation;
- virtual restoration; and
- preservation.

Despite the range of types of materials which are digitised, these remain the four basic justifications. Conservation should be an integral part of digitisation projects, and such activities should necessarily be part of the overall scope of the project in order to counter the diversion of funding, which the experts had observed from traditional collection curation to digitisation.

Treatment of Originals

A relationship should be maintained between the original object and the digital surrogate, but there needs to be an understanding of the separate nature of the two. Even if digital preservation copies are made, it is important that the curatorial responsibilities for the analogue original are maintained.

Standards

Digital preservation standards and best practices should be developed in concert with other activities and initiatives. It is important that there is communication about techniques, choices, and results of other digital preservation activities so that more consistent and permanent solutions can be found to address the problems of digital preservation. The success and widespread adoption of the .wav format for audio files demonstrates that it is feasible to identify the attributes of objects that need to be represented and to ensure that these are represented in the digital file formats.

Quality

The use of policies, procedures, and standards provides a mechanism to support quality of content. It is important that there is communication across projects and digitisation activities so that experiences can be shared. It is also useful to

understand what choices have been made by other digitisation projects and why and how this has affected the quality of their results. High standards need to be maintained by the existence of quality processes, guidelines, safety mechanisms and audits.

Planning

Planning of any digitisation project is essential to ensure that decisions are made from the outset that reflect the needs of the users as well as the abilities and resources of the creators and organisation carrying out the digitisation and preservation activities. Issues of quality, standards, costs, storage formats, and media need to be addressed in the very early stages, and be clearly understood.

Storage

All the participants agreed that preservers need to look beyond the CD-R format as a long-term storage media as the risks associated with it are considered to be too great. It remains an unknown quantity with a lack of suitable information available as to its suitability, and as such puts the materials stored on it at risk. Life expectancy of this medium is estimated to be between three and three hundred years. DLT tape remains one of the best formats for system restoration preservation.

Storage requirements should be built into the archival system at the beginning of any digitisation project by those involved, and not left for another organisation or institution to inherit and preserve without sufficient information. The storage system should be part of the overall management system in place in order to guarantee the long-term preservation of authentic and reliable digital objects. It is important to remember that all storage technologies are at best interim measures.

Migration

Participants agreed that any digitisation project should not be commenced unless there is funding or resources in place for migration, or the ability to provide for this activity over time. Migration is still quite rare in the heritage and cultural environment in the volume discussed during some of the examples provided by participants. This makes it quite difficult to plan. It is important therefore, that the rate of migration is kept as low as possible, and migration to another platform should only be undertaken when there is danger of loss. Action must be a median between necessity and cost. In addition to this, it is important to use and maintain common software (both basic and application) to keep costs down.

Organisations should have migration policies. They must have an awareness that they will need to update their information in the same way that most consumer goods need replacing. It is also vital to be aware that any current storage solution will only have a limited life span.

Position of the Heritage and Cultural Sector

The sector should ally itself with larger, sympathetic organisations that can improve the position and influence of the cultural and heritage organisations. The inclusion of translation in an increasing number of software products seems to indicate a growing awareness of the problem among manufacturers of software products. Unfortunately, there is still a great deal of dependence on the vendors, and there is little control over how changes will affect the digital objects. However, there needs to be agreement within the community (which is still lacking in certain areas) before industry can be lobbied. IFLA and the ICA and similar large bodies should take this issue on board and use it to raise awareness of the problem. Initiatives like OAIS (Open Archival Information System) will mean that there will be less dependence on software vendors and preservers should welcome this.

Costs of Digitisation, Preservation, and Conservation

There needs to be a greater understanding of where the costs of digitisation and preservation come from. The most expensive aspect comes not from equipment, as is commonly understood, but from human resources. This is a significant cost when the importance of creating quality processes is considered. In addition, the cost of transferring the digitised objects to a repository is another huge cost, and this should also be considered in the initial estimate or grant application. A distinction should be made between initial transfer costs and preservation costs.

There are strategies that can be employed to keep costs lower. The use of automation is an example of this. Some of the human resource costs can be reduced. In addition, it is faster to digitise images and preserve them at one resolution rather than at two, where one is used for preservation purposes and the other for access. Costs must be appreciated in terms of outcomes. Depending on the type of digitised object that is to be produced, less or more resources will need to be expended. This is why an understanding of the full range of activities of digitisation from creation to access, as well as users and to long-term storage must be understood at the outset. Slowing down the changes in technology can also reduce costs. Open source software may be a part of this solution to the problem of lowering digitisation costs.

Digitisation Strategies

There should be an overall management strategy in place before commencing any digitisation project. This should include quality processing (including evaluation and maintenance), support strategies, a technology watch, as well as backup and migration policies, and a storage solution. The digitisation activity should also involve prioritisation, and where possible include financial imperatives for digitisation. Where there is financial imperative there is usually a greater likelihood of the objects being preserved in a retrievable and useable manner. Rules should be established at the outset about what treatments and activities are acceptable. Traceability and audit should be built into the system. The Images of England model presents a viable preservation model due to its self-sustaining nature. Images of England has aimed since the outset to keep this resource active by ensuring the images are popular ones which are more capable of generating money because of their long-term appeal. This idea has shaped the whole digitisation process. If the objects are still active and used, they have a much greater chance of survival.

Digitisation and preservation activities should stay as much as possible within formal recognised standards. It is wise to reduce the numbers of formats used. The choices that are made will not be the same ten years from now and it is important to remember that this will not be the final solution.

Preservers must follow the changes in the hardware and software markets. Preservers want to do as little as possible, but at the same time do not want to endanger their digitised materials, and want to be able to access them still. Action therefore needs to be a median between necessity and cost. Good management strategies and clearer vision will also help to reduce costs.

Conclusions (Role of ERPANET)

This workshop and resort has highlighted many of the most pressing concerns of practitioners and preserving organisations. The range of examples and techniques provided by the introductory speakers demonstrated not only some of the challenges and difficulties associated with digitisation, preservation, and conservation, but also brought a range of solutions and techniques to the attention of participants. By examining these different techniques and approaches together, practitioners receive a wide range of potential strategies to employ. While there may not be direct crossovers, the approaches taken, for example, in the audio world provide stimulus and new considerations for those dealing with images or textual objects. Because of the range of solutions available, it is important that practitioners and organisations come together to pool knowledge and strategies and disseminate this information to a greater audience actively involved with the full range of preservation concerns.

There is no doubt that there are still a lot of digitisation and preservation issues to be resolved, and that any object in digital form needs a lot of care. There are also undoubtedly huge management issues involved. It is important that digital information is exploited. Working more closely with other agencies to address some of the issues that seem to be beyond the control of the heritage and cultural sectors is an important step. The cultural and heritage communities have to work outside of its own environment, and this is where ERPANET will play a major role. ERPANET has begun to, and will continue to facilitate the exchange of information and sharing of resources. It is important to have authoritative sources of information that can be applied to a variety of circumstances. Using this forum, information and preservation professionals can extend into their own communities and others, and bring in experts from other fields. It will be important not only to share positive experiences, but also to be aware of the failures and problems that can arise. With this knowledge, digital preservation practices can be improved.

Annexe 1: List of Participants

Aurora Diez Baños	Biblioteca Histórica "Marqués de Valdecilla",
Christine Petillat	French National Archives
David Gómez Lozano	Upper School of Conservation & Restoration, Madrid
Dietrich Schüller	Phonogrammarchiv, Austrian Academy of Sciences
Elena Ruiz Ferrandiz	Fundación Félix Granda
Hans Hofman	National Archives of Netherlands
Hans Petscher	Porträtsammlung, Bildarchiv und Fideikommissbibliothek
Jean-Pierre Teil	French National Archives
Jorge Sebastián Lozano	Departamento de Historia del Arte. Universitat de Valencia
José Luis Esteban	Biblioteca Nacional
Jose Manuel Gonzalez	Ministry of Teaching, Culture and Sports
Julia Craig-McFeely	Digital Image Archive of Medieval Music
Julián Bescós	Informática El Corte Inglés
Maria Jesus Diaz-Albo Chaparro	Universidad Autonoma de Madrid
Marie-Therese Varlamoff	Bibliothèque nationale de France
Max Kaiser	Porträtsammlung, Bildarchiv und Fideikommissbibliothek
Maya Ionova	Universidad Complutense de Madrid
Monica Greenan	British Content Editor, ERPANET
M ^a Montserrat Soca Garcia	Archivo Regional Comunidad de Madrid
Peter McKinney	Coordinator, ERPANET
Seamus Ross	Director, ERPANET
Silvia Pajares Montolío	Fundación Félix Granda
Vikki Fenner	Images of England
Vincenzo De Meo	Italian Content Editor, ERPANET