GUIDELINES FOR DIGITIZATION PROJECTS
for collections and holdings in the public domain, particularly those held by libraries and archives

March 2002

These Guidelines are the result of a joint venture of a group of experts on behalf of IFLA and ICA (International Council on Archives), who had been invited to draft these for UNESCO.

This is a draft. The final version will be published by UNESCO in due course.

They are aimed at decision makers, library and archive managers, and curatorial and technical staff members, particularly those in institutions in developing countries.

The members of the working group were:

- John McIlwaine (IFLA, Chairman)
- Jean-Marc Comment (ICA)
- Clemens de Wolf (IFLA)
- Dale Peters (IFLA)
- Borje Justrell (ICA)
- John McIlwaine (IFLA)
- Marie-Thérèse Varlamoff (IFLA)
- Sjoerd Koopman (IFLA, Secretary)

Preservation and Conservation Section: http://www.ifla.org/VII/s19/sconsv.htm

GUIDELINES FOR DIGITIZATION PROJECTS for collections and holdings in the public domain, particularly those held by libraries and archives

Table of Contents

PREFACE

These Guidelines have been produced by a working group representing IFLA and the ICA that was commissioned by UNESCO to establish guidelines for digitization programmes for collections and holdings in the public domain. The contract specified that the guidelines should so far as possible be particularly applicable to institutions in the countries of the developing world. Members of the group were nominated by IFLA and ICA and their activities were coordinated by Sjoerd Koopman, Coordinator of Professional Activities for IFLA.

The group was aware from the beginning that there exist already many publications and websites offering information and advice in the area of digitization. These have been produced by public and private organizations or co-operatives, by libraries, archives, standards organizations, commercial manufacturers and others. Some are general in scope, others are basically a record of the decisions made and programmes followed by a single institution. Few have emerged from the countries of the developing world, or give much attention to the particular issues of such countries.

The rationale followed by the working group was not to duplicate existing texts but rather to offer a synthesis of available information, drawing upon both published sources and on the operations of specific projects, illuminated by the personal experience possessed by members of the group from their involvement in such projects. It is a summary of the best existing knowledge and practice drawn from around the world.

These guidelines therefore identify and discuss the key issues involved in the conceptualization, planning and implementation of a digitization project, with recommendations for "best practice" to be followed at each stage of the process. A special effort has been made to consider the particular circumstances of the countries of the developing world. Each of the eight sections comprises an introduction that sets the scene and identifies the relevant issues, followed by text which discusses the issues and actions to be taken in more detail and ends with one or more sections of "boxed" text which includes a summary of the main recommendations. As indicated in the Introduction, coverage is concerned only with the paper based documentary heritage, that is with manuscripts, printed books and photographs. It does not include coverage of the special issues relating to sound recordings or motion pictures, which will be treated in another set of guidelines sponsored by the UNESCO Memory of the World programme.

The group naturally realises that no single set of recommendations can possibly fit exactly the particular individual needs and circumstances of any one institution. It is also very conscious that this is a rapidly changing field with new developments constantly taking place in the appropriate technologies and in professional responses to these. It therefore hopes that these guidelines will not be seen simply as standing alone but will also be regarded as providing a gateway to further information. Extensive lists of references are provided in each Section, and these are consolidated into a more comprehensive list in the Appendices, together with URLs for ongoing discussions lists and other sources of current information. There are also URLs for relevant organizations in the library, archive, communications and standards fields and for actual digital projects.

The members of the working group were:

Jean-Marc Comment
Clemens de Wolf
Dale Peters
Borje Justrell
John McIlwaine
Marie-Thérèse Varlamoff

John McIlwaine, Chairman, March 2002

INTRODUCTION

Digital technology opens up a totally new perspective. The World Wide Web holds millions of websites and the Internet is the market place for research, teaching, expression, publication and communication of information. Libraries and archives are society's primary information providers and were early users of the new digital technology with respect to cataloguing and processing management, and later for providing information on their collections to the www-community. Besides preserving and providing access to `born digital material' a great number of archives and libraries nowadays have also turned to creating digital surrogates from their existing resources. It is for those libraries and archives that these guidelines have been compiled.

Definition

These are Guidelines for undertaking digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives. They deal with the paper documentary heritage, manuscripts, printed books and photographs, **not** with sound recordings or motion pictures, nor with artifacts, nor monuments. They are concerned with planning and setting up projects, which means with the selection, management and production processes involved in such projects in well defined, separately financed and usually short term activities, not with programmes as an integral part of an institution's mission or strategy.

Why Guidelines?

Many libraries and archives would like to plan digitization projects but lack experience

There is a need for a practical guide as a working tool for planning digitization projects

This need is particularly felt in developing countries

UNESCO

These Guidelines fit within UNESCO's strategy of knowledge for all. They also have a strong relationship with UNESCO's Memory of the World Programme which is aimed at safeguarding the world's documentary heritage, democratising access to it, and raising awareness of its significance and of the need to preserve it.

Target audience

These guidelines are aimed at decision makers, library and archive managers, and curatorial and technical staff members, particularly those in institutions in developing countries

Why digitize?

The reasons for implementing a digitization project, or more precisely for digital conversion of non-digital source material, are varied and may well overlap. The decision to digitize may be in order to:

To increase access: this is the most obvious and primary reason, where there is thought to be a high demand from users and the library or archive has the desire to improve access to a specific collection

To improve services to an expanding user's group by providing enhanced access to the institution's resources with respect to education, long life learning

To reduce the handling and use of fragile or heavily used original material and create a "back up" copy for endangered material such as brittle books or documents

To give the institution opportunities for the development of its technical infrastructure and staff skill capacity

From a desire to develop collaborative resources, sharing partnerships with other institutions to create virtual collections and increase worldwide access

To seek partnerships with other institutions to capitalize on the economic advantages of a shared approach

To take advantage of financial opportunities, for example the likelihood of securing funding to implement a programme, or of a particular project being able to generate significant income.

Be clear why you are embarking on a digitization project: the purpose will determine the process and the costs. Since digitisation is both labour intensive and expensive (***see*** chapter 4) it is important to capture an image in a way that makes it possible to use it to serve several needs

Before you start, ask yourself

Is the project?

User driven: high demand for (enhanced) access
Opportunity driven: money available so we can do something
Preservation driven: high demand on fragile objects
Revenue driven: we might make some money from it

Do we have?

The money
The skills
The capacity
The technical infrastructure

Carry out

Benchmarking study
Copyright study
Feasibility study
Technical pilot study

Components

The key components of a digital imaging project are:

Selection policy

Conversion

Quality control programme

Collection management

Presentation

Maintaining long term access

All these components are equally important - the chain is not stronger then its weakest link.

Making the decision

Digital technologies are undergoing rapid and continuing development and many issues are unresolved, giving rise to a delusive reliance on the “wait-and-see” approach. The basis of a commitment to going digital is an acknowledgement that the technology will change and change often. The crucial management decision is therefore less about the “when”, or the “whether” to begin. It is rather a question of whether the institution can afford to ignore the opportunity to reach wider audiences in a global community, in a manner afforded by the technology to improve access to and the preservation of cultural and scholarly resources.

Digitization will be a costly exercise, requiring detailed planning and the establishment of an infrastructure to ensure continued access to the digital file. Institutions in countries of the developing world especially should consider whether the costs and time involved will be commensurate with the benefits. Such institutions should for example be prepared to resist encouragement in the implementation of a digitization project by outside donor agencies, when analysis shows that for example the use of microfilm would be adequate, even preferable.

Users

Obviously, the user plays an important role in the decision to begin a project, but which role, is very often hard to define. Indeed the specific demands of the user may be difficult to know. In most cases there is a supposed user’s group, and it is the aim of the institution to increase its services and expand its approach and influence. The user group may differ, depending on the type of institution and the mission of the organisation. Institutions of higher education fulfil faculty staff and students needs. Public and national institutions must satisfy a large and more diverse population. This influences not only selection but also the forms of presentation and accessibility (the user’s interface).

Preservation

Digitization is not preservation: digitization is not cheaper, safer or more reliable than microfilming. Unlike a frame of high quality microfilm, a digital image is not a preservation master. The only way that digital reformatting contributes positively to preservation is when the digital surrogate reduces physical wear and tear on the original, or when the files are written to computer output microfilm that meets preservation standards for quality and longevity. A digitization project is therefore no replacement for a preservation programme based on reformatting on microfilm (or on deacidification, conservation treatment or improved storage conditions).

This is in general true. But there may be specific circumstances, for example in developing countries, that can turn this notion on its head. If an institution with no experience nor facilities for preservation at all, wants to preserve a specific collection, it may decide to invest in digital instead of microfilming equipment, thus avoiding the high expenditure on microfilming cameras and processors and realizing that this digital equipment and the developed staff skills will serve other purposes as well. This shifting from the generally recommended method of preservation microfilming into digitization with its risks in the long term is perhaps not the ideal solution for the problem of nineteenth and twentieth century paper decay but can serve as a practical way of providing protection to certain documents.

Digital technologies offer a new preservation paradigm. They offer the opportunity of preserving the original by providing access to the digital surrogate; of separating the informational content from the degradation of the physical medium. In addition, digital technologies liberate preservation management from the constraints of poor storage environments typical of the tropical and sub-tropical climates in which many developing countries are located.

Cost saving

Digitization does not result in cost savings for collection management. A digital surrogate can never replace the original item or artefact. If an institution wants to save space by deaccessioning the brittle newspapers, it would do better to create microfilm copies rather than digital images (and, even better, decide not to throw away the microfilmed newspaper copy at all).

The whole process, selection, scanning, creating records etc. requires heavy expenditure and the long-term maintenance of the digital assets has its own high costs. An institution may wish to investigate the possibilities of cost recovery by marketing digital copies (see **Sections 3** and **4**).

Urgency of building digital repositories

Preservation of digital information is undoubtedly expensive and requires highly skilled technical staff and equipment.

Individual libraries embarking on digital projects should seek co-operation within regional, national and international agreements and should look to conclude agreements with trusted repositories (see **Section 7**)

Other decisions to be made

Whether to use a digital process which reproduces the image, or to use OCR (Optical character recognition) or actual keying in of the source text. It is likely that users will want searchable texts, and that means OCR or re-keying (in most cases the latter will be cheaper than the former, but there is no rule of thumb on this and a combination of methods may be suitable). On the other hand, depending on the type of users and the kind of text, many users will want to see the page images as well, and experience a touch of the original. This may lead to the conclusion to use both methods but in most cases this would be cost prohibitive. Then the best way is to choose page images.

Whether to produce digital files capable of handling every job traditionally carried out by conventional photographic services (i.e. images for professional publications, displays for exhibitions etc.).

Whether to digitize from the original or from microfilm. The latter represents the so-called hybrid approach investigated in particular by Cornell University and by the Project Open Book at the University of Yale.

Suggested reading

COMMISSION ON PRESERVATION AND ACCESS. *Digital imaging and preservation microfilm: the future of the hybrid approach for the preservation of books*. Washington, DC, 1999. http://www.clir.org/pubs/archives/hybridintro.html

COUNCIL ON LIBRARY & INFORMATION RESOURCES (2001)*. Building and sustaining digital collections: models for libraries and museums*. Washington, DC. (Publication 100) http://www.clir.org/pubs/reports/pub103/contents.html

COUNCIL ON LIBRARY & INFORMATION RESOURCES (2001). *The evidence in hand: Report of the Task Force on the Artifact in Library Collections*. Washington DC. (Publication 103) http://www.clir.org/pubs/reports/pub103/contents.html

KENNEY, A.R. & RIEGER, O. (2000) *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, VA, Research Libraries Group (RLG)

SMITH, Abby (2001). *Strategies for building digitized collections*. Washington, DC, Council on Library & Information Resources (Publication 101)
http://www.clir.org/pubs/reports/pub101/contents.html

SMITH, Abby. (1999). *Why digitise?* Washington, DC, Council on Library & Information Resources (Publication 80). http://www.clir.org/pubs/reports/pub80-smith/pub80.html

Related resources

British Library, U.K. Objectives of digitisation
http://www.bl.uk/about/policies/digital.html

CORNELL UNIVERSITY. DEPARTMENT OF PRESERVATION & CONSERVATION. Moving theory into practice: Digital Imaging Tutorial
http://www.library.cornell.edu/preservation/publications.html (To accompany KENNEY, A.R. & RIEGER, O. (2000) *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, VA, Research Libraries Group (RLG) *see above*)

DIGITAL LIBRARY FEDERATION. Digital Library Standards and Practices
http://www.diglib.org/standardspv.htm

Library of Congress. American Memory "a gateway to rich primary source materials relating to the history and culture of the United States. The site offers more than 7 million digital items from more than 100 historical collections". http://memory.loc.gov/

UNESCO Memory of the World http://www.unesco.org/webworld/mdm/index_2.html

UNESCO Virtual Memory of the World http://www.unesco.org/webworld/en/memoire.html

1. SELECTION

Background

It is important to see digitisation as a series of choices where competing requirements and demands have to be balanced. When selecting source material for digitisation it comes down to three basic questions: whether the source material

> *Needs to be converted?*
>
> *Should be converted?*
>
> *Can be converted?*

The selection therefore has to be conducted in such a way that it will assure that not only issues like the value of the selected material and interest in its content are considered but also demands concerning technical feasibility, legal matters and institutional conditions.

Issues involved in the selection of material for digitisation will be examined from two perspectives:

> *Principal reasons for digitisation (to enhance access and/or preservation)*
>
> *Criteria for selection (based on content or based on demand)*

1.1 Principal reasons for digitisation

1.1.1. For enhanced access

As noted in the **Introduction** there can be several reasons for increasing accessibility:

> Enhancement of access to a defined stock of research material
>
> Creation of a single point of access to documentation from different institutions concerning a special subject
>
> Implementation of the “virtual re-unification” of collections and holdings from a single original location or creator now widely scattered (see also Virtual Collections below)
>
> Support for democratic considerations by making public records more widely accessible
>
> Extending the availability of material in support of educational and outreach projects

The key point is to evaluate the contribution that increased access could make to a defined user community. If the institution planning a digitization project is a private one, it is normal for it to focus on specific needs and to target a specific user group. If however a public institution is involved, it will probably have to satisfy a larger population and more diverse demands.

The way that it is intended to use a digital image is of vital importance in shaping the technical requirements. For example will the amount of information captured in the digital conversion set the boundaries for the usability of the digital images? (see **Section 2**).

1.1.2 To facilitate new forms of access and use

The main purpose in this case is to enable the use of material (original manuscripts and archives, maps, museum artefacts, rare books etc.) :

- that cannot be consulted in its original form other than by visiting its specific repository

- that has been damaged and where technological support is needed to reveal its content or shape (data recovery)

- in an easier and more productive way than by using computer enhancement tools like OCR (Optical Character Recognition) or text encoding for converted texts.

In such cases, the focus may be principally on how to add value to the source material and not on digitization as such. Sometimes costs and technical limitations will make it easier to use solutions other than digital conversion, or hybrid solutions involving both digitization and microforms.

1.1.3 For preservation

When digital conversion deals with source materials which are endangered or damaged, the purpose is, in the first place, to create accurate reproductions of these originals on a long-lasting medium and not to select materials according to demand. These reproductions need to satisfy both users of today and future potential users, and must therefore both be of high quality and possess a physical stability that can be maintained over time.

One method of selecting source materials for preservation is to classify them into three categories:

- *Rare, unique or fragile documents, archives and other objects of artifactual value that need to be retained in their original form*: Digital conversion can provide high quality surrogates with quick and broad access which in most cases will protect this kind of material from handling. This can be difficult to achieve using some kinds of microform.

- *Source material with an important intellectual but relatively low artifactual value, highly used and damaged or fragile*: Digital images are normally good replacements for serving immediate demands. If the source materials are deteriorating and, therefore, need to be replaced permanently, archives and libraries sometimes prefer to produce microfilm for preservation purposes and digital copies for access (a hybrid solution).

- *Mostly brittle source material of high intellectual but low artifactual value and with a low level of use.* This is not material that will be of interest for digitization in the first place. If it is brittle material that needs to be replaced by surrogate copies to allow use, then microfilm is still the normal choice in many countries being stable, cheap and easy to store (but note comments on the situation in some developing countries noted in the **Introduction** above). In the future, when researchers discover this source material and perhaps use it more frequently, there will always be the possibility to digitise the microfilm

Many institutions have not yet accepted digital technology as being stable enough for long-term preservation. The reasons are often that they feel the threat of technical obsolescence of the digital medium and an uncertainty both about the legal status of electronic documents and about the future costs of preservation of such documents (see **Sections 3** and **7**). While waiting for the problem of digital longevity to be solved, most institutions are creating archival images (see above) of what can be called "preservation quality". That means that they:

- can be used for different purposes
- are created at a quality level that will minimise the need for rescanning (see **Section 2**)

The fact that a surrogate has been created is certainly not enough to justify disposal of the originals. Even to be accepted as the text for consultation by the reader rather then the orginal the digital images must:

- have a guaranteed authenticity (see **Section 3**)
- be a part of a preservation plan (see **Section 8**).

Disposal of original source documents after digital conversion is sometimes used in records management programmes but only for documents that have already been appraised and scheduled for disposal, and which have been digitized to facilitate heavy use during their intended life time.

1.2 Criteria for selection

It is useful when planning a digitisation project to look at policies established by other institutions for their own projects. Many of these are now available for consultation on the Web. An example is that of the University of Columbia which has developed a set of selection criteria for digital imaging divided into six categories: collection development, added value, intellectual property rights, preservation, technical feasibility and intellectual control. Another example is the Library of Congress where the selection for preservation digital reformatting is based on value, use, characteristics of the original item, and appropriateness of digital reproduction for use and access. (**See "Suggested reading" at end of this Section** for references to these and other policy statements)

1.2.1. Content

Regardless of the purpose for implementing a digitisation project, the selection of source material will always be more or less content driven. In fact, intellectual value is the basic question in all kind of selection: does the content (the value to the potential reader) of this material justify all the efforts, costs and other resources that will be needed? Therefore, every digitisation project or programme ought to have its own definitions of value based on the goals it trying to achieve.

- *Virtual collections*

During the last ten years scholars have started to build up virtual collections of scanned documents, books, museum artifacts etc. The selection is normally based on the intellectual content of the material, but it could as well be built on the physical appearance or on other factors like age etc. The purposes of building virtual collections may differ. It could for example be to re-unify scattered collections and holdings (see above) or to enhance research by integrating different source material that otherwise would have remained separate items located in different parts of the world. The possibilities of providing widespread access over the Internet plays an important role here.

- *Collecting a critical mass of information*

To make a digitization project worthwhile requires a certain minimum volume of information. Otherwise the research value will be too low to attract enough either planned or potential users. An important question is, therefore, if a selection is being made based on content, should all of a collection be included or only parts of it? Normally the value of archival material, photographic collections etc. is higher as aggregates rather than as single parts taken out of context, but if individual documents or objects have significant research value, even a few of them can form a critical mass of information

1.2.2. Demand

The level of demand is of course of great interest when selecting source material for digitization. If the purpose is mainly to enhance access, the likelihood of significant use of a digitized material will probably govern the selection process. Involving scholars and other researchers in the original decision is therefore a traditional selection methodology.

A basic question, however, is what audience the digitising institution should interact with or maybe give priority to. The answer depends on the mission of the institution in mind, but eventually also on political goals and what society expects from its cultural institutions.

Sometimes an active user group for a specific source material may be spread all over the world and because of that it can be difficult to define or even detect. Materials in special collections often run the risk of being looked upon as little-used, which is not necessarily true since a small specialist group can generate a great deal of important research.

To balance the demands of different user groups many institutions have boards of scholars and other researcher to help them select material that is most urgent to digitise. When an institutions digitising activities are being developed from general proposals to specific projects covering whole collections or types of documents or objects, these advisory boards can be strategically important.

For cultural institutions starting their first digitising project, a good rule of thumb is that selecting the most heavily used parts of their collections will normally give the greatest added value because it will satisfy the majority of the people they try to serve.

1.2.3 Condition

Selection of material for digitization will be affected both by its physical condition and by the existing quality of the bibliographical descriptions available for it. Material which is fragile, damaged and in poor condition may present too many risks of further damage being caused by handling to allow it to be scanned without special care, or some basic conservation treatment. This will involve additional costs, and the institution will need to consider whether other collections in better condition should have priority, or whether the costs of preparation and conservation should be built in to the costs of the overall digitization project. (See discussion below under **Section 4, Budgeting**)

Similarly, if the material being considered as a candidate for digitization lacks detailed cataloguing or descriptive data, it is essential for future access to such material to create such data, and it will therefore need to be considered whether the necessary costs of doing this can be included in the overall budget of the digitization project.

Recommendations

Formulate a policy for the selection of material to digitise at an early stage in the project

Identify the principal reasons behind the project. Is it to enhance access, to support preservation, or a mixture of both?

Decisions about technical requirements, indexing and searching, and preservation of the digital files created all depend on the project's rationale

Create a set of criteria for selection

Consider setting up a special advisory board of scholars or other researchers to represent potential users of the digital files and to help in the selection of what it is most urgent to digitise

Capture an image in a way that makes it possible to use it to serve several needs, and store it as an archival image off line on a cheap and secure storage medium (master file). Surrogate copies of this image can be used for access (access files). Sometimes surrogate copies are made very small and used only as browse images to give an idea of their content (thumb nail files).

.

Suggested reading

AYRIS, P. (1998). Guidance for selecting material for digitization, *in* NATIONAL PRESERVATION OFFICE/RESEARCH LIBRARIES GROUP (1998). *Guidelines for digital imaging: papers given at the joint NPO/RLG Preservation Conference, 1998.* London
http://www.rlg.org/preserv/joint/ayris.html

COLUMBIA UNIVERSITY LIBRARIES. Selection criteria for digital imaging.
http://www.columbia.edu/cu/lweb/projects/digital/criteria.html

COMMISSION ON PRESERVATION AND ACCESS (1999). *Digital imaging and preservation microfilm: the future of the hybrid approach for the preservation of books*. Washington, DC.
http://www.clir.org/pubs/archives/hybridintro.html

De STEFANO, R. (2000). Selection for digital conversion *in* KENNEY, A.R. & RIEGER, O. *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, VA, Research Libraries Group (RLG)

GERTZ, J. (1998). Selection guidelines for preservation *in* NATIONAL PRESERVATION OFFICE/RESEARCH LIBRARIES GROUP (1998). *Guidelines for digital imaging: papers given at the joint NPO/RLG Preservation Conference, 1998.* London.
http://www.rlg.org/preserv/joint/gertz.html

HARVARD UNIVERSITY LIBRARY. Selection for digitization. A decision-making matrix.
http://preserve.harvard.edu/bibliographies/matrix.pdf

HAZEN, D. et al. (1998). *Selecting research collections for digitization*. Washington, DC, Council on Library & Information Resources.(Publication 74)
http://www.clir.org/pubs/abstract/pub74.html

KENNEY, A.R. & RIEGER, O. (2000). *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, VA, Research Libraries Group (RLG)
LIBRARY OF CONGRESS. Preservation Digital Reformatting Program. Selection criteria for preservation digital reformatting http://www.lcweb.loc.gov/presv/prd/presdig/presslection.html

MENNE-HARITZ, A. & BRÜBACH, N. (1997). *The intrinsic value of archive and library material: list of criteria for imaging and textual conversion for preservation* . Marburg, Archivschule.
http://www.uni-marburg.de/archivschule/intrinsengl.html

NATIONAL PRESERVATION OFFICE (1997). *Preservation and digitisation: principles, practices and policies: papers given at the NPO 1996 Annual Conference*. London.
http://www.bl.uk/services/preservation/confpapers.html

NATIONAL PRESERVATION OFFICE/RESEARCH LIBRARIES GROUP (1998). *Guidelines for digital imaging: papers given at the joint NPO/RLG Preservation Conference, 1998.* London.
http://www.rlg.org/preserv/joint

SMITH, Abby (2001). *Strategies for building digitized collections*. Washington, DC, Council on Library & Information Resources (Publication 101)
http://www.clir.org/pubs/reports/pub101/contents.html

UNIVERSITY OF CALIFORNIA (UCLA) LIBRARY. Digital projects. Guidelines and standards.
http://www.digital.library.ucla.edu (especially "Guidelines for Choosing Metadata" and "Standards reference guide")

WEBER, H. & DÖRR, M. (1997) *Digitisation as a method of preservatio*n? Amsterdam, European Council on Preservation & Access.
http://www.clir.org/pubs/reports/digpres/digpres.html

2. TECHNICAL REQUIREMENTS AND IMPLEMENTATION

2.1 Conversion

A digital image is an ”electronic photograph” mapped as a set of picture elements (pixels) and arranged according to a predefined ratio of columns and rows. The number of pixels in a given array defines the resolution of the image. Each pixel has a given tonal value depending on the level of light reflecting from the source document to a charge-coupled device (CCD) with light-sensitive diodes. When exposed to light they create a proportional electric charge, which through an analogue/digital conversion generates a series of digital signals represented in binary code. The smallest unit of data stored in a computer is called a bit (binary digit). The number of bits used to represent each pixel in an image determines the number of colours or shades of grey that can be represented in a digital image. This is called bit-depth.

Digital images are also known as bit-mapped images or raster images to separate them from other types of electronic files such as vector files in which graphic information is encoded as mathematics formulas representing lines and curves.

Source documents are transformed to bit-mapped images by a scanner or a digital camera. During image capture these documents are "read" or scanned at a predefined resolution and bit-depth. The resulting digital files, containing the binary digits (bits) for each pixel, are then formatted and tagged in a way that makes it easy for a computer to store and retrieve them. From these files the computer can produce analogue representations for on-screen display or printing. Because files with high-resolution images are very large it may be necessary to reduce the file size (compression) to make them more manageable both for the computer and the user.

When a source document has been scanned, all data is converted to a particular file format for storage. There is a number of widely used image formats on the market. Some of them are meant both for storage and compression. Image files also include technical information stored in an area of the file called the image “header”.

The goal of any digitisation programme should be to capture and present in digital formats the significant informational content contained in a single source document or in a collection of such documents. To capture the significant parts, the quality assessments of the digital images have to be based on a comparison between those digital images and the original source documents that are to be converted, not on some vaguely defined concept of what is good enough to serve immediate needs. However, the solution is not to capture an image at the highest quality possible, but to match the conversion process to the informational content of the original - no more and no less. This raises two questions: (1) the attributes of the source documents to be digitised and (2) the image quality.

2.1.1 The attributes of the source documents

At capture, consideration has to be taken both of the technical processes involved in digitisation and of the attributes of the source documents. These attributes could be of varying dimensions and tonal range (colour or black and white). Source documents can also be characterised by the way in which they have been produced: by hand (ink), by a typewriter or printer, or by photographic or electronic methods.

The physical condition of the source documents can affect the conversion in different ways. Fading text, bleed-through of ink, burned pages and other kinds of damage sometimes destroy the informational content but more often set physical limitations on the possibilities of catching information during a scan. Therefore, the need for pre-scanning treatment of the source documents has to be identified. Neglecting this can not only be a threat to the documents themselves but can also limit the benefits and results of digitisation and increase the cost. Ordinary steps to prevent this are for example to carry out preliminary elementary conservation treatment, and to use book cradles for bound volumes, and routines to control lighting and other environmental conditions during the actual scanning. If the source documents have artifactual value they will normally need to be examined by a conservator before scanning.

When the risks of damage to the source documents are high and the documents are of special value or in bad condition, it can sometimes be better to scan from film intermediates instead of from the original documents, if such film is available.

2.1.2 Image quality

Image quality at capture can be defined as the cumulative result of the scanning resolution, the bit depth of the scanned image, the enhancement processes and compression applied, the scanning device or technique used, and the skill of the scanning operator.

2.1.2.1 Resolution

Resolution is determined by the number of pixels used to present the image, expressed in dots per inch (dpi) or pixels per inch (ppi). The difference between dpi and ppi is described below in **Section 2.2**.

Increasing the number of pixels used to capture an image will result in a higher resolution and a greater ability to delineate fine details, but just continuing to increase resolution will not result in better quality, only in a larger file size. The key issue is to determine the point at which sufficient resolution has been used to capture all significant details in the source document.

The physical size of a source document is of importance when determining the resolution. When the dimensions of the document increase, the number of pixels needed to capture required details in it will increase too, as well as the file size. Large files can cause problems for users when viewing the images on a screen or in sending them over networks, because the file size has an important impact on the time it takes to display an image. One way to decrease the file size is to decrease the resolution. This is a critical decision, especially if the source document has both a large physical size and a high level of detail, which can be the case with oversized maps and drawings.

2.1.2.2 Bit depth

Bit depth is a measurement of the number of bits used to define each pixel. The greater the bit depth used, the greater the number of grey and colour tones that can be represented. There are three kinds of scanning (digital sampling):

- *bitonal scanning* using one bit per pixel to represent black or white

- *greyscale scanning* using multiple bits per pixel to represent shades of grey; the preferred level of grey scale is 8 bits per pixel, and at this level the image displayed can select from 256 different levels of grey.

- *colour scanning* using multiple bits per pixel to represent colour; 24 bits per pixel is called true colour level, and it makes possible a selection from 16.7 million colours.

The choice of bit depth affects the possibility of capturing both the physical appearance and the informational content of a source document. Decisions about bit depth therefore have to take into account whether the physical appearance, or parts of it, have an added informational value and need to be captured. This can be the case when the purpose of the digitisation project is to produce facsimiles of the source documents.

2.1.2.3 Image enhancement processes

Image enhancement processes can be used to modify or improve image capture by changing size, colour, contrast, and brightness, or to compare and analyse images for characteristics that the human eye cannot perceive. This has opened up many new fields of applications for image processing, but the use of such processes raises concerns about fidelity and authenticity to the original. Image processing features include for example the use of filters, tonal reproduction curves and colour management tools.

2.1.2.4 Compression

Compression is normally used to reduce file size for processing, storage and transmission of digital images. Methods used are for example to abbreviate repeated information or eliminate information that the human eye has difficulty in seeing. The quality of an image can therefore be affected by the compression techniques that are used and the level of compression applied. Compression techniques can be either “loss less”, which means that a decompressed image will be identical to its earlier state because no information is thrown away when the file size is reduced, or “lossy” when the least significant information is averaged or discarded in this process.

In general “loss less” compression is used for master files and “lossy” compression techniques for access files. It is important to be aware that images can respond to compression in different ways. Particular kinds of visual characteristics like subtle tonal variations may produce unintended visual effects.

Digital images reproduced from photographic formats have a wide tonal range, commonly resulting in large files. Another technique besides compression that can be used to reduce file size is to limit the spatial dimension of the digital image (for spatial resolution, see **Section 2.2**). This can be done when the image is intended to be an archival reproduction rather than a facsimile replacement of the original.

2.1.2.5 The equipment used and its performance

The equipment used and its performance has an important impact on the quality of the image. Equipment from different manufacturers can perform differently, even if it offers the same technical capability.

2.1.2.6 Operator judgement and care

Operator judgement and care always have a considerable impact on image quality. In the end it is decisions taken by humans which decide what quality will be achieved.

Recommendations for conversion

A ten step guide to ensure a good conversion process

1. Use scanners that can accommodate:

- *the physical dimensions of the source documents*
- *the type of media involved (transparent or reflective)*
- *the range of details, tones and colours present in the documents*
- *the physical condition of the documents*

2. Examine carefully any requirements for special handling or conservation of the source documents prior to scanning

3. Choose a resolution that will be sufficient to capture the finest significant details required in the group of source documents that is to be scanned. Check that the resolution will not limit the intended use of the digital images. Set the resolution at the chosen level for the entire group of source documents in order to avoid the need for an item by item review

4. Choose a bit depth that is in accordance with the characteristics of the source documents, and of a level necessary to transfer the informational content: bitonal scanning for textual documents consisting of a black image on white paper; greyscale (8 bits) scanning for documents containing significant greyscale information (including pencil annotations on text) and for photographic materials; colour scanning for documents containing colour information, especially when high quality facsimile copies are required.

5. Use enhancement processes cautiously and document carefully all such processes that are used

6. Use "lossless" standard compression techniques for master files and for access files when needed. This means for example:

- *for compression: ITU Group 3 or 4 and JBIG (binary images), lossless JPEG/JPEG 2000 or LZW (multi-bit images)*
- *for exchange: lossless JPEG/JPEG 2000, TIFF 5 or later versions*

7. Try by careful testing of access files to find a balance between an acceptable visual quality for the user and a file size that the computer can access with an acceptable delay.

8. To obtain a stable performance from equipment in use, carefully investigate manufacturers' claims of system capabilities and confirm these through sampling and taking up references

9. Use standards for digital image quality evaluation (see ***2.2. Quality control*** *below)*

10. Implement a continuous quality control programme to verify the consistency of output by individual human operators during the scanning process (see ***2.2. Quality control*** *below)*

2.2 Quality control

Quality control is an important component in every stage of a digital imaging project. Without this activity it will not be possible to guarantee the integrity and consistency of the image files.

Steps need to be taken to minimise variations between different operators as well as between different the scanning devices in use. Scanners most also be regularly controlled to verify accuracy and quality.

A quality control program is needed both for in-house projects and for projects where all arrangements or parts of them are outsourced. An important difference is that in a partly or totally outsourced project the quality requirements often have to be formulated before a contract is signed, due to its legally binding nature. In-house projects can built up their quality control programmes step by step as a part of their project activities.

Although quality control is a crucial factor to ensure the best results, there is no standard way to ensure a certain image quality at capture. Different source documents require different scanning processes, and this has to be considered when developing a quality control programme.

2.2.1 Substance of a quality control programme

2.2.1.1 Scope

An important question for a quality control programme is whether it should include:

- the whole image collection or a sample of images?
- all kinds of files (master files, access files, thumbnail files)?
- other intermediate products like paper facsimiles and microforms?

The answer depends on the purpose of the digitisation project, the required out-put, and the quality levels and reference points chosen. If the digitisation programme is very limited or the quality requirements are extraordinary high, it will make sense to examine the whole collection image by image. However, in most programmes it is enough to set up a sampling plan covering for example 10% of all images produced by each scanning device during a certain time period (day, week, month). If a specified percentage of the chosen images is found to be incorrect then the whole batch will have to be subjected to control.

A quality control programme always covers the master files that are produced and in most cases will also cover other out-puts such as access files, microforms and paper copies

2.2.1.2 Methods

The automated image evaluation tools that are available today are normally not sufficient for materials that are required for cultural and scientific purposes. Therefore, visual quality evaluation has to be done:

- either from on-screen or print-outs
- based on a mix of on-screen evaluation and print-outs (film or hard copies)

Technical limitations that can affect the evaluation must be considered, beginning with the possibilities of getting good quality printed hard copies of grey scale and colour images. Recommended methods for

- on screen evaluation are
    - view scanned images at 1:1 (100% enlargement)
    - use targets to evaluate greyscale and colour reproduction
    - use resolution targets and histograms to evaluate spatial resolution and tonal reproduction
    - use signal to-noise measurement and artifact detection tools

- print-out evaluation are
    - examine by human eye hard copies created from the images to see if they fit the quality requirements
    - compare the print-outs with the source documents

2.2.2 Scanner quality control

Before a scanner is bought, vendors should be required to deliver measurable digital results from relevant digital image quality evaluation tests. When a digital imaging project is running, scanning quality control measures must be set to enable operators to ensure that the scanning device is operating within anticipated tolerances. Issues of main concern in performance are: spatial resolution, tonal reproduction, colour reproduction, noise, and artifacts detection. In projects which are scanning oversized material, such as maps and plans, geometric accuracy is also an important factor.

2.2.2.1 Spatial resolution

A common definition of spatial resolution is the ability to capture and reproduce spatial details. It covers both input and output devices and that is probably one reason why the concept of resolution is one of the most misunderstood and misused technical specifications applied to digitising equipment. Resolution is often specified in terms of dpi (dots per inch). However, dpi should normally be used only for printers, as ”d” always refers to printed dots (e.g. ink jet printers and laser printers). For input resolution (i.e. scanners and digital cameras) and on-screen resolution (i.e. monitors) pixels per inch (ppi) normally should be used. A pixel is in general a much smaller physical unit than a dot.

When it says that a scanner has a maximum resolution of for example 600 dpi, it means in practice that the scanner optically samples a maximum 600 pixels per inch (ppi). But the optical sampling rate of a scanning device only delineates the maximum possible (optical) resolution in the direction of the extension of the CCD unit. It will not guarantee that the scanner in reality can spatially resolve details to the same degree that the optical sampling rate would imply. The reason is that the optical sampling rate of an input device is only one component of the concept of resolution. Other components of importance are for example the quality, focal range and mechanical stability of the optical system (lens, mirrors and filters), the input/output bit-depth, the vibrations of the source document and the CCD, and the level of image processing applied to the image.

There are several methods for evaluating resolution. Commonly used are the following:

- *Resolution targets,* which were originally made for use in micrographic and photographic industries. They are normally used to measure the reproduction of details, uniform capture of different parts of a source document, image sharpness etc. The results can sometimes be not fully trustworthy, but resolution targets are still practical tools to use especially for bitonal conversion.

- *The Modulation Transfer Function (MTF),* in which the spread of light in the imaging process (line spread function) is measured. This is a more reliable and objective way to evaluate how well details are preserved and suits best greyscale and colour systems

- *Spatial Frequency Response (SFR),* which means measuring the ability of the scanner to transmit high-frequency information by means of a specified transfer function (in practice equivalent to MTF)

Examples of targets in use for resolution are:

- IEE (Institute of Electrical and Electronic Engineers) standard Facsimile Test Chart
- AIIM Scanner Test Chart no 2
- PM-189 (A&P International) Resolution Target
- The Scanner SFR and OECF no 2 Target (Applied Image Inc)

A standard for resolution evaluation is ISO 12233 Photography, Electronic Still Picture Cameras, Resolution Measurements

2.2.2.2 Tonal reproduction

Tonal reproduction is the most important of all image quality metrics, because it gives the conditions for the evaluation of other image quality parameters. The effectiveness of these indeed assumes a satisfying tonal reproduction. In practice tonal reproduction determines how dark or light an image is as well as its contrast.

Due to various (electronic) noises in the scanner there will always be losses in the bit-depth during a scanning process. It is therefore important to capture an image with a higher bit-depth than is needed for the digital output, for example at least 12 and 14 raw bits/channel to get an 8 bit output (greyscale).

Tonal reproduction is evaluated by a tone reproduction curve that relates the optical density of a paper document or a microfilm reproduction to the corresponding digital count (tone value) in the digital reproduction. In digital systems this curve is called the Opto-Electronic Conversion Function (OECF).

Tonal values can also be assessed with a histogram which graphically shows the distribution of tones in an image and also the tonal range of it. Clippings in highlights or shadows compared to the tonal values of the source document can indicate that there are limitations in the dynamic range of the scanning device. Dynamic range can be described as the range of tonal difference between the lightest light and darkest dark and its value therefore shows the capacity of a scanner to distinguish extreme variations in density. Normally, the dynamic range of a scanner should meet or exceed the density extremes of the source documents.

It is important that no tonal compression is made in a scanned image at capture. If tonal compression is implemented at this stage of the conversion process, an image can never be restored to fully tonal value again. The scanner gamma value (brightness setting) should therefore always be set to 1.0 in the scanner software. This is sometimes called “linear tone reproduction”

2.2.2.3 Colour reproduction

The main challenge in digitisation of coloured source documents is to reproduce them with maintained colour representation on screen or on printouts. The main problem is that monitors as well as operation systems and system applications display colour in different ways. Human colour perception also differs between individuals

There are several colour models for defining the properties of the colour spectrum. The most used are RGB and CMYK.

RGB stands for red, green and blue and is the model used by monitors and scanners. The idea is to simulate a wide range of colours by combining different amounts and intensities of red, green and blue light. Each of these three colours is defined as a colour channel, and on a 24-bit monitor each channel has 8 bits representing 256 shades. In 1996 a special standard RGB, called sRGB, was created for the Internet, and it is often used for monitors as well as scanners, digital cameras, and printers. However, criticism has been made that sRGB is too limited and cannot reproduce all colours. Therefore, it is important to consider before buying a digital capturing system (camera or scanner) if this limitation is of significant importance in relation to the output required from the digitising project.

The CMYK model is based on cyan, magenta, yellow, and black. It is built on the principle that all objects absorb a certain wavelength from the light spectrum and always reflect an opposing wavelength. Printing and photographic systems are built on the CMYK model which also is called subtractive.

At the beginning of the 1990s a consortium of vendors created the ICC (The International Colour Consortium) with the purpose of developing vendor-neutral and standardised colour management systems. Their ICC standard profile can be used in different operating systems and be embedded in colour images. However, not all colour image management systems support embedded ICC profiles.

Examples of targets in use for colour and grey scale

- Kodak Color Separation Guide and Gray Scale (Q13 and Q 14)
- Kodak Q-60 Color input Target (IT8)
- RIT Process Ink Garmut Chart

2.2.2.4 Noise

In this context noise is the same as fluctuations in light intensity in an image that are not to be found in the source document. In digital imaging systems noise often has its origin in the CCD unit and in associated electronics. Noise is usually measured by calculating the standard deviation of pixel count values over a certain area as a signal to-noise ratio. The greater the deviation, the greater the noise which will reduce the quality of an image. Image software that can measure this is available.

2.2.2.5 Artifacts

Examples of artifacts that can have an impact on the quality of an image are dust, scratches, and streaks. They all create a visible non-random fluctuation in light intensity, but how this affects the image quality differs depending on the out-put requirements. In most cases it is enough to use existing software to detect artifacts, but sometimes visual examination is needed.

2.2.3 Monitor quality control

There are many single elements that can influence the quality of an image when displayed. Firstly, the brightness and the colour purity vary between the centre and the corners of the display. Secondly, the choice of LCD (liquid crystal display) or CRT (cathode ray tube) display can have an effect; the latter is still seen as the best when it comes to image viewing. Thirdly, colour management production needs regular calibration of the monitor according to a standard (the ICC standard profile, see above). Calibration tools are sometimes included in software applications. Calibrating a monitor means setting two values: the light intensity of the monitor (gamma) and the so-called white point (when all three RGB channels are illuminated). Set the gamma value to 1.8 -2.2 and the white point (colour temperature) to cool white (5000 kelvin). It should be remembered that settings of these values which are ideal for the evaluation of image quality may not be optimal for normal viewing by users.

2.2.3.1 Viewing conditions

Image evaluation always needs a controlled environment. It is also important to adapt this environment to the requirements for viewing, which differs between the monitor and the source documents. Monitors are best viewed in low light, but not darkened rooms, and source documents in bright light. Surround effects like reflections can affect the evaluation and have consequently to be minimised for example by using a neutral background (grey) and neutral colours (grey, black, white) on the operator's clothes.

Recommendations for quality control

The conditions for digital image quality can be described as three steps:

- *Identification of the desired end product and production goals. Decide what will be produced and when.*
- *Setting up of standards. Define acceptable levels of digital image quality based on both the attributes of the source documents and the capability of the digital imaging system that will be used.*
- *Decision on reference points. Decide what the output of the digitization process should be judged against.*

Dependent on these decisions select what the quality control programmes should include (a sample of images or the whole image collection; all kinds of files; paper facsimiles; other intermediate products)

Decide upon methods of evaluation of the digital output using both on screen and print-out evaluation and employing available targets for colour and grey scale and resolution and signal to noise measurement and artefacts detection tools

Regularly calibrate the monitors on which evaluation is being carried out, and minimise surround effects in thire viewing environment

2.3 Collection Management

The possibility of being able to use a collection of digital images in the way it was intended depends not only on conversion standards and quality controls, but also on how the collection is managed. If the purpose is to meet not only short term needs but also to provide accessibility over time, steps have to be taken to satisfy both current use and the expectations of future users.

Plans must be made for example to:

- *make scanned images appropriate to the ultimate intended use*
- *up grade distribution of images and user interface functionality*
- *transfer images to new technical platforms to meet increasing capacity for processing and handling of digital information*
- *migrate digital images to new file formats or physical media to ensure long-term accessibility*

To make scanned images usable, great concern should be taken relating to their storage. All image files that are produced by a digital image project must be organised, named and described in a way that fits the purposes of the project.

2.3.1 Organization of images

Before a name and a description of an image file is considered is has to be decided how it should be stored. Normally, the source documents being scanned are physically organised according to principles of archival or library arrangement. Holdings of documents are often divided into series, volumes and issues, and collections of manuscripts and photographic items have numbers.

The easiest way to handle this question is to translate the main principles of how the source documents are physically organised into a logical disc hierarchy in the computer. This should be done as far as possible according to existing standard systems. This is important to ensure the compatibility of file naming structures between different technical platforms. It must also be possible for the collection of image files to grow, and the way of organising them has therefore to be scalable.

2.3.2 Naming of images

Computers are not able by themselves to interpret logical relationships in a collection of source documents as for example sequences of folders and pages. Therefore, this has to be mirrored in the way the scanned image files are named. There are two approaches for this: (1) to use a numbering scheme that reflects numbers already used in an existing cataloguing system, or (2) to use meaningful file names. Both approaches are valid, and what best fits a certain collection or group of source documents should be chosen.

Every digital imaging project has also to adopt conventions for names, for tables of signs and for rules relating for example to punctuation and to the use of capital letters. It is important that these conversions are uniform. A standardised vocabulary is one of the corner stones in managing a collection of digital images.

File extensions are also important when giving names to image files. Many extensions have standard meanings and are employed widely, but care has to be taken when dealing with non-standard extensions. A list of the most common extensions can be found at Webopedia, an online encyclopaedia related to computer technology,
http://webopedia.internet.com/TERM/f/file_extension.html.

2.3.3 Description of images

To describe digital images there is a need for metadata, that is structured data about data. Metadata can also be defined as data that facilitates the management and use of other data. This is nothing new for archivists and librarians. The use of metadata is closely related to rules for the description and cataloguing of printed publications, archival records, and artifacts. The difference is that in the digital world additional categories of metadata are required to support computer navigation and the management of data files. Metadata describing digital images can contain information of different kinds. The Library of Congress project “Making of America II” in 1998 identified three categories of metadata:

Descriptive metadata for description and identification of information resources
Structural metadata for navigation and presentation
Administrative metadata for management and processing

These categories do not always have well-defined boundaries and often overlap. Cornell University in their digital imaging tutorial, *Moving theory into practice* has for each category summarized goals, elements and sample implications in a table which gives a good overview of the concept of metadata. See at:
http://www.library.cornell.edu/preservation/tutorial/metadata/metadata-01.html. Every digital imaging project, however, must choose a designated metadata solution based on its own goals.

2.3.4 Use of metadata

There are two main approaches to metadata solutions: data management techniques and document encoding

2.3.4.1 Data management techniques

The level of descriptive metadata always determines the level or possibilities of retrieval. Therefore, it is of crucial importance *at the time that a digital imaging project starts* to decide the deepest level at which the digital images will be searched. Moreover, existing metadata like finding aids, indexes etc. has to be surveyed, and if adequate, linked to the image files.

It also has to be considered if and how metadata generated in the digital conversion process should be accommodated. Today, TIFF (Tagged image file format) is the most common file format for storing master versions of digital images. In TIFF, but also in other graphic formats such as GIF, the software used by the scanner automatically creates a number of tags with technical and administrative information which is recorded into the file header, in other words directly into the file itself. The information in the TIFF header is stored in ASCII format and in that sense is platform independent.

The practice of recording metadata into the TIFF header is widespread, and the advantages are obvious: it ensures a close connection between the source document, the conversion process and the image file that is the result of the conversion.

Building up collections of digital images also means that efforts have to be made to make them accessible to people. Libraries all over the world have for decades used MARC (Machine Readable Cataloguing) as a metadata standard, but it has not been adopted by other cultural sectors. To meet new demands in retrieval, initially for the Web, the Dublin Core Metadata Initiative at the beginning of the 1990s presented a set of 15 descriptive elements of metadata. These elements are intended to be simple, international and cross-sectoral. Dublin Core is today one of the most widely accepted metadata standards in the world.

2.3.4.2 Document Encoding

Today many digital projects and programmes use SGML (Standard Generalized General Markup Language) or parts of it like XML (eXtensible Markup Language). The purpose is to bind together images and give access to structural elements in single objects or in whole collections. Document encoding can also be used in systems where data are taken from underlying databases and transformed into standardised representation for exchange.

There are some disadvantages with SGML. It is for example not supported by many software applications, and XML has more and more taken over as the most used markup language. A well-known initiative using SGML is TEI (Text Encoding Initiative), which has developed DTDs (Document Type Definitions) for encoding individual texts in the field of humanities. For encoding entire archival collections or holdings of objects, the Encoded Archival Description Initiative (EAD) has developed a DTD for encoding finding aids.

Recommendations for collection management

Organise the scanned image files into a disc hierarchy that logically maps to the physical organisation of the documents

Name the scanned files in a strictly controlled manner that reflects their logical relationships

Describe the scanned image files internally (using the image file header) and externally (using linked descriptive metadata files)

Building a Working Environment for a Digital Imaging Project

Running a digital imaging project means balancing the needs of known and potential users, the technological infrastructure used by the project, and the demands on available human and financial resources. Therefore, the technological capabilities of image capture devices and workstations (scanners or digital cameras, operating systems, internal memory, storage, display quality, networking capability, and speed) must be clear before a project starts, as well as the means of delivering image data to users.

An analysis of technical needs for a digital imaging project is usually conducted within the framework of a **pilot project** or study. It gives a project a chance to investigate, on a small scale, the feasibility to (1) carry through their plans and to (2) introduce digital technology into a library or an archive, if it is the first digitising project for the institution in mind.

Technical needs focus primarily on hardware and software, both dynamic in development for the foreseeable future. Although any list of minimum requirements is almost immediately obsolete and therefore must be fairly general, the **following basic information and communication technology package** should be sufficient to conduct a digitisation project on a basic level:

- An ordinary level PC (Intel Pentium Processor or equal) with the following additional characteristics:
    - At least twice the random access memory (RAM) recommended for current office requirements. The RAM type should also be at least a Synchronised Dynamic (SDRAM)
    - At least 1 Gigabyte free hard drive memory space additional to what is needed for installed software and the operating system (Windows or equal). Image processing is memory hungry, and long delays will negatively affect productivity.
    - A CD-writer, which is an essential peripheral to provide the means to copy the digital product to CD-ROMs, especially if storage space is a problem or if CD-ROMs are planned for resale.
    - A modem or an Ethernet network card for Internet access

- An A3 flatbed scanner optically capable of true resolution of 600 dpi (ppi) or higher. Interpolated results to achieve a higher resolution can result in an unacceptable loss of details. Smaller A4 flatbed scanners are not able to capture the full dimensions of imperial foolscap documents of which many archival collections are comprised. A transparency adapter is another useful optional extra for the capture of film-based materials and glass negatives in large formats.

- An office level black and white laser printer (600dpi) is required if print-outs will be used for quality control. If images are to be printed for sale this will require a high-resolution colour photo quality printer.

- A power management unit (UPS) is recommended in areas that experience uneven electrical power supply.

The acquisition of equipment should also incorporate a maintenance contract for a minimum of three and possibly even five years. It is generally considered that after three years the equipment will require upgrading, and after five years it will be obsolete and need to be replaced.

Software selections should be based on a serious consideration of open source (i.e. publicly available) solutions. Support of open source software under the terms of the GNU General Public License builds a digital library user community independent of market forces, limiting software obsolescence where it is not feasible to sustain expensive licensing structures attached to commercial products. http://www.nzdl.org/greenstone digital library software provides a new way of organizing information and publishing it on the Internet or on CD-ROM. Proposed software developments of an XML base class in the Greenstone software suite are eagerly anticipated as a promising open source solution to collections management of images and related metadata.

The requirements of imaging software focus on the provision of standard file formats offered, notably .TIFF and .JPEG. Sophisticated graphic software tools should be used judiciously to preserve archival integrity, with image enhancement limited to automatic settings of the scanner software

Suggested reading

BACA, M. ed. (2001) Introduction to metadata: pathways to digital information. Version 2.0. Malibu, CA, Getty Standards Program.
http://www.getty.edu/research/institute/standards/intrometadata/index.html

BESSER, H. & TRANT, J. (1995). *Introduction to imaging: issues in constructing an image database*. Malibu, CA, Getty Information Institute.
http://www.getty.edu/research/institute/standards/introimages/

CHAPMAN, S. (1998). Guidelines for image capture *in* NATIONAL PRESERVATION OFFICE/ RESEARCH LIBRARIES GROUP *Guidelines for digital imaging: papers given at the joint NPO/RLG Preservation Conference, 1998.* London.
http://www.rlg.org/preserv/joint/chapman.html

COLUMBIA UNIVERSITY LIBRARIES. (1997). *Technical recommendations for digital imaging projects. Prepared by the Image Quality Working Group of ArchivesCom, a joint Libraries/AcIS committee.* http://www.columbia.edu/acis/dl/imagespec.html

CORNELL UNIVERSITY. DEPARTMENT OF PRESERVATION & CONSERVATION. Moving theory into practice: Digital Imaging Tutorial
http://www.library.cornell.edu/preservation/publications.html

ESTER, M. (1996). *Digital image collections: issues and practice*. Washington, DC, The Commission on Preservation and Access. (Publication 67)
http://www.clir.org/pubs/abstract/pub67.html

FREY, F. & REILLY, J. (1999). *Digital imaging for photographic collections: foundations for technical standards*. Rochester, Image Permanence Institute.
http://www.rit.edu/~661www1/sub_pages/frameset2.html

INSTITUTE OF MUSEUM & LIBRARY SERVICES. (2001). A framework of guidance for building good digital collections, November 6, 2001.
http://www.imls.gov/pubs/forumframework.htm (Formally endorsed by Digital Library Federation, 1 March 2002, http://www.diglib.org/standards/imlsframe.htm)

KENNEY, A. R. & CHAPMAN, S. (1996). *Tutorial: digital resolution requirements for replacing text-based material: methods for benchmarking image quality.* Washington, DC, Council on Library & Information Resources. (Publication 53)

KENNEY A. R. (2000). *Digital benchmarking for conversion and access* in KENNEY, A.R. & RIEGER, O. *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, VA, Research Libraries Group (RLG)

LAGOZE, C. & PAYETTE, S. (2000). Metadata: principles, practice and challenges in KENNEY, A.R. & RIEGER, O. *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, VA, Research Libraries Group (RLG)

LIBRARY OF CONGRESS. Preservation Directorate (1997). *Digitizing library collections for preservation and archiving: a handbook for curators*. Washington, DC.

LIBRARY OF CONGRESS (1999). *Quality review of document images. Internal training guide*
http://memory.loc.gov/ammem/techdocs/qintro.htm

LUPOVICI, C. & MASANÈS, J. (2000). *Metadata for the long term preservation of electronic publications*. The Hague, Koninklijke Bibliothek (NEDLIB report series 2)

OSTROW, S. (1998). *Digitizing historical pictorial collections for the Internet.* Washington, DC, Council on Library and Information Resources. (Publication 71)
http://www.clir.org/pubs/reports/pub71.html

RESEARCH LIBRARIES GROUP (2000). *Guides to quality in visual resource imaging.*
http://www.rlg.org/visguides/

RIEGER O. Y. (2000). Establishing a quality control program *in* KENNEY, A.R. & RIE GER, O. *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, VA, Research Libraries Group (RLG)

SITTS, M. K. (2000). *Handbook for digital projects: a management tool for preservation and access*. Andover, MA, Northeast Document Conservation Center.
http://www.nedcc.org/dighand.htm

SWARTZELL, A. G. (1998). Preparation of materials for digitization *in* NATIONAL PRESERVATION OFFICE/RESEARCH LIBRARIES GROUP. *Guidelines for digital imaging: papers given at the joint NPO/RLG Preservation Conference, 1998.* London
http://www.rlg.org/preserv/joint/swartzell.html

UNIVERSITY OF VIRGINIA LIBRARY. Electronic Text Center. Image scanning: a basic helpsheet. http://etext.lib.virginia.edu/helpsheets/scanimage.html

UNIVERSITY OF VIRGINIA LIBRARY. Electronic Text Center. Text scanning: a basic helpsheet. http://etext.lib.virginia.edu/helpsheets/scantext.html

WITTEN, I. H.et al (2001). Greenstone: a comprehensive open source digital library software system. http://www.nzdl.org/

3. LEGAL ASPECTS

Institutions intending to embark upon a digitization project need to be aware at the outset that they must investigate the copyright situation involved for each item that they intend to scan, and also the legal position affecting access by users to the images that will be created by the project. This will be particularly true if the institution intends to develop a business plan to market access to and copies of the images as a cost-recovery exercise. They need also to consider the issues involved in ensuring the authenticity of the digital images created if they are to serve as surrogates for the original source materials.

3.1 Copyright

Copyright means that an author's right to an original work of literature, music and art is legally protected. The time span for copyright depends on when the work was created and can differ between countries. Copyright gives the owner an exclusive right of disposition over his or her work, in other words to do or to authorize copying and public distribution or performance of any kind. Transfer of copyright must be made in written form and signed by the owner of the copyright.

If the work is (1) made by an employee as a part of his or her employment or (2) by contract defined as a work made for hire, the person or body employing the creator or giving the contract is considered the copyright owner.

Copyright has also a moral element that gives the owner the right to be mentioned, for example when the work is published, and should prevent the work being changed or corrupted.

Many archives, libraries and museums have in their custody collections which have been donated and where the copyright has passed to the heirs of the creator. If the copyright owners are unknown to the institution, it can be impossible or at least very time and money consuming to obtain copyright permission.

One of the advantages of digitization is the possibility of opening up collections and holdings for wider access, which can be in opposition to the protection of intellectual property rights. It is, therefore, recommended that the first issue to address in a digitization project or programme is the legal conditions for making digital copies. To protect institutions from possible litigation where it has proved impossible to identify the copyright holder, it is recommended that access to the digital collection is reliant upon acknowledgement of a copyright disclaimer.

3.2 Authenticity

Opinions about what authenticity stands for and how it can be achieved differ between scholars. The core issue is, of course, that a document or an image is what it purports to be, free from manipulation or corruption. In the analogue world a document is trustworthy when its identity is reliable, which means that the following facts have to be established:

The creator

Time and method of creation

Circumstances of origin

If this trustworthiness is maintained over time, the document is authentic.

When it comes to digital files the situation is more complicated. There is always a risk that something untoward happens whenever such files are transmitted across time or space, in other words when they are stored off line, when hardware and software used to process, communicate or maintain them is replaced or upgraded, or when they are sent between persons, systems and applications. For this reason, a reliable identity is not enough to guarantee the authenticity of digital files. Their integrity must also be intact. It must be established what actions have been undertaken to maintain the files, who has been involved in these actions and what tools have been used. Furthermore, it must be confirmed that no unauthorized changes (whether deliberate or accidental) have been made in the physical representation or in the intellectual content of the files.

Today, there are different kinds of methods in use to prevent or detect unauthorized changes in digital files. Digital signatures and digital seals, built on cryptographic technology, and so called “watermarks” placed inside the images as identifiers are some examples.

If digital images are accepted as replacements for source documents and are intended to serve that purpose, it must be guaranteed before conversion that:

- The source documents have no intrinsic value
- The informational content (and if needed the physical appearance) of the documents has been adequately captured
- The legal requirements are met
- The means for retrieving and preserving the digital images are in place

However, in most digitization projects and programmes source documents or analogue surrogates of them are kept and can be rescanned if the digital images are lost or corrupted.

3.3 Intellectual Property Management (See also **Section 4**)

The high costs involved in digitization suggest the need for cost recovery by the institution as a small compensation, in a manner similar to the provision of a photocopy service. Digital collections are likely to comprise surrogate copies of photographic prints, negatives, slides, transparencies, works of art, texts and book illustrations. Such collections are of high interest to a range of potential markets. While libraries and archives seldom hold copyright in the original works, the motivation for developing a self-sustaining operation is based rather upon licensing the use of images in the protection of the intellectual property of digital assets held by the institution.

The objectives of this policy can be identified as follows:

- To mark the intellectual property of the institution in an unobtrusive manner, as a trademark, using either image manipulation or "watermarking" as a technical protection in order to establish the authenticity and provenance of images.
- To convey adequately the intellectual content of original documents for scholarship, images are provided at no cost on the Internet, at a low resolution that encourages single use.
- To develop a strong business model, including e-commerce, to license multiple use of images according to a market related price schedule as a source of revenue to fund further digitization and digital preservation.
- To provide such copyright information as may be available, along with a copyright disclaimer and seeking a signed copyright indemnity as the responsibility of the user

3.4 Legal deposit

Most countries in the world have legislation that regulates the area of legal deposit for publications offered to the public. There are several motives for this, but the most important one is normally to preserve the cultural heritage. Other motives are to create a base for a national bibliography or a desire to support libraries with published material.

Originally, legal deposit legislation only covered printed publications, but during the last decade publishing in electronic form has grown enormously and forced legislators to start to define such publications also in logical terms. The expansion is not only quantitative but is linked to the fact that new producers establish themselves, and in so doing expand the boundaries of what can be published electronically. Any private individual with a computer and Internet access can take on the role of being simultaneously writer, publisher, printer and distributor of published material.

There are two obvious trends in Internet publishing:

- *convergence*, which means that different media like newspapers, radio, moving images, telephony etc. grow together and give seamless access to their respective contents

- *individualization*, which means that more and more of the information offered to the public has an individual design. So called intelligent agents, "butlers" and "know-bots" are today on the web searching for information according to personal wishes.

Before a digitization project commences it must be made clear to what extent it will be covered by legal deposit legislation.

Recommendations

Investigate the legal position in relation to making digital copies of the relevant source materials as the very first stage of the digital imaging project

Take steps to prevent unauthorised changes in the digital files created during the project

Ensure that the project complies with appropriate local legal deposit legislation.

Put in place a clear policy on access to and use of images within the completed digital project, including, if appropriate, provision of copyright disclaimer forms

Suggested reading

COUNCIL ON LIBRARY & INFORMATION RESOURCES. (2000). *Authenticity in a digital environment*. Washington, DC (Publication 92)
http://www.clir.org/pubs/abstract/pub100abst.html

DE STEFANO, R. (2000). Selection for digital conversion *in* KENNEY, A.R. & RIEGER, O. (2000) *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, VA, Research Libraries Group (RLG)

FEENSTRA, B. (2000). *Standards for the implementation of a deposit system for electronic publications*. The Hague, Koninklijke Bibliothek (NEDLIB report series, 4)

LARIVIÈRE, J. (2000). *Guidelines for legal deposit legislation*. Paris, UNESCO. http://www.ifla.org/VII/s1/gnl/legaldep1.htm

STEENBAKKERS, J. (2000). *Setting up a deposit system for electronic publications: the NEDLIB guidelines*. The Hague, Koninklijke Bibliothek (NEDLIB report series, 5)

WERF, T. van der. (2000). *The deposit system for electronic publications: a process model*. The Hague, Koninklijke Bibliothek (NEDLIB report series, 6)

4. BUDGETING

Building a digital collection is expensive and resource-intensive. Before embarking on digitization projects, some basic planning principles are offered here for projecting the costs underlying the design, implementation, and maintenance of a digital library. Management issues related to the budgeting for a digitization project include the cost of training of staff and integration of new work processes, the acquisition of equipment and provision of a suitable workspace, and the establishment of new systems of digital storage to ensure the preservation of digital heritage. Once these factors have been considered and accepted as an ongoing commitment at the policy level, detailed conceptualization can proceed.

4.1 Cost recovery

Cost recovery can offer small compensation to the high capital expenditure associated with digitization. This is usually conducted in a manner similar to the provision of a photocopy service. It is generally recommended that digitization be limited to source material in the public domain, to avoid complicated copyright litigation. On that basis, a business model can be devised to license the use of digitized material. An equitable model would offer free Internet access to low-resolution images, and escalating according to intended use, to secure for the library or archive a percentage of ultimate commercial profit. This concept may need to be negotiated in the case of parastatal institutions, where centralized accounting systems do not readily identify income-generating activities. In a digital environment, the effective pricing of such a cost recovery service will take into consideration staff time in the creation of the digital file, the cost of disk storage per megabyte, and an additional 10% archiving fee for the long-term storage, maintenance and migration of the file to new media to ensure future access. (See also **Section 3**)

A more detailed discussion of these areas of expenditure follows.

4.2 Areas of expenditure

4.2.1 Staff development

Service providers in the qualification of library and archive staff have been slow to respond in developing countries to the change in traditionally separate user groups of libraries and archives, that have been redefined in the growing use of the Internet.

A significant relationship has emerged between computer literacy and the general information literacy competence of information professionals. Basic computer literacy is a pre-requisite for the provision of an effective information service that includes Web-based resources, yet these skills are still not guaranteed in developing countries. There is a definite need to budget for improved computer literacy, from the level of basic operational skills in a Windows environment, which are readily available in the commercial sector. An annual budget commitment should provide opportunities for staff to avail themselves of professional and technical task-specific training. The tendency in developing countries for Directors to reserve such opportunities for themselves speaks on the one hand of severe financial constraint, but on the other of inadequate consideration of forward planning in preparing an annual budget. Associated is the need to budget for increased availability of computers in the workplace, and the regular upgrade of technical infrastructure, outlined in **Section 2 Technical Requirements & Implementation**.

The implementation of digital projects and subsequent integration of such projects into the institutional structure will necessitate the consideration for budget purposes, of new job descriptions and new posts. The outcome of staff development should be acknowledged and reflected in the transformation in traditional services.

4.2.2 Facilities management

A major concern of libraries and archives in developing countries is the ongoing cost of building maintenance. The need to provide adequate storage infrastructure is not diminished by digitization, since the digital surrogate does not replace the original document or artefact. The attainment of reliable and steady environmental control is often problematic, yet a cool, dust-free working environment is more critical to the rate of deterioration of microforms and digital products than for paper-based records.

Along with buildings and facilities management, is the need to secure a reliable and clean power supply, with emergency backup generators. The problems of maintaining the technical and physical infrastructure can only be met in careful monitoring and planned intervention long before disaster strikes.

Where optimal environmental conditions are not attainable for storage of digital data, serious consideration should be given to the identification of institutions committed to digital preservation, for the development of trusted digital repositories where duplicate copies may be held in safekeeping. See **Section 7 Preservation of digital content** for further details.

4.2.3 Operational expenses

A good starting point for budgeting the operational costs of digital projects is the estimation of cost per image. This should include the current unit price (per Megabyte) for disk storage, and an estimate of staff time in production processes. Guidelines on developing cost estimates towards a project proposal are given in Chapter 8 on project management. In the major budget areas of management, infrastructure and production, the greatest proportion of costs are allocated to staff costs and it is recommended that remuneration of project staff be calculated not on a fixed salary, but rather on the production incentive of a market-related unit price per record. Average production rates per hour are given as performance indicators in **Section 8 Project Management**.

While no two projects are alike, general guidelines on drawing up a budget of costs should consider the tasks in which staff will be involved, developed around the following categorization of production processes in a digital library:

- Selection and preparation of source material for digitization
- Digital conversion
- Metadata capture
- Data management

4.2.3.1 Selection and preparation of source material for digitisation

Criteria for the selection of materials for digitization can be broadly defined in the assessment of user needs, the attributes of the source material and the technical infrastructure for successful conversion.

The selection process requires considerable investment of staff time in assessing the value to users of the informational content of the source material, either as individual documents, or as a collection of documents. The volume of material for digitization and the conversion cost per page as well as the relevance to other on-line resources needs to be evaluated. The selection process may involve lengthy partnerships developed in multi-institutional initiatives for digital collection development.

The time spent in the evaluation of physical attributes of the source material needs to be calculated, first in terms of the relationship of fine detail to the physical dimensions of the documents, whether bound volumes or loose sheets, and the quality and condition of the documents. These factors have budget implications for the purchase of appropriate equipment, and in the establishment and management of preservation procedures where these are not already in place.

Once the selection process has been finalized, a further ongoing cost is that of preparation. An estimate of staff time should take into consideration the retrieval of materials for digitization and their return to the shelf. This function should therefore include the cost of preservation and conservation required to protect integrity of source materials, including documentation, microfilming, flattening, cleaning, repair of minor tears, or the disbinding of volumes and possible subsequent rebinding or protective enclosure of source material.

4.2.3.2 Digital Conversion

The cost of the technical infrastructure required for the conversion is determined by the media selected. Bound volumes may need to be scanned face-up on a planetary scanner, loose documents on a flatbed scanner. Transparent media, (slides, negatives) may be captured with a transparency adaptor on a flatbed scanner, but optimal image quality might be secured in the inclusion on the budget of a film scanner. Included in the hardware cost estimate is the maintenance contract to support maximum production. The wide range of hardware products available indicate a need to select the technical infrastructure on the basis of the image quality determined according to the Guidelines outlined in **Section 2 Technical Requirements & Implementation.**

The objective of digital conversion for cultural heritage institutions is authentic representation rather than image enhancement for desktop publishing. Image capture software is normally bundled with the capture device, and subsequent image management can be effected with high end products like Adobe PhotoShop, Corel PhotoPaint) or demonstration packages on the Web (PaintShopPro).

Once a digitization process has been chosen for the source material selected, the cost per image can be analysed on the following basis:

- *Source type:* Turning pages and repositioning of bound materials will take longer to scan than loose sheets; the large pixel dimensions in scanning oversize maps or newspapers slow production rates, and may need to be outsourced where technical infrastructure is not available.

- *Quantity:* total volume of images to be scanned.

- *Process*: direct scans or intermediate, OCR conversion to ASCII text.

- *Standard:* resolution, bit-depth, tonal range will affect resultant file-size, and ultimately, the cost of disk storage.

- *Cost per item:* Where resolution is constant, cost per item is affects by physical dimension of the source material, resulting in variations in file-size, and cost of disk storage.

4.2.3.3 Metadata Capture

The processes to make collections accessible, either in a catalogue or on the Web are determined by the selection of a metadata standard, based on the following factors:

Extent of existing collection-level description.
Need to modify metadata for various user audiences.
Compatibility to make collection visible through a Gateway.

A collection of photographs may be catalogued in MARC Standard, either at collection level or, at great cost in terms of staff time, at item level entry. The images in each album might rather be described at the item level in the Dublin Core Metadata Initiative, a much simpler set of fourteen elements. Similarly, one might need to create an EAD finding aid to provide hyper-linked hierarchical access from collections to individual documents in a digitized manuscript collection.

The cost of metadata or indexing processes are disproportionately high (60% of total cost), as they are conducted by qualified information specialists, who often need to be re-skilled in the use of new standards.

4.2.3.4 Data Management

Post capture processing includes quality control against the conversion standard selected and re-scanning where necessary. This can either be conducted on each file, or at specific capture intervals, to ensure consistency of image quality.

The creation of smaller derivative files from master tiff files can be automated, to provide low-resolution images for Web presentation.

Digital archiving comprises electronic records management functions of providing security, authenticity, and integrity for long-term preservation and access. While many document management systems will offers these features, it is important that proprietary file formats are avoided at all costs. Files should be stored in a standard file format (.TIFF, .JPEG, ASCII text) that can be migrated to a new platform as required, without loss of data and resultant costs incurred to the library or archive.

The identification of a local agent to provide ongoing software support in developing countries is important for cultural heritage institutions. Digital asset management is becoming of increasing importance to the commercial sector, and the strength of the growing market will positively affect pricing, and enhance the availability of local support.

4.2.4 Managing storage and delivery systems

The application of digital technologies in providing open access to information demands high levels of capacity in information technology. Where this capacity is lacking in developing countries, reliance on consultants must be calculated into the budget at market-related prices.

The implementation of a storage system to manage document images should enable the management of file relationships, audit trails, version control, and disposition scheduling. The selection of a suitable system requires some investigation of commercial software products for budget purposes. Software evaluation may be effectively conducted by a specialized consultant, working in conjunction with staff to identify the needs of the institution. Apart from the many useful features that many software packages offer, an additional point of budgetary consideration is the license fee, usually an annual commitment to maintenance and updating of the software. A valuable lesson drawn from an early digitization project, Project Open Book conducted at Yale University, was that once renewal of the license was discontinued, the data soon became inaccessible as the platform became obsolete.

The design of a user interface and management of the delivery system is integral to access. Staff training in HTML for Web presentation is increasingly important, and further details are provided in Chapter 6 on access. Budgeting for software is open, with solutions ranging from highly sophisticated HTML editors (Dreamweaver, Front Page), to shareware products available on the Web (Arachnophobia, Front Page Express).

The budget considerations in managing the storage and delivery system will include the software requirements outlined above, systems administration functions of server acquisition and maintenance, network infrastructure and access control (firewall), backup hardware and media (tapes, CD's etc.). Storage of backup copies and microfilm masters in off-site low temperature and low humidity storage is recommended for disaster recovery purposes.

Modest solutions to managing the storage and delivery system can be applied in developing countries. One solution can be found in hiring the services of a commercial Internet Service Provider (ISP), rather than assuming the technical challenge and ongoing costs of server maintenance. The budgetary considerations in this scenario would be a dedicated telephone connection, and a monthly subscription to the ISP. The limitations on storage allocations should be carefully negotiated in this case.

Another modest budget solution that has proved successful in the creation and management of digital data has been identified in the establishment of national or regional consortia, in which joint grant funding proposals can be sought to cover expenses. This solution offers immense additional benefits in the shared experience of staff training, the development of common policy in a nascent area of activity and additional security in collaborative data management.

Recommendations

Investigate cost recovery options of income-generating activities

Build business models to support income-generation

Reflect the transformation of traditional services in staff establishment posts

Where digital collections are vulnerable, form partnerships for the development of trusted digital repositories

Budget operational expenses as unit costs, i.e. cost per image

Delegate responsibility of data storage and delivery to a commercial ISP

Form consortia for collaborative development and shared expenses

Suggested reading

ARCHAEOLOGY DATA SERVICE. *Digital archives from excavation and fieldwork. Guide to good practice*. 2nd Ed. Section 5. (Costs of digital archiving).
http://ads.ahds.ac.uk/project/goodguides/excavation/sect54.html

ASSOCIATION OF COLLEGE & RESEARCH LIBRARIES, U.S. Information literacy competency standards for higher education
http://www.ala.org/acrl/ilcomstan.html

CHAPMAN, S . (2000). Project planning; creating a plan of work and budget, *in* SITTS, M. K. *Handbook for digital projects: a management tool for preservation and access*. Andover, MA, Northeast Document Conservation Center.
http://www.nedcc.org/dighand.htm

CONWAY, P (1996). Yale University Library's Project Open Book: preliminary research findings. *D-Lib magazine*, February 1996.
http://www.dlib.org/dlib/february96/yale/02conway.html

JONES, T (2001). An introduction to digital projects for libraries, museums and archives. http://images.library.uiuc.edu/resources/introduction.htm

PETERS, D. & PICKOVER, M. (2001). DISA: insights of an African Model for Digital Library Development. *D-Lib magazine*, 7 (11)
http://www.dlib.org/dlib/november01/peters/11peters.html

WATERS, D & WEAVER, S (1992). *The organisational phase of Project Open Book*. Washington, DC, Council on Library & Information Resources.
http://www.clir.org/pubs/reports/openbook/openbook.html

Related resources

Australian Co-operative Digitisation Project 1840-1845. Appendix 4 . Budget.
http://www.nla.gov.au/ferg/append4.html

DISA: Digital Imaging Project of South Africa. http://disa.nu.ac.za

Dublin Core Metadata Initiative http://dublincore.org/

Internet Library of Early Journals. Final report, March 1999.
http://www.bodley.ox.ac.uk/ilej/papers/fr1999/

MARC Standards http://www.loc.gov/marc/

RESEARCH LIBRARIES GROUP (1998) Worksheet for estimating digital reformatting costs. May 1998. http://www.rlg.org/preserv/RLGWorksheet.pdf

5. HUMAN RESOURCE PLANNING

Because of resource constraints, many libraries and archives in developing countries tend to be behind the digital technology curve. Service providers in the education and qualification of library and archive staff have been slow inform students of the new skills they will need to respond to the digital environment. These include not only technical skills, but proposal writing and project management skills applied to the development of technical services. The successful application digital technology is not a matter of hardware or software, but a problem of access to opportunity, which goes far beyond technology.

Directors of libraries and archives may fear that because they do not understand the technical details of digitization, they cannot effectively plan for the implementation of digitization projects. It is more important for managers to understand the impact of digitization on the organization and its goals. Three main areas of consideration are change management, capacity building, and in developing countries, the social implications of digital technologies.

5.1 Change management

Opportunities for staff development in the implementation and use of digital technologies require managerial support, often less than enthusiastic when faced with the reality of trimming budgets to support new initiatives.

Change is basically about people. It may be necessary to analyse the problems of interaction within the organisational culture for obstacles related to territoriality, a lack of informed managerial support and fear of change within the line management, including technophobic barriers to technological innovation. These issues are often underestimated.

The functional units of organization within the institution may need to be deconstructed to enable change by focusing less on procedures and more on common goals of providing an information service. It is inevitable that existing lines of authority and responsibility will be relaxed. The level of seniority that is age-related in the traditional societies of developing countries has no place in the digital arena, where individuals must be fearless of risk and change, and be self-motivated in learning the limits and opportunities of information technology and communication. In the absence of formal training in developing countries, managers can nevertheless provide leadership in seeking aptitude in these areas to empower the right people in the organization.

For example, a simple manifestation of managerial support for changing institutional cultures might be in making time available to staff, who show an aptitude, to familiarize themselves with computers. Financial assistance in the form of institutional loans for personal computers, modems etc. will serve the institution by extending the learning curve beyond office hours, while taking the threat of change out of the workplace.

5.2 Capacity building

Even when opportunities abound, people and organizations have a natural aversion to change, especially where it is perceived as daunting, complicated or costly. At the same time there is a natural human tendency to desire what others have. Capacity building is therefore effectively achieved by forming partnerships with early adapters, either institutions or individuals with experience in the use of the technology, and who in their commitment to making it work, ensure the transfer of skills and increase the chances for a successful outcome of the project.

The development of partnerships with similar cultural heritage proposals to collaborate with experienced institutions or individuals on joint initiatives can leverage human development beyond the seniority and gender constraints of the particular institutional culture.

Formal training opportunities that might be available include commercial training for the basic office environment, or short courses offered by universities and colleges, some even on-line, aimed to deliver successful technology. Most institutions began operating on the information highway by sending highly motivated delegates to intensive training courses. In the developing world, the training should be appropriate to the particular needs of operating independently with limited IT support. It has become clear that in providing grant-funded specific digitization training courses that the acceptance of such opportunities also bears with it a level of accountability. Follow up progress reports should be submitted by participants at regular intervals, both to their managers and to the course organizers or funding body. Capacity building then becomes self-motivated, if the individual is empowered to affect change in developing digital technologies.

Intensive instruction in digitization should assume a basic level of IT competency in a Windows environment, and aim instead to provide key skills for digitization:

> image capture: to capture a digital image from a physical object
>
> OCR (Optical Character Recognition): to convert imaged text into machine-readable format
>
> markup languages: standard protocols for adding metadata, e.g. HTML, XML
>
> metadata: standard schema of administrative, descriptive, structural and preservation information, e.g. Dublin Core
>
> indexing and database technologies to search and retrieve digital resources
>
> intellectual property management: the risks and responsibilities of disseminating electronic information
>
> user interface design: the interpretation of user interactions with the data
>
> web technology: encompasses basic delivery mechanisms of digital data via HTML, XML and use of search engines
>
> project management: to achieve goals within set periods of time, and within a specific timeframe

In addition new managerial skills are required in the broader areas of project management, systems implementation and increasingly, in fundraising. The reliance in developing countries on inadequate funding from the national government is broken in providing an information service to the global community. The goals and objectives of digital projects have to be clearly identified and the implementation carefully planned in order to attract grant funding. Whether digital projects are to be ultimately outsourced or production conducted in-house, the need to develop both technical and managerial skills is essential for effective quality control.

5.3 The social contract

An important component of capacity building in developing countries is the opportunity provided to create new job opportunities for the people of the country. Partnerships that support human development are preferable to those that offer quicker, and often cheaper off-site conversion, but which ignore the social upliftment of job creation.

Human resource development is essentially aimed at breaking the digital divide. The Internet provides global information sharing, changing the way in which users interact with information resources. The boundary to knowledge dissemination is no longer owned and distributed in an unequal hierarchy from the dispenser (the librarian or archivist) to the user. The value of information is no longer vested in its ownership, but in the trusted value-added services of skilled information professionals to guide and direct the user in turning the overwhelming volume of electronic information into knowledge.

Recommendations

Provide leadership in embracing change

Empower the right people

Form partnerships with early adaptors for capacity building

Develop new technical and managerial skills

Create new opportunities for social upliftment

Suggested reading

ARMS, W. Y. (2000). Digital libraries for digital education: editorial. *D-Lib magazine*, 6 (10), 2. http://www.dlib.org/dlib/october00/10editorial.html

GARROD, P. & SIDGREAVES, I. (1997). Skills for new information professionals: the SKIP Project. Plymouth, Academic Services, University of Plymouth. http://www.ukoln.ac.uk/services/elib/papers/other/skip/

HASTINGS, S. K. (2000). Digital image managers: a museum/university collaboration. *First Monday*, 5 (6), 9pp

Related resources

Canadian Heritage Information Network. Capture your collections. http://www.chin.gc.ca/Resources/Digitization/English/index.html

CORNELL UNIVERSITY. DEPARTMENT OF PRESERVATION & CONSERVATION. Moving theory into practice: Digital Imaging Tutorial http://www.library.cornell.edu/preservation/publications.html

HUMANITIES ADVANCED TECHNOLOGY & INFORMATION INSTITUTE (HATII), University of Glasgow. Digitization summer schools for cultural heritage professionals.

http://www.hatii.arts.gla.ac.uk/SumProg/

TOWNSHEND, S. et al. (2000). *Digitising history: a guide to creating digital resources from historical documents*. Colchester, Arts and Humanities Data Service. (Also at http://hds.essex.ac.uk/g2gp/digitising_history/index.asp)

6. DEVELOPMENT AND MAINTENANCE OF WEB INTERFACES

The rush to get on the Web

The digitization of our cultural heritage brings together various sectors of the global community in an unprecedented manner. The user groups of traditional library, archive and museum structures have been redefined by the growing use of the Internet. Scholars are creating or using electronic resources to further their research; distance-learning models prompt teachers to gather Web resources in an online learning environment and publishers are integrating print with digital editions to reach wider audiences. The support of computer and information specialists in the application of new technologies to develop and manage online information is increasingly sustained by libraries and archives seeking to improve access to digital information that represents rich resources of culture and scholarship. The unique properties of the digital medium give visual form to cultural heritage information. The interactive techniques of the Internet that appeal to the cognitive senses provide a new opportunity for libraries and archives to develop a global user community.

The organizational issues involved in creating and maintaining online information resources can be grouped into four areas:

- Developing digital content
- Building a Web team
- Website production and management
- Introducing Web-based services

6.1 Developing digital content

The previous **Sections 1. Selection** and **2. Technical Requirements & Implementation** offer guidelines to create an enabling environment for the development of digital content.

6.2 Building a Web team

A useful starting point for creating a website is for staff to visit sites of similar institutions, to see what features work well to make the pages interesting to users. The variety of tasks involved in three identified areas of activity in creating a website suggest the formation of a team to draw upon different human resources in the institution and to contribute a range of skills.

- The authoring and content management for the Website is a general institutional function. Different sections of the website can be apportioned to nominated individuals to update regularly, under the co-ordination of a Webmaster.

- The systems administration functions, which include site maintenance, access management and network connectivity are a responsibility best delegated to a designated systems administrator.

- The third area of activity of the Web team is that of Website management. This function includes developing an institutional policy on the role of the website in marketing and promotion; the development of corporate branding in the graphical representation of the institution; and the development of a consistent design style. Graphic design skills are not normally represented in libraries and archives, and it may be necessary to appoint a specialist consultant in this capacity.

General computer literacy, and a working knowledge of HTML (Hypertext Markup Language) are required for Website development work. The importance of HTML for librarians and archivists, and increasingly of XML (Extensible Markup Language), is growing rapidly. Although these skills may still be rare in developing countries, the design and maintenance of interesting interfaces to Web-based content does not require computer-programming experience. It is likely that designated staff will be self-taught in the various activities, and will take on the responsibility in addition to their normal tasks.

6.3 Website production and management

There are few tested principles when producing websites for libraries and archives. The real challenge is to design a website that has reliable and up to date content and a user interface that is easy and intuitive to accommodate the needs of different users, both scholarly and the general public. Some basic technical guidelines exist to ensure a consistent, high standard is maintained in the production and management of a website.

6.3.1 Website Production Guidelines

These guidelines are designed to meet the following needs:

- To assist staff with little previous experience to develop responsibility for the website
- To provide guidance to specialist design consultants
- To evaluate the products of external consultants.

6.3.1.1 File and folder structure

- Create a folder for each section of the website
- Store graphics files in separate folders from the HTML files
    - Common graphics used across the site, such as those used for the program template should be stored in a graphics folder in the root directory.
    - Images unique to a certain section can be stored in either:
      (a) a graphics folder within the relevant directory, or
      (b) a labeled subfolder of the central graphics directory
- Store downloadable documents (i.e. zipped Office documents, PDF documents) in separate folders from the HTML files.

6.3.1.1 File naming conventions

- Select a single file extension to use for all HTML files across the website - either .htm or .html or .shtml (depending if Server Side Include (SSI) is used)
- Use lower case for file names
- Do not include the space character, “&”, “*” “/” “\” in the name
- Use meaningful file names for URLs
- Keep file names 8 characters or less

6.3.1.2 Page layout and design

- Use a standard design template for the HTML content input by various contributors. Include in the template the graphical layout, logos and contact information for the institution, a predetermined colour scheme, text formatting, hierarchy of headings, and a set of bullets and lines.
- Hard code the page width to within a standard screen setting, e.g 800 x 600 to avoid horizontal scrolling.
- Select 216 “web-safe” colours.
- Do not use frames to ensure accessibility to the visually impaired.

6.3.1.3 Web-ready Graphics

- Limit the use and size of graphics in the interest of download times and text-only browsers.
- Use only standard file formats, .GIF for simple pictures and graphics, and .JPEG for complex colour images.
- Code the image dimensions in the <img src> tag of the HTML as the browser can then format the page before loading the graphics. Use the <alt> tag to describe the image for the visually impaired.

6.3.1.4 Minimum requirements

- Each page should have an unique title, preferably a meaningful expansion of the filename
- Include Metatags for 5-10 Keywords and a Description of 250 characteristics to identify the content to a Web search engine
- Provide links to the Homepage and other main sections
- Include a highlights section that is regularly updated with news and topical events
- Provide a mechanism for feedback, either mailto: or a feedback submission form, such as that offered by Active Feedback.

6.3.1.5 Site maintenance

- Test the website development across platforms, and across browsers to ensure that it views well for most users.
- Check the site monthly for broken internal links, broken external links and orphan files that are not linked from anywhere.
- Subscribe to a commercial service, e.g. Netmechanic, to automate error checking, or run a manual check with specific tools, e.g. “Check links sitewide” in Dreamweaver.

6.4 Introducing Web-based services

The ultimate challenge for libraries, archives and museums lies in applying digital technologies to the development of Web-based services. Developing digital content requires image capture, the description and indexing of images, and the management of access to digitized collections. Guidelines on image capture are presented in **Section 2 Technical Requirements & Implementation**. The description and indexing of images requires a new approach to the cataloguing or archival description methods traditionally employed in libraries and archives.

6.4.1 Indexing digital content

Collections have been traditionally documented in various ways by accession registers, card catalogues and more recently in databases, which offer the advantage of automated search functionality. The accessibility of collections in a Web environment relies therefore on the ability to search from a remote access point. The creation of digital records demands new methods of knowledge organization and data management in a digital, distributed, multimedia environment. Digitization, and the automation of the associated record describing the digital object, by metadata capture, cataloguing or encoded archival description (EAD), have become tools for interaction with Web-based content.

International technical standards are emerging to ensure interoperability across the Internet, in a manner similar to the z39.50 protocol for interoperability between databases. Current Internet standards models are available, such as those offered by W3C (World Wide Web Consortium) and the IETF (Internet Engineering Task Force). These standards include various versions of HTML from HTML 1.0 to HTML 4.0 and CSS 2, and W3C is now encouraging the use of XML, providing for the current development of schema, based on important metadata standards such as RDF (Resource Description Framework) and Dublin Core.

Standards allow more freedom, interoperability and accessibility for users. They also avoid reliance on a software vendor to maintain digital collections. Schemas enable knowledge structuring and electronic data management at collection, document or record level in digital libraries. Used as interactive information services on the Internet they have an increased potential to support the description, discovery and retrieval of heterogeneous and distributed information resources.

6.4.2 Access management

Access may be achieved by means of Websites, or CD-ROM, or both. The advantage of CD-ROMs is that they fulfill the legal requirement in some countries for physical evidence. In developing countries, where the networking infrastructure and bandwidth are limited, access to information can be greatly assisted by the use and distribution of CDs.

The hosting of a Website in developing countries suffers from many limitations. Where an internal server cannot be maintained, the website may be successfully hosted for a fee by a local commercial service provider. Websites may be developed to serve a single institution, or as portal sites linking related information resources. Portal sites carry an additional responsibility for long-term preservation of linking mechanisms. A collaborative digital repository will serve the joint responsibility of related digital content.

Access management issues are essentially those of electronic records management. These functions may be categorised as follows:

- *To ensure that records can be exported from the software application.*

The ability to move data to new software versions across time will ensure long-term preservation. The use of standard file formats (.TIFF, ASCII text) will provide software independence.

- *To preserve security, authenticity, integrity.*

Access policies and permissions are intended to limit undue manipulation and possible corruption of archived electronic records. Any changes made to the file are recorded for future reference, and assist in maintaining information integrity. Authenticity requirements form part of provenance, by maintaining records in their original format, and managing groups of records according to their security markings.

- *To associate contextual and structural metadata.*

The association of contextual and structural metadata with the image as a single digital object ensures that all elements are displayed as a unit on retrieval.

- *To manage appraisal audit trails.*

Appraisal audit trails follow disposition schedules set at creation. Schedules are normally set chronologically, or conditionally. Digital content management entails the regular review of disposition decisions pending, and the selection to override disposition schedules for permanent preservation. Finally, list of records for transfer or destruction should be maintained.

Recommendations

Identify employees who have a little technical, managerial and editorial skill and develop around them a Web team whose skills complement one another

Develop an institutional policy on the intended use of the website

Adopt technical guidelines to ensure a consistent, high standard in the production and management of the website

Acquire new skills in the application of metadata standards for indexing of digital content

Embrace principles of electronic records management to secure long term access to digital content.

Suggested reading

DAWSON, A. (2000). *The Internet for library & information service professionals.* 3rd ed. London, Aslib.

Digital Imaging Group DIG35. Metadata specification
http://www.digitalimaging.org/links_metadata-digital-images.html

Digital Imaging Group DIG35. Metadata specification MARC / AACR2

http://lcweb.loc.gov/marc/umb/um01to06.html

European Union. DLM Forum. Guidelines for using electronic information
http://europa.eu.int/ISPO/dlm/documents/guidelines.html

INTERNATIONAL COUNCIL ON ARCHIVES (1999). *ISAD (G): General International Standard Archival Description.* 2nd Edition. Paris**.**
http://www.ica.org/eng/mb/com/cds/descriptivestandards.html

UNESCO. Communication and Information Sector (2001). *Website production guidelines*. Paris.

Related resources

Active Feedback Online Feedback Management Solutions http://www.activefeedback.com/af/

Dublin Core Metadata Initiative http://dublincore.org/

Encoded Archival Description http://www.loc.gov/ead/

Hypertext Markup Language http://www.w3.org/MarkUp/

IETF (Internet Engineering Task Force) http://www.ietf.org/

Internet manual for Librarians http://www.epnet.com/lrc_ft/interman.html

MARC / AACR2 http://lcweb.loc.gov/marc/umb/um01to06.html

MICROSOFT CORPORATION. Improving Web Site Usability and Appeal
http://msdn.microsoft.com/workshop/management/planning/improvingsiteusa.asp

Resource Description Framework (RDF) http://www.w3.org/RDF/

World Wide Web Consortium http://www.w3.org/

XML (Extensible Markup Language) http://www.w3.org/TR/REC-xml

7. PRESERVATION OF DIGITAL CONTENT

Digital technologies offer a new preservation paradigm. They offer the opportunity of preserving the original by providing access to the digital surrogate and of separating the informational content from the degradation of the physical medium. In addition, digital technologies liberate preservation management from the constraints of poor storage environments typical of the tropical and sub-tropical climates in which many developing countries are located.

The preservation advantage of digital content lies in the possibility to create and store multiple copies in various locations without informational loss. In an electronic environment, the physical location of the document becomes irrelevant, and remote storage options are a normal feature of backup procedures, rather than the distressing relegation of traditional collections. Multiple copies stored off-site increase the rate of preservation of materials threatened by environmental and operational shortcomings of the institution.

Ultimately, the higher optical quality of the digital surrogate together with the convenience of online access satisfies the research requirements of the user, and results in reduced handling of original material.

7.1 Preservation challenges

Librarians and archivists are primarily concerned with the intellectual issues of preserving integrity and authenticity of the information as recorded in their collections while providing long-term access to both physical and electronic records.

Digitization practices should therefore be integrated with existing preservation services to ensure the physical preservation of objects is not overlooked in their treatment prior to scanning and protective enclosure following scanning to extend the life of the original materials.

The preservation of digital information presents new challenges:

7.1.1 Technical support

The concept of long-term access is not supported by the IT industry, where dynamic market forces work against standardization. Concerns for media preservation persist as outdated media quickly become obsolete, but performance improvements in developing storage media such as tapes, disks and CD-ROMs, encourage digital preservation by media migration.

7.1.2 Technology obsolescence

The greatest challenge lies in technology preservation, which entails not only the migration of the data itself, but also the migration and emulation of the technology platform, including devices and the data formats in which the information was created to ensure that it will continue to be accessible on emerging new platforms.

Just as in the physical environment, there is no complete solution. Some recommended strategies to meet these new challenges are offered:

- Policy development at the point of capture.
- Application of international standards and best practice
- Application of non-proprietary models
- Persistent archive management
- Collaboration on developing trusted digital repositories

7.2 Policy development at the point of capture

Decisions regarding digital preservation need to be made at the outset, for conformity of capture and the management of digital objects. The policy should formulate goals of the digitization project, identify materials, set selection criteria, define the means of access to digitized collections, set standards for image and metadata capture and for preservation of the original materials and state the institutional commitment to the long terms preservation of digital content.

7.3 International standards

Establishing digitization procedures according to appropriate standards for managing electronic information facilitates access, use and long-term preservation. The role of standards has been critical to interoperability and to automating processes. Adherence to standards can facilitate preservation in managing the transfer of information between hardware and software platforms as new technologies evolve.

Where possible, one should adhere to established, internationally accepted standards and where such standards do not yet exist, it is advisable to adopt international best practice.

7.4 Non-proprietary models

Platform independence is an effective strategy to prevent technical obsolescence, and is achieved in developing practices that support open systems and non-proprietary IT standards to ensure long-term access. It is especially important for institutions in developing countries to avoid the expense of annual software license fees that might render the data inaccessible where these fees are not sustainable. Instead, the Extensible Markup Language (XML) offers a non-proprietary technology neutral interchange protocol.

File format of archival master image files should also be interoperable, like .TIFF and .JPEG; and metadata schema should require no specific software to be intelligible, i.e. ASCII text, rendered in XML.

7.5 Persistent archive management

The context of document creation is easily altered in a digital environment. This context must be carefully described and stored if the document is to be retained as a record. The archival concept of records management has supported the development of digital libraries, in describing the content of the information, its structure in relation to other records, and the context of its creation, storage and migration. Information integrity is achieved in the management of audit trails, version control, access policies, disposition scheduling, and maintaining file relationships.

A persistent archive is built on an infrastructure to organize and store large collections of electronic records, to support information discovery, and to create trusted digital repositories. The information architecture of a persistent archive integrates both the digital objects and the metadata required to access the digital objects, encapsulated together as a collection of electronic records. Persistence is achieved in the assignment of relevant administrative, descriptive, structural and preservation metadata to all digital objects and to the organization of the collection.

Standardization of the archive format for persistent archive management is developing around the imminent ISO standard, the Open Archival Information System Reference Model (OAIS), developed by the Consultative Committee for Space Data Systems (CCSDS). Further work is recommended on the development of the standard to include preservation description information better to support the preservation function.

7.6 Trusted Digital Repository

The preservation strategies mentioned above culminate in the current consideration of assigning responsibility to designated repositories for the long-term maintenance of digital resources, as well as for making them available over time to the communities of users agreed upon by the depositor and the repository.

Digital preservation is addressed in concept as long-term maintenance of data and access through time and changing technology. The attributes of a trusted digital repository are identified to assure the digital library community that certified institutions are committed to the long-term management of digital resources. Currently the community is basing systems and procedures on the OAIS Reference Model.

Recommendations

Associate preservation and access as organizational objectives

Set digital preservation policies before you begin scanning

Adhere to international standards and adopt current best practice

Avoid dependence on proprietary software

Assign administrative, descriptive, structural and preservation metadata to all digital objects

Identify a trusted digital repository committed to the long- term management of your digital resources

Suggested reading

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS) (2001). *Reference model for an open archival information system (OAIS). Red Book. Issue 2* (No. CCSDS 650.0-R-2). Washington, DC, National Aeronautics and Space Administration.
http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf

DEEGAN, M. & TANNER, S. (2002). *Digital futures: strategies for the information age.* London, Library Association.

DOLLAR, C. (2000). Electronic archiving: requirements, principles, strategy and best practices. in *PDA/FDA Conference on Technical Implementation*, Philadelphia, PA, Cohasset Associates.

GOULD, S. & EBDON, R. (1999). *Survey on digitisation and preservation*. The Hague, International Federation of Library Associations and Institutions (IFLA).

HEDSTROM, M. & MONTGOMERY, S. (1998). Digital preservation needs and requirements in RLG member institutions. http://www.thames.rlg.org/preserv/digpres.html

HODGE, G. & CARROLL, B. (1999). Digital electronic archiving: the state of the art and the state of the practice: a report to the International Council for Scientific and Technical Information and CENDI. http://www.dtic.mil/cendi/proj_dig_elec_arch.html

JONES, M. & BEAGRIE, N. (2001). *Preservation management of digital materials*. London, British Library. http://www.jisc.ac.uk/dner/preservation/workbook)

MOORE, R. et al. (2000). Collection-based persistent archives; part 1. *D-Lib magazine*, 6 (3) http://www.dlib.org/dlib/march00/moore/03moore-pt1.html; part 2. *D-Lib magazine*, 6 (4) http://www.dlib.org/dlib/april00/moore/04moore-pt2.html

Open Archives Initiative (OAI). http://www.openarchives.org/

RESEARCH LIBRARIES GROUP (2001). Attributes of a trusted digital repository: meeting the needs of research resources. http://www.rlg.org/longterm/attributes01.pdf

ROSS, S. & GOW, A. (1999). *Digital archaeology: the recovery of digital materials at* risk. London, British Library Research & Innovation Centre. (Report 108)

ROTHENBERG, J. (1999). *Avoiding technological quicksand: finding a viable technical foundation for digital preservation.* Washington, DC, Council on Library and Information Resources (Publication 77) http://www.clir.org/pubs/abstract/pub77.html

ROTHENBERG, J. (2000). *An experiment in using emulation to preserve digital publications*. The Hague, Koninklijke Bibliothek. (NEDLIB report series, 1)

WATERS, D. & GARRETT, J. (1996). *Preserving digital information: report of the task force on archiving digital information*. Washington, DC, Council for Library and Information Resources.(Publication 63) http://www.clir.org/pubs/abstract/pub63.html

Related resources

CORNELL UNIVERSITY. Project Prism: information integrity in distributed digital libraries http://prism.cornell.edu/main.htm

INTERPARES Project (International Research on Permanent Authentic Records in Electronic Systems. http://www.interpares.org/

LIBRARY OF CONGRESS. Preservation Digital Reformatting Program.
http://lcweb.loc.gov/preserv/prd/presdig/presintro.html

NEDLIB (Networked European Deposit Library). http://www.kb.nl/coop/nedlib/

8. PROJECT MANAGEMENT

The conceptualization of a digitization project is a process of consultation between users and information providers. It is important to structure that process by involving the advisory board, staff, scholars and other stakeholders in preparing a pilot project proposal. This initial conceptualization process can become protracted and therefore very expensive unless it is conducted within a set timeframe.

The conceptualization process will firstly analyse the situation by identifying the need of the institution or group of institutions, and formulate ideas to meet that need. The project design stage will paint a broad vision to influence that need, and finally, the project plan will define the steps required to achieve the vision.

This Section will offer some guidelines on structuring the process of designing a digitization project by developing consensus around a project proposal, by drawing up realistic cost estimates and in effectively managing the project by breaking the tasks into manageable pieces as links of a chain.

8.1 Proposal writing

A well-formulated proposal resulting from a consultative conceptualisation process will clarify digitisation selection policy decisions that form the basis for operational decisions in implementation. See **Section 1 Selection**. An institution will keep referring back to the decisions made in terms of institutional commitment, managerial support, the comparative value of collections, the identified criteria for the selection of material to be digitised, technical infrastructure and staff training. The proposal may also assist in securing funding for the project. A proposal might have the following draft format:

8.1.1 Introduction

Brief outline of the project background
Participant analysis
Joint venture agreement

8.1.2 Vision and mission

- *Development goal*
  Outline in a single sentence, the long-term benefits to which the project will contribute.
- *Immediate objectives*
  Specify changes in participants or in their context that will be achieved by the end of the project. Identify your users, and what they should be able to do when the work is completed, that they cannot do now

8.1.3 Needs assessment

- Existing policies and practices (in terms of preservation and access)
- Identified shortcomings
- Project outputs:
  How it is intended to meet objectives in the specific digital products, on-line services and skills that the project will provide

8.1.4 Activities

Steps in which staff will engage to achieve project outputs. These might include:

- General computer literacy training
- Training in digital conversion techniques

- Digital conversion of source material
- Collating distributed source material
- Modifying cataloguing procedures for indexing and metadata capture
- Web design and publishing
- Negotiating copyright

8.1.5 Performance Indicators

- Collect evidence to inform the project of progress towards outputs
  This might take the form of a certification, or of breaking down each task into time units: e.g. high production levels on a flat bed scanner might achieve 90 scans per hour. Heavy, large or fragile source material might slow production to 30 scans per hour. The rate of metadata capture is dependent upon the complexity of the schema used; unqualified Dublin Core records can be input at an average of 15 per hour
- Calculate average production rates per hour as performance indicators towards remuneration of contract staff per record, rather than on a fixed salary

8.1.6 Responsible people

- Establish skills, experience and capabilities required to carry out each activity
- Allocate responsibility for each activity to a single person, who is accountable to the project for carrying out the task according to plan
- Actual implementation may be delegated to others

8.1.7 Time-frame

- Identify the beginning, the duration and the date by which each activity should be completed
- Link activities which one can only begin on completion of the previous activity. e.g. scanning and OCR
- Set a timeframe to implement and complete the entire project
- Limit project timeframe of a pilot project to two years. Small projects may require a shorter timeframe
- Limit the conceptualization phase
- Average lifespan of hardware is two years, of software five years. If the project timeframe is too long, maintenance disruption and software migration result in delays.
- Digital conversion targets of 20,000 pages per year are realistic, with consideration of production loss in the initial 6-month startup period.

8.2 Cost estimates

The proposal should reflect the itemised budget as an indication that due consideration has been given to the feasibility of the project. The cost estimate will form the basis for a grant funding application, while the breakdown of costs into operational expenses, organisational costs and staffing costs will assist in managing the expenditure over the duration of the project.

8.2.1 Operational costs

- *Materials*
  Stationery, archival quality boxes, boards and preservation storage materials, printer cartridges, computer software
- *Equipment*
  Computers, scanners, printers, disk storage
- *Transport*
  Transport to project meetings

- *Services*
  Maintenance contracts, conservation of originals, staff training, expert consultancy fees, couriers for source materials, accommodation and catering for meetings/workshops

8.2.2 Organizational costs

- *Management*
  Project manager salary, travel to meetings with donors, attending workshops and conferences.
- *Administration*
  Partial compensation for institutional administrative staff and services of host institution.
- *Organizational development*
  Strategic planning meetings, team-building exercises, project reviews
- *Overheads*
  Office space rental, refurbishment, office furniture, cleaning

8.2.3 Staffing costs

- *Full-time staff*
  Can be called upon to take up new tasks and responsibilities. Although familiar with the operations within the institution, they are not always flexible due to other commitments. Their time may also be more expensive than other options, including salary, benefits and overheads.
- *Part-time staff*
  Can contribute specialist capabilities, and hired for flexible time periods to assist on the project. They may have other commitments outside the project, limiting their ability to work additional time when required.
- *Contract staff*
  Recruited for fixed period, as long as their capabilities are required on the project, or on a phase of the project. Their specialist skills are however, not absorbed into the organization, and their availability may disrupt project scheduling.
- *Consultants*
  Bring specialist capabilities to specific activities, and are paid only for the work they contribute. The management of consultants is challenging, as much of their input takes place outside of the project office, and the contract must be closely specified for a successful outcome.

8.3 Managing the digitisation cycle

The project manager should possess an awareness of the whole process, including the role of the digital learning environment, technical capture parameters, indexing and retrieval, electronic publishing, intellectual property management, digital archiving and preservation. Collaboration with partner institutions, IT specialists and vendors, will also form an important part of project management.

A project planning matrix will allow for effective management by identifying the need to consider the issues involved in the following operational tasks, tasks that form the links of chain in the production process:

8.3.1 Source material

- Handling on a flatbed scanner to prevent damage to source documents
- Disbinding or scanning face-up on a planetary scanner
- Re-housing once surrogate copy has been made

- Preservation treatment prior to scanning and/or protective enclosures following scanning

8.3.2 Data management

- Where will the data be stored, and who will manage it?
- Establish a data architecture to accommodate digital objects for delivery and collections management functions

8.3.3 Imaging standards

- What resolution, bit-depth, tonal range, etc. (see **Section 2**) meet functional and aesthetic requirements?

8.3.4 Extent of metadata

- Decide on objectives in catalogue enrichment, or
- collection-level or
- item-level metadata suitable for discovery, use and management

8.3.5 Reformatting as publishing

- Decide on objectives in creation of facsimile reprints, or print on-demand
- Full text databases
- e-Books
- Contribution to collaborative collections/union catalogues

8.3.5 Delivery systems

- Web site design and maintenance
- Navigation and display
- Programming scripts for maximum automation of work processes
- Systems security and authorization

Once these decisions have been made, the pilot project aims to test and evaluate the feasibility of introducing digital technologies into the institutional workflow.

8.4 Managing the workflow

The co-ordination of the workflow is achieved in three ways:

- *In the supervision of a quality control programme*
  The quality control function sets a consistent standard of image capture, and tracks the status of processing
- *In the regular documentation of progress at established reporting intervals*
  Monthly reports introduce a level of accountability to the project team
- *In the establishment of a tracking system*
  The tracking system will provide a useful tool for project review.
  It should co-ordinate and record the workflow in a database, to reflect:
    - Beginning and end date of each activity
    - The processing steps of each record created, and by whom, e.g. date of capture, indexing, quality control, Web publishing.
    - Further management metadata elements should document the capture environment, the change history, file path and digital preservation record.

The expectations of decision makers formulated in the conceptualisation phase may differ from the reality of project management. A successful outcome requires the periodic review of project goals, based on the data gathered around the workflow co-ordination.

Recommendations

Prepare a project proposal to develop consensus around the digitization project

Prepare an itemized budget of organizational, operational and staffing costs to assess the feasibility of the project

Remunerate contract staff on the production incentive of a unit price per record, using known performance indicators

Develop a planning matrix to manage the operational tasks

Establish a tracking system to monitor and report on production

Suggested reading

CONWAY, P. (2001). Project management*,* in *Preservation options in a digital world: to film or to scan*. Andover, MA, North East Document Conservation Center.

PETERS, D. & PICKOVER, M. (2001). DISA: insights of an African Model for Digital Library Development. *D-Lib magazine*, 7 (11)
http://www.dlib.org/dlib/november01/peters/11peters.html

SITTS, M. K. (2000). *Handbook for digital projects: a management tool for preservation and access*. Andover, MA, Northeast Document Conservation Center.
http://www.nedcc.org/digital/dighome.htm

Related resources

ASSOCIATION OF COLLEGE & RESEARCH LIBRARIES, U.S. Information literacy competency standards for higher education
http://www.ala.org/acrl/ilcomstan.html

Colorado Digitisation Project. Digital Toolbox.
http://coloradodigital.coalliance.org/toolbox.html

Digital Project Management, New School University
http://www.nootrope.net/newschool2.html

HARVARD UNIVERSITY LIBRARY. Selection for digitization. A decision-making matrix.
http://preserve.harvard.edu/bibliographies/matrix.pdf

UNIVERSITY OF CALIFORNIA, LOS ANGELES (UCLA). Digital projects. Project Management.
http://digital.library.ucla.edu/about/estimating/projectmanagement.html