good practices handbook

good practices handbook

version 1.3

edited by the Minerva Working Group 6
Identification of good practices
and competence centres

3 March 2004

Minerva Working Group 6
Identification of good practices and competence centres

Coordinator
Borje Justrell (Riksarkivet, Sweden)

good practices handbook

version 1.3 (3 March 2004)

edited by the Minerva Working Group 6
Identification of good practices and competence centres

General co-ordination
Rossella Caffo (Minerva Project Manager),
Borje Justrell (Coordinator of Good Practices Working Group)
Antonella Fresa (Minerva Technical Coordinator)

European Editorial Committee
David Dawson, Karl-Magnus Drake, Borje Justrell,
Muriel Foulonneau, Antonella Fresa, Anna Maria Tammaro

Secretariat
Marzia Piccininno (Ministero per i beni e le attività culturali, Italy)

Texts
Ciaran Clissmann (Pintail Ltd.), Karl-Magnus Drake
(Riksarkivet, Sweden), Borje Justrell (Riksarkivet, Sweden),
Anna Maria Tammaro (University of Parma. Italy)

Web version
Maria Teresa Natale and Andrea Tempera
http://www.minervaeurope.org/publications/goodpactices1_3.htm

Design
due_pavese

Acknowledgements
This handbook had not been more then a good idea without Ciaran Clissman. Besides all his knowledge about structuring information, his way of putting good practice into proper English has been an invaluable contribution to this work. Remaining peculiarities in the text should be blamed only on the editors, because we have not always followed Ciaran Clissman´s advices.

We will also acknowledge the members of the Minerva project and the National Representative Group who have shared their experience and great knowledge with us. Last but not least, a special thanks to the Italian national Working group on good practices for their great interest in the handbook and their valuable contributions to its content.

table of contents

introduction

1.1 Document Overview

This document is a result of the Minerva project's good practices working group. It presents a practical handbook to the establishment, execution and management of digitisation projects, with particular focus on the cultural area (libraries, museums, archives). The target audience of this handbook is teams within and across cultural institutions who are contemplating, or are already executing, digitisation projects. The document reflects the outcome of the work carried out by the Minerva project, including the substantial research represented by the national questionnaires completed in connection with the National Representatives Group (NRG) meeting in Alicante, May 2002.

1.2 Document Structure

This document is composed by two chapters:

- Background
- Practical Guidelines

Background - This reviews the relevant aspects of the Minerva project, and states the role of this document in the overall progress of the project. It also covers the work carried out to date, in order that the reader shall have a clear picture of the context in which this document should be considered.

Practical Guidelines - The most important practical lessons learnt and information collected by the Minerva project best practice team are presented. This focuses on a significant number of practical 'rules of thumb' which should be considered by organisations which are establishing, executing or managing digitisation projects in the cultural sphere. The guidelines are divided into the following areas, each of which reflects a stage in the life-cycle of a digitisation project:

- Digitisation Project Planning
- Selecting Source Material for Digitisation
- Preparation for Digitisation
- Handling of Originals
- The Digitisation Process
- Preservation of the Digital Master Material
- Meta-data
- Publication
- IPR and Copyright
- Managing Digital Projects

The guidelines are presented in a pragmatic manner, aimed at the hands-on project team.

Complementary and supporting information are available on the Internet at the Minerva Web site

www.minervaeurope.org

On this Web site you can find relevant references to examples of good practices (**http://www.minervaeurope.org/listgoodpract.htm**), competence centres (**http://www.minervaeurope.org/competencentre.htm**) and role models which are being carried out in the European cultural field, as well as by global links to appropriate and useful online resources are provided.

It may be noted that there are several other sources of guidelines on digitisation and the creation of digital cultural content. The most important ones are noted in the selected list of digitisation guidelines (**http://www.minervaeurope.org/guidelines.htm**) which is available on the Minerva Web site as well.

It is also important to mention that in depth, technical issues have been intentionally removed from this Handbook, in order to keep is as simple and short as possible. For complementary technical information, including standards, please refer to the *Minerva Technical Guidelines* (**http://www.minervaeurope.org/structure/workinggroups/servprov/docindex.htm**), which are also available on the Minerva Web site.

2.1 The Lund Principles

On 4th of April 2001, representatives and experts from the European Commission and Member States met at Lund in Sweden (under the Swedish Presidency) to discuss how to coordinate and add value to national digitisation programmes, at a European level. The meeting resulted in the publication of a set of general principles to govern public digitisation initiatives and their coordination. These principles, called the Lund Principles, were transformed into the Lund Action Plan, which establishes a list of actions to be carried out by Member States, by the Commission, and by Member States and the Commission jointly, to improve the digitisation landscape across Europe.

2.2 The Minerva Project

This document is an output of the Minerva project, which was established in 2002 under the leadership of the Italian Ministry of Culture (IST contract 2001-35461). The project includes representatives of the relevant government ministries or central state agencies from many EU Member States, with the common objective of promoting a shared approach and methodology for the digitisation of European cultural material. The project recognises the unique value of the European cultural heritage, and the strategic role which it can play in the growing digital content industry in Europe. It also recognises the value of coordination of the efforts of national governments and cultural organisations, in order to increase the level of synthesis and synergy between and among digitisation initiatives.

The Minerva project has a number of focused working groups within the overall consortium. Each working group is made up of experts nominated by the project partners, working together on a particular aspect of the project objectives. The objectives of each working group are described on the project Web site at **http://www.minervaeurope.org/structure/workinggroups.htm**. The working group structure allows the project to examine a number of the most important areas of the digitisation sphere, in parallel.

The following working groups exist within the project:

- Benchmarking framework;
- Interoperability and service provision;
- Inventories, discovery of digitised content, multilingualism issues;
- Identification of user needs, content and quality framework for common access points;
- Identification of good practices and competence centres.

The activities of the working groups include meetings, public workshops, publications (such as this handbook), international coordination and cooperation, etc.

3.1 Introduction

This chapter presents the core of the handbook. It provides practical guidelines for organisations and bodies contemplating, or involved in, digitisation projects. The emphasis is on the cultural sphere; however, the material is to a large degree relevant to other spheres (e.g. tourism, general document management).

The material in this chapter is broken down in accordance with the stages in the digitisation life-cycle. This means that a reader can easily identify material which is relevant to his work, regardless of how far his own project has progressed. It is anticipated that many users of this handbook will be at the first stage of the project (planning); however, at least some of the material provided here should be of value to any digitisation project.

The digitisation life-cycle stages are used as the basis for the breaking down of the guidelines and mirrored in the table of contents.

Each guideline description is structured in the following way:

- A Guideline Title
- An Issue Definition, which sets the scene for the guideline and/or introduces the problem which the guidelines addresses
- A set of Pragmatic Suggestions which aim to facilitate the relevant aspect of setting up or executing a digitisation project
- Notes or Commentary, where any additional information is provided. This is sometimes empty

Supporting references are available on the Internet Web site:
www.minervaeurope.org

Neither the guidelines nor the references are exhaustive – however they provide the most important information needed by a project which is addressing a particular task or tasks within the life-cycle of a digitisation project.

3.2 Digitisation Project Planning

Planning of the project is the first step in any digitisation project. Time spent on planning the project will pay dividends in the easier management and execution of the project. Normally, the following questions have to be answered:

- What (work needs to be done)?
- Who (should do it)?
- Where (should it take place)?
- When (will it take place)?
- How (will it be done)?

A digitisation project should have clearly specified goals and objectives – these will impact directly on the selection, copyright and publication. The project should have suitable personnel, with appropriate knowledge and skills, as well as a training plan in place to provide any additional expertise that the project may require.

A project should not begin until some research has been carried out into other projects in the same area. Such research will identify issues which need to be addressed, will stimulate new ideas and areas which might not yet have been considered, and will add value and credibility to the project output.

Research will also help to indicate the amount of work which may be planned for the execution of the project, by meeting or talking with organisations which have completed similar projects. Such interactions will help to establish whether your organisation has the personnel, the skills and the technology infrastructure to carry out the project, or whether significant training and preparation will be required.

Some time may profitably be invested in ascertaining the copyright status of the material which is to be digitised. Failure to secure permission to digitise and to publish on the Web can cause the failure of a digitisation project, despite any technical expertise and experience.

A technical pilot may also be considered, at the start of the project, in order to ensure that any anomalies or problems with the technical workflow are resolved before commencing the main project.

The Reasons for the Project

Issue Definition
Each digitisation project has its own reason for being executed. Often, the reasons involve providing access over the Internet to cultural holdings which would otherwise be underused, or protecting fragile holdings from the wear and tear of hands-on access. In other cases, the projects are exercises in inter-body cooperation, and involve the establishment of portals, networks, etc.

The reasons for the project will have a profound effect on the criteria for selecting the material to be digitised. They will also affect the project management, the meta-data, the online publication (if any) of the project output, the quality control etc. 'Why' is the most important question to raise before starting a digitisation project.

Pragmatic Suggestions
- The project must have concrete, explicit aims, and these aims must be documented.
- The aims of the project should be realistic, when compared with the resources available.
- All steps of the project should be validated against these aims, in order to ensure that work carried out in the project contributes towards the achievement of the guidelines.
- The project aims should document the value which the project will bring to the institutions involved in the project. If time and effort are to be invested in the project, the justification for the project, from an institutional point of view, must be clear.

Human Resources

Issue Definition
Before a project can start, it is important that the personnel required to work on the project is available. Many cultural bodies do not have large corps of staff who have a great deal of free time to carry out digitisation projects, over and above their usual duties. Also, the knowledge required for digitisation projects may be different from the skills necessary to carry out the tasks for the daily operation. Hardware and software solutions required for a digitisation project, therefore, need to be identified.

Pragmatic Suggestions
- Ensure that sufficient staff is available to carry out the project.
- Assign staff to each task or work-package of the project plan.
- Identify training requirements, including information technology training and education in the handling of delicate artefacts and documents.
- Carry out, if possible, training by using the hardware and software solution which will be used during the project, before the project commences (vendors sometime offer technical solutions free of charge for training or relevant equipment can be hired short-time).
- Aim for a small core of skilled staff dedicated to the project, rather than a large group of 'occasional' staff.

Notes/Commentary
Even if the material presented in this guideline is common to all project management scenarios, it is worth to reperat this matter: in fact there is a possible risk to irreplaceable artefacts and documents if the resourcing is not properly handled.

Research

Issue Definition
Regardless of the scope of any particular project, it can be assumed that similar projects have been carried out in the past. There is a strong likelihood that information about such projects will be available on the Internet, or else published in appropriate journals, etc.

Researching the area as part of the project planning process can help to identify candidate hardware and software solutions, to plan workflow and process, and to avoid issues and obstacles which have been experienced by other projects.

Pragmatic Suggestions
- As early as possible in the planning process, carry out research into any other projects which are addressing similar issues to the project being planned. This handbook provides a starting point; however the amount of material available on the Internet is the largest and most comprehensive resource.
- Research helps avoiding mistakes. It can also put the project team in contact with others who have completed similar projects, and give the opportunity to learn from their experiences.
- Having carried out research adds credibility and value to the output of any project. Assurance that your project has not been carried out in a vacuum, by taking into account the work of others, enhances the results of your project.

Notes/Commentary
Many cultural digitisation projects are funded with public funds, and have a requirement to publish their findings and their reports. Such publication is almost always on the Internet, as well as using other appropriate media.

Project teams are usually very happy to share their experiences and their results – this adds value to their work.

Risks

Issue Definition
At the start of any project, plans have to be made to guarantee a successful outcome. However, the goal is not to eliminate all risks but to prepare for them by creating a project framework which responds to the unforeseen in a resourceful and effective way. The aim is to create a project with staff and procedures that can accommodate changes. Therefore, all project planning need to have a risk analysis.

Pragmatic Suggestions
- Distribution of digitised images over the Internet is a form of publication, and is by this reason covered by laws of copyright and Intellectual Property Right (IPR). Examples of questions in a risk analysis are:
  - What could be the consequences of using the material without specific permission?
  - Has attempts been made to find the rights holder?
  - If copyright infringement does occur, what would the impact be on the project?
- For public information the legal value of the information is an important issue. What steps have been made to guarantee that a digitised source material is not corrupt and has actually been produced by an authorised institution?
- The authenticity must also be guaranteed. What actions have be undertaken to maintain the image files, and what tools have been used?
- Financing the project could be a problem and, therefore, a potential risk for the possibilities to reach the goals of the project.
- A key question is the level of skill in the project. Is it possible to hire new highly skilled and experienced persons? If not, will it have an effect on the work plan of the project?

3.3 Selecting Source Material for Digitisation

The selection of the material to be digitised is an important decision for any digitisation project. Typically, the ideal choice is to digitise all the material in a collection or holding; however, this is rarely feasible, so choices must be made. The criteria for selection will differ, depending on the goals of the digitisation project; an online resource for schools may choose to digitise material in line with a syllabus, while a museum may digitise its best-known holdings in order to stimulate visitor numbers or its most fragile artifacts in order to minimize demand for 'hands-on' examination. These are of course not the only issues to be addressed in the selection criteria – the reasons for choosing to digitise particular material will vary from project to proposal, as will the reasons for deciding not to digitise. Examples of other reasons include legal constraints, institutional policies, technical difficulty of digitisation, already-existant digital copy, etc.

Establish Selection Criteria

Issue Definition

When planning a digitisation project, the choice of which material to digitise is critical. The criteria for selection will depend on the goals of the project, as well as on technical and financial constraints, copyright and IPR issues, and the activity of other projects in the area.

Pragmatic Suggestions

- It is essential to establish criteria for the selection of material to be digitised. The selection criteria must reflect the goals of the overall project. At least the following criteria may be considered:
  - Access to material which would otherwise be unavailable, or of limited availability;
  - Wider and easier access to very popular material;
  - Condition of the originals;
  - Preservation of delicate originals, by making digital versions available as an alternative
  - Project theme
  - Copyright and IPR
  - Availability of existing digital versions
  - Cost of digitisation
  - Appropriateness of the source material for online viewing
- The criteria for selection should be explicit and discussed with, and endorsed by, all relevant stakeholders, prior to selection or digitisation.
- The selection criteria should be fully documented, so that the reasons for any decisions to digitise or not to digitise are clear throughout the project.

Notes/Commentary

Most commonly, cultural bodies have a core of high-value, high-user-interest material which is, by default, included in any digitisation project which is meant to represent the institution.

A large proportion of all digitisation projects have online Web publication as a goal. This means that the copyright and IPR issues which surround any material which may be digitised must be considered before selection.

Selection against the Criteria

Issue Definition
Having established the criteria against which material is selected to be digitised, the actual selection process can take place. This guide suggests how to manage this process.

Pragmatic Suggestions
- Each candidate for digitisation must be evaluated against the selection criteria. In case that any selection criterion is not met, this should be noted. In the event that this results in the rejection of important or critical objects, it may be necessary to review the selection criteria. Should this occur, the new criteria should be noted.
- Once an object has been selected for digitisation, its details should be entered into the digitisation management knowledge base (see chapter *Managing Digitisation Projects*).

Notes/Commentary
At this stage, the project is engaging with each of the items to be digitised, for the first time. This is the optimum opportunity for the project to create a knowledge base of all the items in the scope of the project. Having such a knowledge base will support the management of the project, and help to ensure that, for example, the appropriate expert knowledge is acquired for handling rare artifacts, as well as more mundane issues such as the location of originals.

3.4 Preparation for Digitisation

An appropriate environment and hardware/software system must be in place before digitisation can begin. The elements of such an environment include hardware for the digitisation process itself (e.g. scanners, digital cameras, copy stands, other hardware), a computing infrastructure to which the hardware is connected, software for image capture and processing software, as well as software for meta-data and quality control. The working environment should be appropriate to the material being digitised, paying special attention, for example, to light, humidity, vibration, disturbance, movement of the originals, etc.

Hardware

Issue Definition
The appropriate technical equipment must be in place for the digitisation to go ahead. Typically this will consist of digital image capture equipment (digital cameras, scanners for books, documents or microfilm, audio and video hardware, if appropriate) connected to an appropriate computing platform (computer, operative system, network, etc**).** Two different digitisation methods, using different hardware, can be distinguished: scanning and the use of digital cameras.

Pragmatic Suggestions
- Appropriate hardware must be installed and its quality and functionality controlled before digitisation begins.
- Relevant test targets should be used for the evaluation of digital image capture devices.
- No source material should be present until the hardware environment has been fully established and tested with non-sensitive materials.
- Most digitisation projects will require a flatbed scanner, for material which is not harmed by being pressed flat against a hard surface (e.g. unbound printed material and manuscripts).
- The largest possible scanner should be acquired by the project. The folding or mosaic canning of materials should be avoided. The project should bear in mind that the transportation of large (e.g. A0) scanners is not trivial.
- Usually, a flatbed scanner should only be used where the material is already flat, and will not be damaged by being held against a hard, flat surface. A scanner with a book cradle may be appropriate for many bound articles, up to the appropriate size limits. Most digitisation projects will require a digital camera, for capture of material which cannot be flattened or held on a scanner book cradle.
- If a scanner is used, it should ideally be at least as large as the item to be scanned.
- Image capture (by scanner or digital camera) should be carried out at the highest reasonable resolution. This will often result in very large master files; smaller files can be extracted from the master, for purposes such as Web delivery. However, a higher-quality image can never be derived from a lower-quality image.
- The definition of a 'reasonable' resolution will depend on the nature of the material being scanned, and on the uses to which the scanned image will be put. For example, if the scanned images are only ever to be used as thumbnails, this can allow scanning at a low resolution. Equally, the resolution must capture the most significant details of the item – if scanning at a high resolution yields no more information than at a lower resolution, the high resolution scanning is difficult to justify.
- Image capture should create a file format which is loss less, i.e. not compressed. Typically, the Tagged Image File Format (TIFF) is used. It should be considered to keep a master copy of the image in the camera's or the scanner's raw format, if possible; this can be important for future processing of the images, such as for high quality cuts, zoom etc.
- If a digital camera is used, a project shall choose the most powerful and flexible one which can be afforded. The limitations of the digitisation hardware cannot be overcome by any subsequent processing. It should be noted that 'digital zoom' does not provide a better quality picture; it merely displays less pixels per unit of view. In order to capture detail, three parameters are most important – the number of pixels in the image, the bit - depth, and the quality of the optical lens being used.
- It is important to have appropriate stands for holding material while it is being digitised.
- A digital camera with a dedicated copy stand should be used. The camera should be mounted on a tripod or on an elevation column, and have supplementary lighting, filters, etc, as appropriate. Consultation with an experienced digital photographer with a background in similar projects is advised, if at all possible, before setting up the hardware environment.
- The photographic plane and the plane of the material being digitised must be exactly parallel, if the image is not to be distorted.
- Appropriate lighting must be part of the photographic set-up when using a digital camera; it is very rare for ambient light to be sufficient.
- Suitable filters should be used in order to reduce colour distortion.
- A computer with significant storage should be connected to the devices. This computer should be backed up very regularly – this requirement reflects the high costs in time, technology and possible wear on the originals, of the digitisation process.
- If an item must be scanned in multiple parts, an overlap of several centimetres should be provided, in order to ensure that there are no gaps between the parts. The same settings, light, etc should be used for all parts, in order to avoid any 'patchwork' effect.

Notes/Commentary
The hardware used is a major constraint on the quality of the end result of any digitisation project. Unless the project is digitising only flat materials which can be scanned without damage to bindings, frames or the

source material itself, the use of a digital camera will be important. While an analogue camera can be used, and the slides or prints scanned, the advantages in terms of time, effort and quality of a high-specification digital camera are many.

If the project has a limited life-span, renting hardware may be appropriate. Another alternative is the use of external agencies to carry out the digitisation on behalf of cultural bodies involved in the project.

Software

Issue Definition
Having created a digital version of the object, the resulting file is likely to require processing before it can be used. Colour may need correction; extraneous detail may need to be cropped (removed) from the edges of the image, etc. Also, the master files are typically very large, so a smaller file in a compressed format will often be needed (e.g. as a thumbnail image, or for Web delivery).

Pragmatic Suggestions

- When the scanner or the digital camera is turned on, a calibration routine should start automatically.

- Suitable image processing software will be needed to utilise the master files for whatever the purpose of the digitisation project may be. While digitisation hardware will typically be provided with some software included, this is usually not of sufficient power and flexibility for many projects.

- The requirements on the software depend on the aims of the project. It is worthwhile to note that, once the master files are not modified in any way; various different types of software can be used to process them. However, the cost in time and effort may be significant, and will usually overshadow the cost of a more powerful software package.

- The project should acquire the most appropriate and powerful software package which it can afford.

- As an absolute minimum, the software must be capable of:
  - opening very large image files;
  - modifying the resolution and the colour depth;
  - saving multiple different versions, in different file sizes;
  - selecting and copying a part of the image, and saving this as another file;
  - exporting images in different file formats, including the Web common standards JPEG and GIF.

- Several free software packages provide this level of functionality; however investing in a commercial product is likely to pay dividends in time, effort, documentation and technical support.

- In the event that the digitisation project has an OCR component, the choice of software is also critical. Any OCR exercise has a certain amount of manual editing and

correction; the manner in which this is supported by the software product in use can have a significant effect on the time and effort required by the project. Better OCR packages may enable review and editing on a single screen, suggest possible corrections for mis-read words, support the use of multiple text columns (e.g. newspaper layout), etc.

The evaluation of multiple OCR packages is likely to be worthwhile, if the project exceeds, for example, one person-year in size

Notes/Commentary
The right software will save a digitisation project a large amount of time and effort. If the project is of significant duration (e.g. more than two persons for more than six months), evaluation of several software packages may be worthwhile, in order to establish the best match for the requirements of the project.

Environment

Issue Definition
Many rare or delicate materials require a particular environment. It is critical to any digitisation project that the digitisation process has the least negative effect on the source materials. An appropriate digitisation environment is important to many digitisation projects.

Pragmatic Suggestions
- The environment in which digitisation takes place is of considerable importance.
- Expert opinions should be sought in order to ensure that all aspects of handling of original material are addressed as well as possible. These include the environment for digitisation.
- The area used for digitisation should be dedicated to the digitisation project for the duration of the project. Excessive movement, rearrangement etc. of the workspace can lead to damage, loss or other negative effects on the source materials, as well as to loss of time by the project.
- If the source materials have particular requirements in terms of light, humidity, etc., then these should be replicated as closely as possible in the digitisation environment. For certain materials, such as leather documents, a short-term increase in humidity may assist in relaxing the materials prior to flattening for photography or scanning.
- In almost all cases, direct exposure to bright light (e.g. sunlight) for extended periods is not recommended. Smoking, eating and drinking in the vicinity of the items should of course not be permitted – keep coffee away from the work area!

Notes/Commentary
Depending on the size and budget of the project, a dedicated digitisation environment may not be feasible. However, the aims outlined here, to minimize movement, disruption and handling of the materials, should be kept in mind.

As with the handling of heritage material, no references should be taken as a substitute for discussion with those whose responsibility includes the care of the material.

3.5 Handling of Originals

This section considers how a digitisation project should treat the material which is being digitised. In many cases, the source material is rare or valuable; the negative effects of digitisation on the source material must be minimized.

In every case, it must be emphasized that the specialist knowledge of the individuals who are responsible for the source material on a day to day basis will be valuable to the project team.

Moving and Manipulating Original Material

Issue Definition
In many cases, the material to be digitised is of particular sensitivity or fragility. Replacing hands-on access with online access is often an important reason for digitisation projects in the first place. It is critical that any digitisation project takes steps to ensure that no damage is done to the original material during the digitisation process. These steps may range from the use of the correct hardware to the establishment of a suitable micro-climate or the movement of the digitisation centre of operations to the location of the material, rather than vice versa.

Pragmatic Suggestions
- Consult the person usually responsible for the source material, before moving or handling it. Include any information on appropriate handling, in the digitisation project knowledge base (see chapter Digitisation Process Management)
- Be prepared to be flexible – an inconvenience to the digitisation project can be overcome, while damage to a unique artifact may be irretrievable.
- If necessary, bring the digitisation equipment (e.g. digital camera) to the source item, rather than transporting the item itself.
- Avoid unbinding of bounded books and records. Use instead of a flatbed scanner a scanner with a book cradle or a digital camera.
- Always remove staples, paper clips, and other fasteners; they can damage both the digitisation devices and the source material.
- Expert advice (e.g. from the curator of the item to be digitised) should be sought before any handling of the original.
- This advice should be sought prior to digitisation, ideally at the time that the article is selected for digitisation. The advice should be recorded in the digitisation management knowledge base, and consulted before movement or digitisation of the article. If necessary, the expert should be briefed on the capabilities of each possible hardware solution.

Notes/Commentary
While much of this material is quite obvious, it is important to establish and maintain a discipline while handling the source material.

3.6 The Digitisation Process

This chapter provides some practical guidelines for the actual digitisation process. The technical solutions for digital capture can differ. Scanners, digital cameras or software applications for optical character recognition (OCR) are areas covered in some detail, as being most relevant to the largest number of projects. The digitisation of transparent originals like microfilm is not considered.

Using Scanners

Issue Definition
Flatbed scanners are a very common digitisation tool. The most common A4 and A3 models are relatively cheap, require limited skills to use, and can manage a fast throughput of material, once a workflow has been established. Larger models (up to A0) of flatbed scanners and scanners equipped with book cradles are very expensive and thus require long-term projects/programs, high-volume digitisation, or oversized source material.

Pragmatic Suggestions

- Only scan material on a flatbed scanner which will not be damaged by being pressed flat onto a hard surface. Consult the experts, if in doubt.
- Ensure that the glass scanning plate is completely clean at all times. This both leads to better image quality and also protects the source material from soiling.
- If possible, scan only items which fit on the flatbed scanner or the scanner equipped with a book cradle in one piece.
- If it is necessary to scan an item in multiple parts, ensure that there is sufficient overlap to allow the image to be reassembled, during post-processing (by using mosaicing software).
- Test the scanner, and its output, on non-sensitive material before beginning to scan original source material. Train users with the same non-sensitive material.
- Establish a file-naming convention for the files produced by the scanner, for example by using the existing cataloguing system or giving them meaningful names. The file name should allow mapping between the file and the source item.
- In order to maximise the portability of files across computer platforms, a file name with a maximum of eight characters, followed by an extension of at most three characters, should be adhered to.
- Before establishing workflow or work-batching process, carry out some end-to-end scanning and image processing, in order to ensure that the end result of the workflow will be what is anticipated.
- Scan at the highest resolution and bit-depth that is feasible given the reasons for the project, the limitations of the scanner, the conditions for data storage, and the attributes of the source material.
- Scan with the maximum appropriate colour depth, given the same limitations.

- Back up the hard disk where the data is stored, on a daily basis.
- Quality control of the digital image and of meta-data is important – at scanning time is the most convenient time to address any quality issues. The following points may be borne in mind:
  - Establish minimum resolution and colour parameters (mainly the spatial resolution and the bit-depth) for groups of items to be scanned.
  - Examine the scanned output on screen, on paper and in any other format that you expect it to be used for (e.g. on a mobile device).
  - Ensure that the screens (monitors) being used are reliably calibrated. Avoid having other material on and around the screen, which may affect the perception of the item
  - Master images must be created with visible scaled rulers, and colour or greyscale images must include also a standardised colour/greyscale reference target.

Notes/Commentary

Scanning is in itself a relatively simple operation. However, in order to increase efficiency and minimize errors, having a workflow system in place will be worthwhile.

Scanning of oversize items, or very high quality scanning, takes a significant investment of time and effort per item. This can be reduced by using hardware appropriate to the item (e.g. a larger scanner, a book cradle); in the event that large hardware resources are not available, allow plenty of time. Training on oversize or irregular materials should not be neglected.

Using Digital Cameras

Issue Definition

The use of digital cameras is becoming increasingly common in digitisation projects. This reflects their flexibility in terms of being able to digitise non-flat objects, such as bound books, folded or wrinkled manuscripts, and 3D objects. However, a scanner equipped with a cradle is normally to prefer when digitising bounded books and over-sized material like maps and drawings.

Pragmatic Suggestions

- Consider renting a high-quality camera, if the scope of the project is limited.
- Put up the digital camera on a motorized carriage on a column and place the items to be digitised on a steady copy stand board with specially tailored lights.
- Organise training from a specialist digital photographer – the difference in quality between pictures taken by an amateur and the same photos taken by a specialist can be striking.
- Ensure that backgrounds will show the item clearly.
- Avoid changing the light conditions between shots, and between images of different parts or sides of an item – this can lead to erroneous impressions of colour variation.
- Use apochromatic lenses and appropriate lens filters to combat colour misregistration and image distortion.

Notes/Commentary

The increasing use of digital cameras in digitisation projects reflects their availability as a mainstream consumer product, and the resulting decrease in price. However, there remains a significant difference, in both price and quality, between specialist digital cameras and mass-produced low-end consumer products.

Software Applications for Optical Character Recognition (OCR)

Issue Definition
Many digitisation projects involve the digitisation of printed documents, such as books and newspapers. This occurs most often (though not exclusively) in tandem with the use of scanners. The use of OCR software is a popular way to extract the information from such scanned information, and to open opportunities for processing the information. OCR software recognises the letters and numbers which make up the scanned image (bit mapped image file), and exports them as ASCII text files, rather than as image files. This enables searching, indexing, format conversion, and other data processing operations to be carried out.

Pragmatic Suggestions
- Evaluate multiple OCR software offerings before selecting a particular product. While OCR software is often included with the sale of a scanner, more powerful software is typically sold separately.
- A major element of any OCR project is the identification and manual editing of mistakes, ambiguities and locations where the text could not be processed. An OCR package which provides a friendly user interface for carrying out this task can save considerable time and effort.
- OCR works best with documents which are in good condition – folding, wrinkling and discoloration of the source material will increase the number of errors and faults in the OCR process. Pre-treatment, where possible, of the source material should be carried out to avoid this.
- The use of image processing software, to remove discoloration and improve contrast, before the use of OCR software, should be considered for material which is not in perfect condition.
- The availability (or not) of dictionaries in the language of the source material, as part of the OCR package, should be verified.

Notes/Commentary
Relevant products in this market include:
- OmniPage
- TextBridge and
- Adobe Capture.
- Abby FineReader

The last of these has excellent editing and fault resolution functionality.

3.7 Preservation of Digital Master Material

In the longer term, it is an important goal of any digitisation project to protect and keep accessible the data which it has created. This involves dealing with the inevitable obsolescence of digital file formats and various types of computer storage media.

Preserving the digital master material and corresponding meta-data helps to avoid having to re-digitise any items, thus protecting the fragile source material and avoiding repetition of the labour-intensive digitisation process including generating meta-data.

File Formats

Issue Definition
The digital output of the digitisation process is usually a master file in uncompressed TIFF format with some meta-data embedded (see chapter The Scope of Meta-data Used for Object). The file format as well as the compression used will have a major impact on the usability of the digitisation output. At this time, issues such as file format, standards file size, network transmission time, and different kind of outputs (monitor or printer) need to be taken into account.

Pragmatic Suggestions
- Before deciding on a file format, take into account the relevant standards, the established global user base and the degree to which file formats are supported by software in use by your organisation and your target audience. The size of the global user base is a good indicator of the future, ongoing, support for a particular file format. It also indicates the likelihood of sustainable migration paths, when file formats change.
- The default digitisation output file for digital images is Tagged Image File Format (TIFF). Unless your project has a clear, justified reason for using some other file format, digitisation output, and so master files, should use this format.
- The output file will typically be quite large. It is common to have a large master file, which is stored locally but not transmitted over the Internet. From this master file, smaller versions can be created using image processing software, either in TIFF, or more commonly in a delivery format such as JPEG 2000, PNG or GIF.
- Regardless of how attractive a proprietary or national format may appear to be from a technical standpoint, it is important to bear in mind that failure to use standard formats and media will act as a major obstacle to international exchange of raster image files and corresponding embedded meta-data as well as the creation of networked resources.

Notes/Commentary
File format choice must be governed by the imperative to create the highest quality digitisation output, and by the availability of migration paths for future preservation of the digital master. The role of standards in this area is very great.

Media Choices

Issue Definition
The issue of media choice is an important one for projects which wish to maintain their digital collections over a several-year period. Important projects such as the UK Domes day book initiative have been lost due to media obsolescence.

Pragmatic Suggestions
- The output of the digitisation project will be held on server machines, including those which serve digital content to Internet users. However, these machines need to be backed up. Also, if a server is not dedicated to a digitisation project, the digital content should be stored on removable media, separate to other data on the server.
- All master files (including meta-data) should normally be backuped on two kinds of media separately stored.
- Currently (early 2004), the use of CD-Rs as a common backup medium is in the process of being replaced by the use of DVDs. DVDs offer significantly larger storage, and the hardware needed to read them is becoming ubiquitous on new PCs and laptops. DVD writers remain more expensive, but are already well within the means of all but the smallest projects.
- However, DVDs are not expected to replace magnetic tape media like Digital Linear Tape (DLT) as the storage medium of choice for backup of computer storage, in the near future. Both of these technologies should be seriously considered as candidates for actual storage, but carefully considered for long term preservation.
- Regardless of the choice of medium, it must be borne in mind that the medium will become obsolete in near to mid-term future. Within five years, migration to new storage media is likely to be a necessity.

Notes/Commentary
The rapid change of media layouts, driven primarily by the consumer electronics industry, has had major effects on digitisation projects in the past.
However, the increasing trend to store data “on the Internet” on large server machines, and as data on mobile hard drive units, facilitates the migration of data from place to place and from medium to medium. Once servers are backed up and migrated to new servers over time, the dependence on removable media as the only record of a digitisation process can be expected to decrease.
In the meantime, the issue of media selection is still an important one. There is no indication that the limits of compressed, small-footprint digital storage are being reached.

Migration Strategies

Issue Definition
As noted above, the choice of file format and storage medium must take into account the feasibility of moving data to a new file format and/or a different storage medium, in the foreseeable future.

Pragmatic Suggestions
- Examine the relevant standards for file formats and storage medium, as noted in the previous two guidelines. Compliance with standards is a reasonable indicator that a particular format or medium will have some support into the future.
- Proprietary file formats and non-standard media formatting should be adopted only with great care.
- Migration from one format to another should avoid migrating from a lossless file format (e.g. TIFF in the image domain) to a lossy one (e.g. JPEG), for master digital material. Once information is lost, it cannot be replaced.
- Bear in mind that any choice of file format and/or storage medium will become obsolete in the foreseeable future (possibly less than five years, probably less than ten years).
- The size of the market for storage media provides an indication of how likely it is that migration from one medium to a new one will be feasible, as the medium becomes obsolete.
- Having created the digitised material, storage media (e.g. CD-R, DVD) should be refreshed periodically (once every two to three years), to combat data loss. This involves copying all data to new media.
- The status of digitised material, including when it was last refreshed, should be recorded in an appropriate log.

3.8 Meta-data

The area of meta-data is one of the most actively researched and dynamic in the whole digitisation area, as well as in areas such as information retrieval, Web searching, data exchange, enterprise application integration, etc.

The selected meta-data model is of particular importance as it influences the choice of attributes to describe an object. Related to this is the choice of a standard model, as it is described in the following chapter on standards.

The Scope of Meta-data Used for Object Description

Issue Definition
Before selecting a meta-data model for a digitisation project, the material to be described with the meta-data should be reviewed. This will help to identify existing meta-data models, as well as to pinpoint any omissions or gaps between what is covered by an existing meta-data model and the important meta-data for your project.

Pragmatic Suggestions
- The use of appropriate meta-data is very important for enabling search and retrieval of material from digital collections. This is even more the case when searching across multiple collections, stored in different locations, is the overall objective (logical union catalogues, virtual combined museums, etc.).
- There exist already many meta-data models. Therefore, each project has to choose as meta-data model based on its own goals. It is advisable to avoid creating a new one, unless the requirements of your project are badly underserved by all existing standards.
- Time spent modelling the important characteristics of the material being digitised, and identifying its key attributes and descriptors will be well invested. Such a model can then be compared with the scope and features of existing meta-data models.
- Possible controlled vocabularies (e.g. to describe a location, or an artist) should be identified. Several such vocabularies already exist and can greatly increase the success of searches, etc. See the section on meta-data standards and controlled vocabularies, below, for details.

Notes/Commentary
The Making of America II project (Library of Congress) used three categories of meta-data:
- Descriptive - for description and identification of information;
- Structural - for navigation and presentation;
- Administrative - for management and processing.

Of particular importance are the meta-data models to be selected for a digitisation project – the choice of which set of attributes that will be used to characterize the works and items to be digitised, the resulting images, the description of the undertaken processes, techniques and technology, the rights management, etc.
The National Library of Australia has a powerful model for this.
The plethora of existing models and competing standards for meta-data has led to projects which focus purely on translating from one standard to another.

Appropriate Meta-data Standards

Issue Definition
Certain important standards already exist for meta-data. In the bibliographic domain (and increasingly in non-library cultural domains), the Dublin Core standard is of great importance.

Pragmatic Suggestions
- Review existing meta-data models and standards before creating your own.
- Creating a totally new meta-data model for cultural collections should be avoided.
- The meta-data work carried out by similar projects in the past is likely to be relevant to your project – meta data models travel well between projects in the cultural area.
- Unless your project has good reason not to do so, the Dublin Core fields should be included in the meta-data model. While museums may find the CIMI model better fits their holdings, a common core set of attributes should be aimed for, which will enable cross-collection searching.
- If a proprietary meta-data model is to be used, a mapping from this model to the Dublin Core should also be developed.
- While a naming scheme or national naming convention may be very useful, a full meta-data model is better, both in terms of the amount of data that can be stored about an item, and also to enable more powerful searching and interoperation with other projects and other countries.

Notes/Commentary
There is an impressive number of existing standards, covering various aspects of meta-data. However, there is also significant overlap across standards, and a very large population of institution-specific models, where sectoral or cross-domain models have been neglected.

3.9 Publication

At this stage of the project, the digital master material has been created and stored/backed up. A suitable meta-data model has been identified, and the meta-data associated with each article has been created.

Preparation for publication involves processing the newly-created material prior to publication. Typically, publication means display on the Internet, and processing means reduction in image/audio/video file size, quality, and downloads, to fit the operational characteristics of the Internet.

Image Processing

Issue Definition
The TIFF files created during the digitisation process are typically very large (a few to many megabytes). Such files are not appropriate for Internet publication, due to the great length of time that they would require to download to the end user.

Pragmatic Suggestions
- Create delivery versions of master material. As a minimum, there must be one delivery version. A second version, a 'thumbnail', may also be useful, depending on the layout of the Web site on which the material is to be published.
- Delivery versions are created by opening the master TIFF file in an image processing package, and exporting it in JPEG, PNG file formats.
- Typically, colour resolution can be reduced, to 256 colours. If this shows an appreciable loss of quality, a higher colour resolution can be used. Choosing the right colour resolution usually requires some subjective decision to be made.
- An image created at 72 DPI will show at approximately its original size on many computer monitors. This makes 72DPI a reasonable choice for many images which are to be viewed on-screen. For lower resolutions, a subjective decision of 'acceptable quality' will be required.
- Choosing file format, colour resolution and pixel resolution involved deciding on what is 'acceptable' quality. A balance must be found between quality and file size.
- In general, the total image files on a Web page should not greatly exceed 100 kilobytes. Larger images can certainly be published; however, these should be accessed via a link from the Web page, with suitable warning text that the download may be prolonged.
- Unless material is being streamed, video and audio material will typically involve large file sizes, with the file downloaded before viewing offline. However, the download time can be adjusted by changing the frames per second of the video, the sampling rate of the audio, etc.

Notes/Commentary
Decisions regarding image processing depend to a large degree on personal judgement. The guidelines provided here may be considered too strict or too lax, depending on the project and the end user audience.

Image processing software such as Paint and Paintshop is freely available online. More powerful image processing

software may save sufficient time and effort to justify the expense of the software.

Audio and video editing software is also available freely online. Equally, audio and video hardware is usually supplied with the software required to edit and process the data created.

3D and Virtual Reality Issues

Issue Definition

The guidelines provided above for image publication are not immediately applicable to digital renderings of 3D and virtual reality material. However the balance between quality and file size is a common one on the Internet.

Pragmatic Suggestions

- Viewers for 3D and VR material are not yet widely distributed with operating system software. This contrasts with image, audio and video, which are commonly provided with Windows software.
- Ensure that viewers for any 3D or VR material are readily available. Make the viewer software available from the same site as the material. This helps to overcome any issues with other software download sources becoming unavailable.
- Evaluate multiple viewers before endorsing one or another. Compatibility across file formats and viewers is not as standardized as in the still image domain.
- Modern PCs, with a focus on games, will often have hardware accelerators and increased graphics memory. This can have a profound effect on the VR viewing experience.

Notes/Commentary

- A VRML viewer which has been successfully used in one of the reference projects (the Irish ACTIVATE project) is the Blaxxun Contact viewer.

Online Publication

Issue Definition
Many digitisation projects in the cultural area lead to the creation of online cultural resources, usually a Web site with images, meta-data, 3D artifacts, etc. They range from the simplest content sites to complex, software-driven portals and viewing engines. A large body of knowledge covers the creation of Web sites; only a few guidelines are provided here, as well as links to examples of Web sites nominated as best practice examples by Minerva partners.

Pragmatic Suggestions
- Web sites should be easy to navigate – links to the front page or to a table of content should be available throughout.
- Due attention should be paid to universal access and to the utilisation of Web sites by the partially sighted and other disabled persons.
- Web pages should be short enough to minimize the amount of scrolling necessary by the user.
- Images should be small enough not to disrupt the browsing experience. Larger images should be linked to from the Web pages, with a note to the effect that the image is large and download may be slow.
- The use of animations, pop-ups, pop-unders, Flash and similar technologies should be treated with care. It should be possible to bypass lengthy introductory animation sequences.
- Web sites should ideally be multilingual, with at least the host country language and one or two other languages (commonly including English, as the de facto online language standard) supported.
- Links to external resources should be verified on a periodic basis, in order to minimize dead links and the annoyance associated with these.

Notes/Commentary
The actual process of making material available on the Web is one which is widely understood and documented. This handbook does not provide guidance on how the create Websites, program in HTML, build Web-enabled databases and carry out the other tasks which are needed to create and maintain a Web presence. It is anticipated that many of the cultural institutions which utilise these guidelines will already have some Web server functionality availability, which they will exploit for their digitisation project.

3.10 IPR and Copyright

The publication of any material online must be accompanied by some consideration of the Intellectual Property Right (IPR) associated with the material. For material which is in the public domain (such as particularly old books or newspapers, or material placed explicitly in the public domain), there is relatively little difficulty. However, many cultural institutions derive revenue from the use of images of artifacts or images in their collections, and so are defensive of copyright. Material, the copyright of which is held by third parties, can only be published with the consent of such third parties.

Fortunately, a range of technical options are available to protect the copyright of material placed on the Internet. These are surveyed here.

Establishing Copyright

Issue Definition
The initial step when exploring the copyright situation for a cultural item is to establish the ownership of that copyright.

Pragmatic Suggestions
- Establish the legal situation with regard to copyright and publication in the country where the project is being carried out. Each country has its own copyright laws, usually dating back to at least the 19th century. Such laws usually apply to all forms of publication, including online publication. They may, or may not, cover the act of digitisation, which may be construed to be an archiving process, or may be considered copying.
- On no account should online publication go ahead without copyright being sought.
- Certain items, e.g. old newspapers, have clear copyright rules governing them. Typically these allow free copying once the papers are of a certain age. Items which fit into this category can be freely digitised and published.
- For items whose copyright is vested in the institution carrying out the project, internal permission will be required for digitisation and online publication.
- For items whose copyright is held by a third party, such as the lender or donor of a collection of historical items, that party's permission must be sought, in writing. Only when such permission has been received, should publication go ahead.
- Securing permission to digitise and publish may involve payment. The amount of payment must be balanced against the value of including the relevant item(s) in the online resource.

Notes/Commentary
The copyright situation varies from country to country.

Safeguarding Copyright

Issue Definition
The publication of items online on the Web is an open invitation to make copies of the items. It is infeasible to prevent some level of copying of material displayed on the Web. However, there are a number of possible procedures which can be considered, each of which has some effect in the safeguarding of copyright.

Pragmatic Suggestions
- Establish whether or not copyright must be safeguarded.
- Agree the procedures to be used to safeguard copyright, with the copyright holders.
- The following procedures are among those which could be considered:
  - Addition of a visible watermark or copyright stamp on each image.
  - Addition of an invisible digital watermark on each image. Such marks can be used to prove the ownership of a 'stolen' image, as well as to track the use of the image across the Internet.
  - Encryption of images, with the issuing of the appropriate key only to registered users. This, of course, reduced the value of the online image to the rest of the public.
  - Restricting publication to low-resolution images, such as 75 dpi for screen viewing. This restricts the degree to which images can be used in other domains, such as printing, clothing, etc.
  - Restrict publication to only small parts of an image.
- Display images only to registered, authorized members of a particular community.
- Test the results of the copyright protection process using the first few items, in order to ensure that the process does not have any unexpected or unwanted effects.

Notes/Commentary
The approach which is most appropriate for any one project will depend to a large degree on the goals of the project and the cultural institution, as well as on the nature of the material. In general, the publication of a small selection of images, at low resolution, is a common approach for online galleries and museums. The relative uniqueness of many cultural holdings provides proof of ownership of copyright in many situations.

3.11 Managing Digitisation Projects

The success of any project, including digitisation projects, is influenced to a large degree by the management of the project. This section provides a small number of guidelines specific to the management of digitisation projects in particular.

Digitisation Process Management

Issue Definition
A typical digitisation project will involve dozens, hundreds or even thousands of items to be digitised. In order to achieve an efficient project, it is important to establish a work-flow that maximises the through-put of the digitisation team. In addition, information resources such as a digitisation project knowledge base will be of significant importance.

Pragmatic Suggestions
- Establish and document each of the steps that an item must go through during the digitisation process. These will include, for example,
  - retrieval from storage / usual location
  - cleaning or preparation
  - scanning or photography
  - return to usual location
  - file naming
  - file storage
  - creation of online delivery versions of large master files
  - backup of servers / storage media
- Develop a digitisation project knowledge base that can be used to track the object through the digitisation process, and enables the status of the project to be reviewed at any time. This knowledge base may take the form of a database (e.g. in MS Access, Oracle, MySQL, etc), or may use a simple spreadsheet or even a collection of documents. The important issue is not the format of the knowledge base, but the process which ensures the recording of actions which are carried out.
- The name, identifier and other relevant information for each item to be digitised should be entered in the digitisation project knowledge base, as soon as the item has been selected. The status of the item (i.e. which step it is has last completed) must also be recorded, on an ongoing basis.
- Procedural choices must be made – for example, should items be collected at the digitisation workstation at the start of each day, each week, or on a per-item basis.
- Articles which require similar activities or hardware setups should be digitised together. This reduces time spent setting up digital cameras, configuring scanners, etc. The parameters for hardware setup should be documented, in order to allow any digitisation to be replicated in the event of file loss, etc.
- The location, phone numbers and backup staff of key service delivery personnel (e.g. IT support) should be

noted at the start of the project, and remain available throughout.

Notes/Commentary
The larger the project, the more worthwhile it is to establish a process and workflow. The efficiencies which this introduces will greatly repay the time spent setting them up. The references below include some projects which concentrate purely on this aspect of digitisation.

Team Development

Issue Definition
Digitisation projects often expose the staff of cultural institutions to new technologies for the first time. Such technologies include digitisation hardware, Web publication, image processing, meta-data tagging, database development and population, etc.

Pragmatic Suggestions
- If possible, include at least one person with appropriate information technology skills in the project team.
- Assess the state of knowledge of the personnel to work on the project, and the IT skills that they will need, well in advance of the project. Identify training needs and fill these before the project starts.
- IT skills are not the only ones which may be needed. Specialist skills may be needed, as noted above, in the handling of delicate documents, artifacts, etc. Appropriate training maybe available from the individuals whose responsibility includes the source material.

Notes/Commentary
It is better to have a small core of skilled personnel working on a project than a larger population of occasional participants. However, while developing and using a particular skill is efficient for the project, staff may prefer to be exposed to the full digitisation life-cycle. Digitisation and meta-data tagging is not in itself particularly rewarding work – exposure to other elements of the project will increase staff satisfaction.

Staff Training

Issue Definition
Unless the staff working on the project has significant experience from prior projects, there will be a requirement for staff training. This will include two quite different areas – the technology to be used, and the handling of the source material.

Pragmatic Suggestions
- Do not assume that no staff training is required, nor that archives, library or museum staff automatically has all the relevant expertise.
- Ensure that the training requirements of the staff on the project are identified at the start of the project, i.e. already in the planning phase. These training requirements should be included in the digitisation project knowledge base, and acted upon before the training is needed in the project.
- Certain training, such as the use of the digitisation technology, may be to learn 'on the job'; other training, such as handling of source materials, requires training in advance.
- A smaller core of personnel, who are trained and develop experience during the whole project, is to be preferred to a larger, more casual group which changes its membership more frequently.
- Technology training may be well delivered from another project in the same institution; alternatively an outside digitisation agency may be able to provide training.
- Curator training may best be provided by the individuals who are responsible for the care of the original material.

Notes/Commentary
A lack of staff training can lead to unfortunate and irreversible accidents or incidents early in the project; the same may result at any time if staff is removed from the project and new personnel start to work. A small, well-trained core is a desirable aspect of such projects.

Time invested in training at the start of the project should be repaid in extra productivity and less problems during the life of the project.

Working with Third Parties for Technical Assistance

Issue Definition
It is often appropriate for a digitisation project to engage the services of one or more third parties during the project. The services which are most commonly provided include the actual digitisation itself, the management of the project, integration with third party systems, software development, etc. This allows the cultural body to concentrate on its own areas of expertise, without need to train and retain staff with advanced IT or other skills.

Pragmatic Suggestions
- As with any other project, the relationship between technical partners and other project members should be governed by clear, strict contracts. A documented and signed specification of the products or services to be provided should be agreed before any work is carried out.
- The work being carried out should be reviewed on a regular basis, to ensure that what is being delivered is in fact what the project wants or needs.
- While the use of third parties can be convenient, it should be borne in mind that any expertise or experience to be gained during the execution of the outsourced work will be lost to the cultural institution at the end of the project. This also applies to temporary staff who is employed for the duration of a project. It may be better to dedicate a long term member of staff to a project, while replacing him in the short term with a contractor.

Notes/Commentary
Certain large projects, such as the French national digitisation programme, have identified a preferred supplier, the relationship with whom may stretch for several projects and several years. Having established a working relationship with a supplier, the value of changing supplier between projects may need to be questioned.

Working with Third Parties in Cooperative Projects and Content Sharing

Issue Definition
Many digitisation projects are either cooperative efforts which involve two or more cultural bodies, or else EU-funded Framework projects, which almost always have multiple partners in multiple countries. The guidelines for establishing and managing multi-partner projects are many, and go beyond the scope of this document. However, a few pointers are included.

Pragmatic Suggestions
- Ensure that all partners are aware of, and have endorsed, their roles and responsibilities within the project. Refresh this knowledge on a regular basis.
- Establish a common mode of communication across partners, and ensure that all partners receive the information which is aimed at them. Electronic mail is ideal for this purpose, so long as partners read and reply to such mail.
- Subcontractors should be governed by strict commercial contracts, with their deliverables clearly and unambiguously defined.
- The IPR of all partners should be clearly documented and formally signed by all partners. A partnership agreement which clearly states the IP Rights covering material which is being brought to the project, and material which is created by the project, should be agreed in advance of the project commencing.
- Each partner should have a clear role in the project – if a partner's role is not clear, review whether or not the partner is necessary to the project.

Notes/Commentary
The notes above are only a small part of the possible material that could be provided on the establishment and management of multi-partner projects. Partners and suppliers are a major source of delay and confusion within a project – clear agreement and common endorsement of the roles and responsibilities of all partners at all times can help to avoid this.

Costs

Issue Definition
Building a digital collection is normally very expensive. Projects, therefore, have to take into account all start-up and infrastructural costs as well as costs for running the project. That means costs for initial planning, data specifications, tracking and documentation systems, staff training, and so forth, as well as the incremental cost for digitising the actual source material.

Pragmatic Suggestions
- Digitisation projects should consider the following costs involved in the design, implementation and maintenance of a digital collection:
  - Staff development
  - Facilities needed
  - Operational costs
  - Costs for storage and for delivery systems
- Staff development includes salaries for project management, Web programmer, educational officer, technical support, etc. but also travel costs and training
- Costs for facilities are often connected to questions concerning required functionalities and the need for tradeoffs. Projects have for instance to decide whether keeping costs on a low level is more important to the overall success of the project than achieving the highest possible standard for image capture.
- Operational costs to consider are:
  - Time for handling source material (from shelf to image capture device and back again) as a percentage of total salary cost per day.
  - Preparation of source material (conservation, cleaning etc).
  - Capture time (from set-up to naming and saving) provided as a percentage of the operators total salary costs per day.
  - Cataloguing and handling of meta-data as a percentage of total salary costs.
  - Hardware and software cost per digitised item (preferably based on depreciations or replacements costs rather than acquisition costs.
  - Quality assurance time as a percentage of salary costs.
  - Hardware and software maintenance.
  - Technical support time related to capture.
  - Project management time related to capture.
  - Training related to capture.
- Be aware of that image capture often is the least costly part of a digitisation project. On average, one third of total costs are connected to digital convention, slightly

less than one third to creation of meta-data, a bit more then one third to administrative and quality assurance tasks. The rest is long term maintenance costs.
- Storage costs to consider should normally be total costs for maintenance per gigabyte.

Notes/Commentary
Sustainability in the long term is often pushed down the list of priorities by more immediate and pressing concerns. Regardless of the quality and robustness of the digital resources created by a digitisation project, they will not last long if the project in mind cannot find funds for their maintenance.

Printed by La Tipografia, Roma, Italy
in the month of June 2004