

An Chomhairle Leabharlanna Cultural Heritage Project

Cork Archives Institute Digitisation Pilot Project

Final Report

Author: Brian McGee, Archivist, Cork Archives Institute

& Peter Foynes, Contract Archivist

<u>Table of Contents</u>	2
INTRODUCTION	3
DELIVERABLES	4
1 Selection Methodology	5
2 Copyright Confirmation Report	7
3 Imaging & Handling Specifications	8
4 Scanning/Imaging Methodology.	12
5 & 6 Tagging/Metadata & Preservation of Digital Master	15
7 Report on Web Content Creation	19
EVALUATION	22
RECOMMENDATIONS	24
Project Management	24
Selection	25
Copyright/ IPR	26
Imaging Specifications & Equipment	27
Imaging Methodology	27
Metadata and Digital Preservation	28
Web Content	29
REFERENCE	31

Introduction

The Cork Archives Institute and Clonakilty Museum Pilot Project was chosen for funding under the Ask About Ireland Cultural Heritage Project, a six month initiative co-ordinated by the Branching Out Steering Group Cultural Heritage Panel, managed by An Chomhairle Leabharlanna and the Department of the Environment and Local Government and funded by The Heritage Council. The Pilot Project is a contribution to the development of experience and expertise in Ireland the area of digitisation for Archives and Museums.

The Cork Archives Institute is the local archives service for the City and County of Cork and is the designated repository for local authority records. The Institute is funded by Cork City Council, Cork County Council and University College Cork and is an important centre of cultural heritage and a centre of research for students and academics, local historians, genealogists and the general public. Public exhibitions and other outreach projects are undertaken on aspects of local history.

The content for the Pilot Project was made up of a small selection of documents from our numerous collections of estate records. Many of these records originated from solicitors' offices and generally consist of the records made during the day to day running of an estate, such as legal and financial documents, leases of land, wills and marriage settlements, rental account books showing the details of tenants' rent payments and terms, correspondence with land agents, and maps and drawings of segments of the estate by cartographers. Occasionally found in these records are more personal papers relating to the family who owned an estate, for example, private correspondence with other families or family photographs.

The material chosen was a small cross-section of the records available, and examples were selected of many of the different types of archival documents which may be found, in order to look at the particular challenges each type would pose for digitisation.

Clonakilty Museum is a regional museum for the West Cork area. The Museum displays material relating to aspects of the history of the locality, including industrial heritage, education, the West Cork railway, the War of Independence, Michael Collins, GAA history, and lace and crochet. The Museum also holds photographs, correspondence, artefacts and ephemera relating to the Hungerford Estate, Inchadoney, in West Cork, and it is these items which were used in the Pilot Project under the general theme of the landed estate in Cork.

Deliverables

Deliverable No	Deliverable title
1	Selection methodology
2	Copyright Confirmation Report
3	Imaging & Handling Specifications
4	Scanning/Imaging Report on and Methodology.
5 & 6	Tagging/Metadata and Preservation of digital master copy. Strategy and Framework.
7	Report on web content and publishing.

1 ***Selection Methodology***

Tasks Documented by this Report : Selection

This report describes the use of criteria for selecting original archival documents for digitisation and how these criteria were recorded and implemented. It explains how Project objectives influenced the selection process and the choice of criteria. Existing finding aids in the Archives were used to narrow down a preliminary list of suitable collections, and then a document sample survey was carried out. A digitisation database was created to enable the recording of the appropriate archival, physical and other data. The management of the selection of documents required a defined workflow and review process which ensured compliance with the criteria and the Project objectives and also ensured sufficient quality of the data entered.

The Report has two main sections covering Selection Criteria and the Selection Process.

1. Selection Criteria

All selected documents were examined under the following criteria:

- **Relevance To Themes And Objectives.** A selection framework was the first step in the selection methodology. This was in order to have a point of reference for content when initially selecting collections and documents from the collections. It was necessary to decide on size/scope parameters when selecting documents in order to create viable content covering each topic in the thematic structure in sufficient detail and breadth, and because it would not be possible to select all eligible documents. The selected documents would be a representative and accurate sample of our holdings of papers relating to the landed estate. (See Intellectual Content & Historical Value). Documents of outstanding physical/visual 'presence' were to be selected, in order to add to the appeal of the final published web content.
- **Intellectual Content & Historical Value.** Information of particular interest historically i.e. A notable historical theme, event, place, time, organisation, or person. Rarity, uniqueness, age, representing the types of archival document.
- **Conservation & Preservation Considerations.** Digitisation will not damage the item. Noting of conservation requirements, possible damage treatments, and the feasibility of undertaking conservation.
- **Copyright & Other Legal Restrictions.** Copyright should be held by the Archives Service or Copyright Permission should be obtainable. Records should not contain private personal information likely to affect living persons.
- **Technical Considerations.** Number of Images/Scans required to copy, degree digitised version can represent full content of original, level of image quality required, display characteristics, unusually high resolution, colour fidelity, oversized dimensions.
- **Available Organisation and Documentation.** Information/Knowledge available about collection/document (Finding Aids) from collection Database, collection box or summary lists, indices, published information, and degree of cataloguing and processing required

2. Selection Process

This Section of the deliverable covers the management and implementation of the physical selection process.

- Staff, Time and Archival Equipment. Review and training sessions were held with the relevant Staff regarding the implementation of the selection criteria and complying with the Project objectives. The handling, conservation and security procedures in selecting the original documents were set out. Training was necessary for certain staff in the handling of original archival documents and basic damage treatment and listing techniques. An archival work station was used for selection and a temporary storage location for the selected documents was allocated on archival shelving adjacent to the work station. Sufficient acid free containers and folders, and labels and other archival equipment were made available.
- Finding Aids Survey. The existing collections database, descriptive lists and indices and other finding aids were surveyed to generate a Preliminary List of collections containing material relevant to the theme.
- Documents Survey. A survey was carried out on a small sample of items from each collection on the Preliminary List of collections in order to record the general types and descriptive characteristics of the documents held in the collections on the Preliminary List.
- Digitisation Database. A relational database in Microsoft Access format was created (the 'Digitisation Database') in order to allow the documentation and control of selection, the recording of descriptive data and the allocation of imaging specifications, and physical batches for the scanning/imaging task later on. This database was relatively simple and was created in-house. For project management purposes fields were included in the database relating to why the document was selected and recording a possible place in the thematic structure, plus an item note with general observations on the contents which may be of relevance to the creation of web content or the digitisation process
- Descriptive Data. It was necessary to record the appropriate Archival-standard descriptive data about each selected document. This information relates to the contents and the context of the document, for example, date, creating agency, the type of archival document, and the provenance of the collection from which it derives. The skills involved in the task of producing, managing and editing this descriptive data are those of a trained Archivist and are governed by professional standards such as ISAD(G). Other areas of the database included physical information, concerning the dimensions, materials, and other physical features of the document, copyright information, imaging specifications and imaging batches.
- Physical Selection of Documents & Data Entry. Individual documents were physically selected from each collection based on the Preliminary List. The list of criteria was consulted during the selection process. It was decided to physically remove and separate selected items from their original storage location.
- Review of Selection and Editing of Data. All selection decisions were reviewed weekly in order to ensure that the criteria were being used properly and that the Project objectives were being met. The descriptive data was also reviewed and edited. Information on the known copyright status of documents was checked and entered. The review stage was particularly important in view of our use of non-Archivist staff for some of the selection task
- Final Selection. A final overall review of selected items was completed and the final set of selected documents was chosen. It was necessary to adjust the number of documents involved based on an estimate of the time required for the other tasks, particularly the Imaging/Scanning task.

2 **Copyright Confirmation Report**

This report concerns how the Copyright issue was dealt with during the Project, firstly with regards to determining and recording the Copyright status of each item/originating collection selected for possible digitisation and secondly in terms of protecting that Copyright during and after the completion of the Project.

- **Copyright & Archives.** The legal basis includes the Irish Copyright Act 2000. Under this Act, Copyright restrictions remain in force for about 70 years from the death of the author of a work, although there are variations for different types of works such as photographs or sound recordings. Copyright is a property right which is distinguished in law from 'ownership' of a work. Copyright of a work may be retained or transferred to another person independent of the issue of the ownership or custody of the work.
- **Process of Determining Copyright.** Once a set of documents for digitisation had been initially selected, it was necessary to check the Cork Archives Institute Accessions Register, which records the circumstances of donation for each collection. This Register records the details of the person donating the collection, and conditions, if any, of the donation, and also whether copyright had been transferred to the Archives. Copyright was either owned by the Archives. not at issue due to age of the item (e.g. over 70 years since death of author), owned by the donor of the collection, owned by another party (e.g. the photographer in the case of a photograph) or copyright ownership was undetermined.
- **Record copyright status in database.** Fields were set up in the Digitisation Database to allow us to record the copyright status of each document/originating collection selected for digitisation. Seek specific permission of copyright owner/ seek copyright holder. This would have been necessary if copyright of any of the material selected was held by a party other than the Archives. The Archives held copyright for all the selected material or copyright no longer applied due to the age of the material. However, if copyright was in doubt it would have been necessary to post notices in newspapers and on the internet and to write enquiry letters in order to make contact with the copyright owner.
- **Protecting Copyright:** By providing information with each image of the ownership of copyright. It should be clear to those using the digital images, whether via the internet or otherwise who exactly owns the copyright. Preventing wholesale, illegal or unauthorised copying or use of items. The copyright protection measures should not interfere with the easy viewing of the digital images, render illegible parts of an image, substantively or negatively alter the physical or visual integrity of the digital image or create additional technical problems for the long term accessibility and preservation of the digital master images
- **Options for Protecting Copyright** including Digital Signature/Invisible Watermark, Digital Encryption, Visible Watermarking, Copyright Notice on Image. Measures chosen to Protect Copyright on the web included a copyright notice saved onto each image and a copyright declaration on the internet site. It was necessary to facilitate those who wish to make or obtain authorised copies for private research or study or another limited purpose such as teaching. In this case a copyright declaration will signed by the requester. The copy will then be supplied either as a print out on paper or as a digital image.

3 Imaging & Handling Specifications

This report details the imaging specifications which were used to digitally image the selected items. The general parameters of the specifications were set down based on professional expertise, research on existing standards and practices, the aims and objectives of the project, the features of the selected items, the capabilities of the imaging equipment available, and handling and conservation requirements. Equipment was selected. This included an A3 scanner, a digital camera, and computer hardware and software. Benchmark imaging specifications were created covering file formats, resolution, pixel dimensions, bit-depth, colour, equipment settings and any operator instructions for the person doing the imaging/scanning. The imaging specifications were then allocated to the selected items by grouping items with similar imaging requirements together. This information was recorded in the Digitisation Database.

A number of imaging specifications were required due to the fact that the physical characteristics and intellectual characteristics of the selected items were not uniform.

1. Establish General Parameters

It was important that the imaging specifications and equipment used would be compliant with the overall aims and budget of the Project and with the scope of the project in terms of

- Aims, budget and scope of the project

Available resources and expertise. Facilitating the task workflow of the Project. Producing master images of adequate quality for both access and preservation purposes.

- Intellectual features of the selected items

Items were examined to see which intellectual features of the original documents would need to be captured in the digital imaging process. Documents must be digitally imaged in a way which preserves the evidential integrity of the item by including all of the original information, and by retaining as much as possible the original order or arrangement of the information. It was important that the text and other information in the final images was as legible as possible, while maintaining the visual features and visual integrity of other parts of the original document. Each document selected for imaging was a complete archival document, for example an account book or letter, however, archival documents may contain more than one single physical item, for example loose enclosures in ledger. Archival records have an 'evidential' quality which should be retained as far as possible in a good digital copy. This evidential quality is formed from a combination of the physical and intellectual features and the context and content of the document, for instance, the layout or organisation of textual information internally, the relationship to other documents in a collection, the size and type of materials and inks used, and the use of seals and signatures.

- Physical features of the selected items

The imaging types for the selected items were either manuscript, printed text/typescript, or continuous tone (Photos). Some items had very small, faded or partially illegible text or fine details, and these required higher resolutions to produce good quality legible images. Most items

contained colour features. Documents varied in size from a single page to a multiple page document such as a large bound volume (see table above). The size of the document would in part determine the number of scans required and the imaging specification needed to capture a document in the most efficient manner possible while retaining a good image quality. Equipment and techniques which could handle items of A3 size and above were required. The original page order and layout of the original needed to be preserved for imaging.

- Capabilities and limitations of imaging equipment

The major constraint to be considered in selecting equipment and imaging specifications was the time and the cost involved in capturing, processing and storing large numbers of larger images.

- Handling and conservation requirements

The equipment or techniques used must not put documents under stress or risk of damage. Handling procedures must be set out to avoid damage. Conservation, including flattening and cleaning would be carried out prior to imaging on items which were creased, folded, or otherwise damaged. This would be carried out by a trained Archivist (see Imaging Methodology).

- Existing Standards and Case Studies

Existing practices, standards and case studies were researched to see if their imaging specifications might be relevant to the Project. Two programmes were identified as being of particular relevance to the present project. The **Electronic Access Project, National Archives & Records Administration and the Library of Congress National Digital Library Programme**

2. Select Equipment

The equipment selected for the Project was as follows:

- ❑ A3 flatbed scanner. To be used for flat and loose documents less than or equal to A3 size. Necessary for photographs, small items and items of limited legibility. The IT Dept of the City Council provided advice regarding the purchase of a scanner. An Epson GT10000 A3 flatbed scanner was selected. It was capable of high resolutions, had a wide dynamic range, a high colour capability and a fast throughput (imaging speed/downloading) using a fast 'SCSI II' connection.
- ❑ High resolution digital camera. To be used for most bound volumes, and all other loose items greater than or equal to A3 size. The Nikon D100 was the camera finally selected for the Project. This is a fully digital professional SLR (single lens reflex) camera. Its features include a high 6.1-megapixel resolution, interchangeable high quality lenses, a full set of photographic exposure controls and autofocus. Also, the camera had the facility to control photographic exposure and storage from the computer using 'Capture Control' software. This software would allow us to send images directly to the computer hard drive, assisting the workflow of the imaging task.
- ❑ Photo. copystand & lights. Necessary for placing of digital camera and support of items during imaging, and to supplement natural light. A large Meopta MG4 copy stand was purchased. The size of the base was approx. 400mm x 350mm, which could be augmented with boards to support items up to about 600 x 500 mm. Two free standing Portaflash 336VM flash units were bought for use in conjunction with natural light to produce correctly-exposed digital photographs.

- ❑ Desktop PC. The platform for capture, viewing, storing and editing of images. Advice was provided at an early stage from the IT Department, Cork City Council, regarding the minimum system required to carry out the Project, the types of systems available and the cost factors involved. A PC PRO Desktop was purchased with a 60Gb hard drive, 512 RAM, 2.4GHz Processor; plus a 19 inch monitor.
- ❑ Tape Backup A magnetic DDS4 (Digital Data Storage) tape drive was installed in the desktop computer before delivery. A single tape could to store 20 gigabytes of uncompressed data.
- ❑ Colour printer with high resolution. To assess print quality. On advice from the IT Dept. an Epson 1290 A3 Colour printer was selected, which had high resolution capability.
- ❑ PhotoShop 7.0 was chosen as the image editing software. The programme is an industry standard which has extensive editing and viewing features including image browsing, metadata, conversion and batch actions.

3. Benchmark Imaging Specifications

This section details the benchmark imaging specifications used in the Project.

The benchmark specifications were initially drafted with reference to the general parameters of the Project, existing standards/practices and the hardware and software environment.

- File Formats & Compression. Uncompressed TIFF version 6.0 (Tagged Image File Format) was chosen as the master image file format. JPEG was chosen as the file format for access images and images on the internet. JPEG is the most widely used web format with high compression and low file size and 24 bit-colour support. The software to produce PDF (Adobe Acrobat 5) format files was purchased. This format is useful for online access purposes, particularly for the online display and printing of multi-page documents.
- Colour & Bit Depth, and Colour Management. It was decided to capture all documents in high colour (24 bit RGB.) To have consistent colour management between the different sets of software and hardware, all colour settings (monitor, digital camera, scanner and image editing program) were calibrated to the Adobe RGB (1998) colour working space. This colour setting provided a wide range of RGB colours, and was suitable for conversion to CMYK colour format for high quality printing purposes. The monitor was calibrated to the Adobe RGB setting, True Colour 32 Bit with a colour temperature set to the recommended 6500K. For those items to be imaged using the digital camera, colour was also a function of lighting and exposure conditions, and these would have to be carefully set up to produce good quality colour images.
- Resolution

300dpi was selected as the main benchmark resolution setting, giving the required mix of quality and smaller file size, with the option of 400dpi on the A3 scanner for a few smaller, faded or less legible items. A 300dpi resolution will capture all but the smallest and least legible items at a level adequate for good quality printing purposes and more than adequate for on-line viewing. It was found that the larger file sizes generated at resolutions higher than about 300dpi added considerably to the time (and therefore the resources) needed to capture, to edit, to manipulate, and to store digital images, and to convert the images for use on the internet.

- Pixel Dimensions

It was decided that the pixel dimensions of the image would be the default pixel dimensions produced when an item was captured at 100% (actual size) at the given resolution (300 or 400dpi).

- Equipment Settings

The equipment also had internal adjustment settings which could effect final image quality. When setting the benchmark specifications it was also necessary to specify the equipment settings for the Monitor, Camera and Scanner, covering settings such as, screen resolution, colour setting, colour temperature, exposure / lighting settings, internal image size, quality and sharpening.

- Operator Instructions

In addition to equipment settings, other software options/other techniques would be used during the imaging process at the discretion of the operator. This included brightness/contrast adjustments, sharpening of the image using 'Unsharp Mask', used at a low level. Natural light to be used for photographic exposure with fill-in flash if necessary. Border/margin to be included in shot if possible. Rotation of image to upright position, cropping of image to establish consistency of appearance across all the scanned images.

4. Allocate Imaging Specifications

The imaging specifications were allocated to individual items by putting each item into an 'imaging specification group' using the digitisation database, containing documents requiring similar specifications and equipment. Any adjustments of imaging specifications could then be made for the group rather than for every item individually. These groups reflected the benchmark specifications but were more specific in terms of the exact specifications, in particular for equipment settings and operator instructions. The following Imaging Specification Groups were used:

- Unbound Document less than A3
- Faded/Fine Unbound Text Document less than A3
- Loose Photographs
- Bound volume
- Bound volume (Soft bound, suitable for scanner)
- Large Bound Volume
- Photographic Album with Text
- Unbound Text/Map Document greater than A3
- Deed Greater than A3
- Outsize Map/ Plans/Posters (> 620 x 430mm)

4 Scanning/Imaging Methodology.

This report describes the methodology and the process of the document digitising (image capture) Task. In order to facilitate a good workflow, items were put into physical batches of items with similar imaging specifications and size/handling characteristics. Basic conservation treatments were carried out including cleaning and flattening. A system of item management was required to track and control the physical items and the digital items. Measures to ensure quality control were also set out. Images were then edited and adjusted minimally to produce the final master images.

1. Prepare Work Areas

The provision of adequate workspace was important in both the workflow and the control of documents. Three spaces were required for the Task.

- Cleaning area; equipped with necessary cleaning equipment. This area must be well ventilated, separate from the rest of the staff area and away from the archives storage area to prevent dust and spores being spread.
- Flattening area; tables large enough to accommodate a number of different types of document including maps and large deeds. Paper weights, archival acid free paper and tissue, and adequate space to circulate material are needed.
- Capture area; this room included the capture equipment arranged in a manner that minimised unnecessary movement, sufficient workspace to prepare documents for digitisation and sufficient storage space (such as shelving) to hold documents awaiting digitisation and those which had been digitised. The room has to be clean and dust free, with adequate electrical sockets, good natural lighting (for photography) and with a good floor to minimise vibrations.

2. Pre-Capture Preparation

This section outlines the pre-capture preparation involved in the digitising task.

- Inspection of Documents, Document Batching and Data Collection. The first phase of the Digitisation Task was the further inspection of each selected document. The inspection had a number of objectives. The first was to establish basic physical information and to update the digitisation database accordingly. Data collected included document dimensions (in mm, overall condition, physical format, handling requirements, materials, conservation requirements, presence of colour, level of fragility, faded text. The second objective of the inspection process was to correct and examine for accuracy the existing archival item description including date, title and content. The final objective of the inspection process was to collect data relevant to the image capture process, specifically the number of scans required and the appropriate imaging specification, which was then entered into the digitisation database.
- Physical Batching of Items. Inspected documents were then grouped into a physical batch appropriate to their imaging specification. This was recorded in the database. A report could then be generated listing the items in each batch for the use of the operator.
- Conservation and Handling, Cleaning and Flattening. To ensure that no harm was caused to the items during the imaging/scanning Task a professional Archivist was employed to carry

out the Digitisation Task. Documents presented a few physical handling problems during the digitisation task, including bound volumes with tight spines, documents which were difficult to maintain in a flat position and the supporting of documents of large dimensions. Appropriate handling procedures were put in place including no inks or any other damaging materials allowed in the areas where original documents were being handled. Light exposure to be kept to a minimum. Documents were only removed from folder/container or cover to allow conservation or imaging to take place and then returned immediately. Fragile documents were identified during the batching process and handled with extreme care during imaging and other handling. Cotton gloves used when handling items and lint-free cotton gloves when handling photographs. Bound volumes were supported by a book support cushion at the point of capture by the camera.

- Conservation Treatment; included dry cleaning with soft brushes, dry flattening and the removal of metal clips and dog-ears. The dry flattening of documents required 4-6 weeks. Parchment deeds were flattened on a large archival workstation table by opening the deed out fully and placing a few sheets of archival tissue and heavy archival wrapping paper over the deed. Other items were flattened on a large archival workstation table under flat metal shelves and heavy bound volumes taken from the collection for this purpose, interleaved with archival tissue and wrapping paper. Small single sheet items required about 1-2 weeks, folded multi-page paper documents required approximately 2-3 weeks, large heavily creased paper items took over 6 weeks.
- Conservation Workflow. Items for conservation were identified from the digitisation database, cleaned and laid out for flattening. Reports based on the batch types were generated from the database. These reports were then manually annotated to record conservation status and location of each item. On removal from its container/folder, the project item number was written on each item. Documents marked as being fragile were identified and, where appropriate, were prepared for other treatments which would not risk damage to the document.

3. Item Capture

This section relates to the process of capturing the digital images of the selected items.

- Management of the digital object: File Naming and Directory Structure; Both the document and scan number had three numeric characters. Three characters were necessary to denote the document number, as more than 100 documents had been selected for the project. Three characters were necessary for the scan number to allow for up to nine hundred and ninety nine scans for each document. Each element was separated by an underscore which would ensure cross-platform compatibility, in particular for use on the internet. In a number of cases the item/original archival document was comprised of a number of individual documents, or 'single items', and the numbering system had to be modified accordingly, to include lower case roman numerals in brackets to denote the 'single item' number. A single directory was created to accommodate all the items associated with the project. Sub-directories were created to cater for the different contents of the project. The most important subdirectory held all the final images. Other subdirectories contained copies of images due for adjustment and metadata.
- Management of the Physical Object: Folders, numbering, digitisation database. Upon selection each document was removed from its original location and placed in a folder, which was annotated with the project item number and a brief description of the document. The folder was then the basic handling unit for the project. Relevant information was also

entered into the digitisation database (See Deliverable No. 1, Appendix 1). Documents were then separated by batch type (see above). Each batch was boxed and each box appropriately labelled. Workflow was based upon the batch as the batching system enhanced workflow by grouping similar type documents. On completion of the imaging, the digitisation database was updated with the scanning status for each document and the number of scans taken for each document.

- **Quality Control.** Image quality was primarily a function of the benchmark imaging specifications, for example resolution, and colour; however, the correct operation and use of imaging equipment and handling of the items during imaging was also important. This included the positioning of items, cleaning of equipment, minimisation of vibration, correct photographic exposure and lighting. A significant amount of photographic knowledge was required to produce quality images using the digital camera.
- **Quality Checking Process.** Constant quality checks were made during the imaging task and covered on-screen quality, fidelity, legibility, brightness, contrast, correct imaging specifications, equipment settings, numbering and saving, file size and pixel dimensions. The print quality was examined periodically for images in each imaging specification/ batch.
- **Typical Data Capture Workflows** for the scanner and digital camera included up to 20 steps covering capture, setting adjustments, quality checking, editing, and saving.
- **Image Editing and Manipulation.** Once the raw TIFF image had been captured at an acceptable level of quality it was saved on the computer hard drive and opened in Adobe Photoshop. The editing procedures included rotating the image to upright screen position, cropping the margins, viewing the image at various screen magnifications, adjusting brightness & contrast slightly, applying auto tone level adjustments, and unsharp mask at a low level to improve the legibility of small or faded text.
- For bound volumes (digital camera) the saving and naming process was automated using built in file saving software in the Nikon Capture Control Software. The auto-file saving and auto-naming facility greatly increased workflow for bound volumes, as it was only necessary to turn the page and to release the camera shutter.
- **Batch Editing/Adjustment.** Adobe Photoshop allowed the recording of an 'action', similar to a macro in Microsoft Word. The action can be applied to a batch of images in a directory. This was particularly useful for larger items if the images of each page were identical in size and visual characteristics.

5 & 6 Tagging/Metadata & Preservation of Digital Master

This report outlines the metadata captured during the Project, and the strategy used to try to ensure the future preservation and accessibility of the digital master images and metadata.

Metadata is essential for making digital images available to users and for making sense of how they are arranged and created. Metadata is made up of descriptive metadata, technical metadata, structural metadata and administrative metadata. Descriptive metadata describes the 'content' in a meaningful way and includes archival descriptive/cataloguing and physical information about the original item. Technical metadata describes the technological specifications and equipment used and included 'embedded' metadata generated during the scanning/imaging Task. Structural metadata describes how the digital master images are arranged and stored and how the metadata fits together with the master data. Administrative metadata is information which concerns the management of the Project and the organisational context of the creation of the data and metadata, and information to facilitate future migration and any changes to the data.

The digital preservation strategy had the following elements: Physical preservation; and maintenance of long term accessibility and usability.

The physical preservation strategy involved saving the data (and metadata) onto reliable media, selecting appropriate storage conditions and equipment, the use of multiple redundancy and distribution of copies and the setting of long term physical maintenance and migration procedures.

The long term access strategy involved measures to minimise the effects of hardware and software obsolescence and organisational change so that the digital master images would remain useable for a long period. The concentration was on using non-proprietary and platform independent formats as much as possible. It was noted that the data would require migration into another digital format at some stage in the future.

Introduction

Outlines of some of the issues involved in digital preservation, challenge of keeping this data accessible to a constant degree over time is not just a matter of preserving the original data files but also of extensively documenting the technological and administrative context of their creation and of having a strategy for overcoming hardware and software obsolescence.

METADATA

Description of the types and contents of Metadata captured in the Project including:

- Descriptive data concerning the intellectual content of the Project and the documents captured;
- Technical data on the imaging specifications, the equipment settings and file formats used;
- Structural data on accessing the documents, their arrangement and the framework within which the descriptive data was created;
- Administrative data for future management of the collection

- ‘Meta-metadata’ describing the content and structure of all the other metadata files.

All metadata files were converted into ASCII files using the “Export “ facility within the MS Access software.

Physical Preservation & Long Term Access/Migration Strategy

This section outlines the steps taken to try to ensure the physical preservation of the digital master images and metadata and their long-term accessibility.

Storage Media

Media selected for preservation, CD-R (Compact Disk Recordable) and other options. Optical media, in particular the CR-ROM and CD-R format are used widely to store master copies of important data due to the fact that CD’s are less susceptible to accidental data erasure and many environmental effects than magnetic tape.

CD-R (Compact Disk Recordable) was chosen as the selected media because of the following features: Protected WORM (Write Once Read Many) format: Data difficult to erase. Reasonable physical stability: The CD, if written, stored and handled in the appropriate ways is likely to last at least until the data needs to be migrated to another medium in 5-10 years. Prospect of medium/long term hardware accessibility/readability: The technology is an industry standard and in very widespread use. Budgetary and Project considerations: CD-R drives and disks are quite inexpensive and easy to use.

A defining requirement for long term digital storage is that the storage medium (or media) chosen can reasonably be assumed to have a reasonable life expectancy within the review date for migrating the data to a new medium.

CD’s produced by a major and well-established quality manufacturers are seen as better than those produced perhaps more cheaply by less well established manufacturers. Gold and silver based CD’s appear to be a safer bet, due to their stability however it is not clear if the high cost of special Gold-based media is justified due to the lack of open standards regarding their longevity and the fact that CD-R technology is likely to become obsolete well before the disk itself physically deteriorates.

Writing the Master Copies

- The manufacturers recommendations from both the CD and the CD writer were implemented, regarding handling and write speed.

Copies/Levels of Redundancy/ Physical Distribution

- It was decided to have 3 sets of master copies of the data on CD-R and to store the copies in different locations. 2 sets will be designated ‘Master Copy 1’ and ‘Master Copy 2’, one to be stored at the Archives Institute and one in the City Library. These sets will not be used except for storage/testing and migration purposes. The third set is to be a ‘Working Master Copy’, which also to be stored at the Archives Institute and used for access purposes where a derivative copy of a high quality master image is required.

- In addition to the above, Access copies had been created (in JPEG format) for use by public or staff in the course of research, or for an outreach, education or exhibition purpose. Copies of the JPEG images are also held on CD-R.

Storage Conditions

- Storage containers. The CD-R masters are kept in inert archival polyester pockets stored in specially designed flip-top acid-free boxes available from a company in the U.K. Each box can hold up to 50 CDs and each pocket has an index label/card, so writing on the CD itself was not necessary.
- Environmental conditions. It is hoped to store the CD's under standard archival environmental conditions, at a constant humidity in the range 40-55% and a constant temperature in the range of 16-18 Degrees Celsius.

Long Term Access/Migration Strategy

- CD Testing/ Technology Review. A sample of the CD's will be tested every 2 years to check that the data and metadata is intact and accessible, to be made part of the preservation management programme of the Archives Institute. A review will also be made every 2 years of the wider technological environment to identify trends which may impact on the accessibility of the stored data.
- Hardware & Software Migration Framework. Physical migration (Media migration) will take place at that time where there is a risk of the existing format/technology becoming completely inaccessible or obsolete within 5 years or if the storage media (CD) appears to be deteriorating resulting in possible data loss or possibly if new media becomes available which is independently guaranteed, according to international standards, to offer long term/permanent data preservation. At migration, the data in its entirety, and including all metadata will be migrated according to the best available international standards. A complete record of all migration and maintenance activities will be kept and included in a separate metadata file.

File Formats

Outlines choices of file format for Image Files (Content) and Metadata Files. The choice of file format may have a major effect on future accessibility and preservation of digital records. This means that the chosen formats must be portable across different operating systems, allowing greater flexibility in accessing data at a later date and must not be tied to a particular vendor, as vendors may go out of business or discontinue support for particular software.

Three criteria were identified as being relevant to identifying the best formats for long-term storage. 1. Open source - the relevant code is accessible to all. 2. Platform Independent – the format can function on any operating system. 3. International or Industry (de-facto) Standard – the format has achieved wide acceptance in the industry

TIFF 6.0 (Tagged Image File Format) (Uncompressed version) was the image file format chosen by the project for long term storage.

ASCII flat text (American Standard Code for Information Interchange and International Standard ISO 646.) was chosen as the format for the metadata files due to the simple and platform independent nature of the format. 'Flat Text' means that the file is saved as a basic text file, without word processor formatting.

File Naming

ISO 9660 "Information Processing – Volume and File Structure of CD-ROM for Information Interchange" was looked at for file naming on CD's. The current Project was compliant with ISO 9660 Level 1; a basename of eight alphanumeric characters and an extension of three alphanumeric characters separated by a required full stop. Level 1 is compliant with the file naming rules for the major operating systems (Windows, Linux, Mac OS, Unix).

7 Report on Web Content Creation

This Report details the creation of web content. A sample of images suitable for use on the web site were copied from the master images based on Project objectives and information gathered during previous tasks. Textual/Descriptive information captured during the selection process, was used for the HTML/ web version. Web-ready images were created in JPEG format

Selecting Items for the Web. Over 340 digital images and associated metadata/descriptive data was generated by the Project. According to the aims of the Project it was decided to use a representative sample of these items which would include an example of most of the range of physical or intellectual types of archival documents from Estate Papers, and in particular items of interesting and appealing visual or intellectual content.

Web Site Structure:

- Correspondence/Letters
- Photographs
- Maps, Plans and Drawings
- Deeds and legal documents
- Account Books and Rentals
- Estate Reports and Sale Catalogues
- Other Records & Ephemera (Including posters, notices and a few printed items)

Creating Web Content

The textual descriptive content about each item for the web was mainly generated during the selection process. Also recorded in the database at this time was a note concerning any interesting features about the Item. This text had already been extensively corrected and edited and was ready for use on the web site. It was 'cut and pasted' directly from the database into the HTML page.

Other web content, such as introductory text and links and descriptions of extract images was created during the Web Content Task.

HTML

The HTML pages were created using Microsoft FrontPage Express from the given Pilot Project Site Templates. This involved downloading a blank copy and inserting the relevant text and images, including the relevant heading and header images.

Images

Many of the items selected for digitisation in the Project required more than 1 digital image to represent them, for example a bound volume of accounts or a multi-page letter. It was decided to use in most cases a single image on the web as an illustration of each item, rather than attempting to place entire multi-image items online. One extract of a few pages of a multi-image item was created in PDF format. (See File Formats Below)

Production of Web Ready Images

It was decided to produce web images which would have the lowest download time while maintaining quality, colour fidelity and legibility of textual features.

- **File Formats:** JPEG format, Adobe Acrobat (PDF) Format. Most images for the web were in JPEG format. Adobe Photoshop 7.0 was used to create the images, and has a good 'Save for the Web' capability. A small set of images were delivered using Adobe Acrobat (PDF) Format. This format can combine more than 1 image in the same file, and is useful for displaying multiple-page documents. A series of JPEG images were directly imported into a PDF file.
- **Resolution:** Reduced to 72 dpi; Reduce Image (Pixel) Size by around 50-70%. Most images selected could be reduced by about 50% while still being fully legible on screen. Images scanned at the higher 400dpi resolution had higher pixel dimensions and could be reduced by an even greater percentage, often 70-80%.
- **Compression setting:** Best mix of low file size and good quality. (Range of compression 5:1 - 50:1). Sometimes a high level of compression used on a larger image pixel size produced more legible results than low compression on a smaller-sized image.
- **File size:** always less than 100K, and mostly in the range 10-60K

Thumbnail Images. Smaller 'thumbnail' JPEG images of each item were generated from the master image, normally at a size of about 5 or 10% of the original for inclusion on each content page. It was decided to produce more detailed thumbnails in order to give users a flavour of the appearance and contents of each item by just scanning a content page. The user would have the option of viewing a larger image, but it would not be absolutely necessary. In fact, it was hoped that these small but not completely illegible images are quite enticing because they are semi-legible and will hopefully encourage the user to download the larger item.

Linking and Presentation Each content page consisted of a series of images with the attendant descriptive text. A link was placed on each small thumbnail image to the larger image with the image text 'Click for Larger Image'. The descriptive text was either wrapped around the image or a table was made up in HTML with 2 columns, images on the left and text on the right. The table tended to produce the best layout best with sections containing larger thumbnail images and those images with shorter descriptive texts.

Outline of Issues Associated with Particular Types of Item/Document.

- **Correspondence/Letters.** Long text descriptions, wrapping the text around the thumbnail image more satisfactory than HTML tables. Proved difficult in a few of these images to produce a legible, good quality JPEG with a low file size due to faded text in the original document.
- **Photographs** produced the most satisfactory results on the web in terms of low file size and quality. The JPEG file format is particularly suited to continuous tone items such as photographs, and high compression ratios were used with little loss of quality.
- **Maps, Plans and Drawings.** It proved time-consuming to produce a legible, good quality JPEG with a low file size which would fit on-screen, due to the small or less-legible visual and textual features in the original maps. Intricate colour features were easily lost using high compression ratios. Some of the resultant JPEGs tended to be slightly larger than the screen, necessitating the use of the scroll bars. In the end, some of the items appeared more satisfactory as physically smaller but less compressed images.

- **Deeds and legal documents.** Generally large documents with small, dense text. Posed major problems for producing satisfactory web-ready images, due to the delicate legibility of the text in many of the master images. It was decided to use extracts of the main image in most cases, resulting in small, legible images, although the thumbnail images were usually derived from the original image as a whole.
- **Other Types** Account Books and Rentals, Estate Reports and Sale Catalogues, Other Records & Ephemera (including posters, notices and a few printed items). Most of these items posed few except for similar problems as those with deeds for example where the original document was quite large with delicate legibility.

Testing and Evaluation

Uploaded pages and images examined live on the Web in the Internet Browser for final editing and evaluation. Download speeds, appearance, and ease of navigation and use were noted and then certain pages and images were revised.

Evaluation

- The Project overall was successful, with all master images of the selected items being created and stored and the vast majority of the metadata and web content created within the required timescale.
- The number and wide range of items selected may have been slightly over-ambitious for a 6 month Pilot Project of this kind, given that there is little detailed information available on the specific solutions used in imaging the types of archival document used in the Project.
- Project management might have been more successful in the area of reporting, with delays mainly due to unforeseen circumstances. In retrospect, it would seem that such a Project demands the exclusive attention of a Project Leader as well as project Staff and a Project Manager who is not too heavily committed to other work in the Institution
- The selection task was much more vital to the success of the Project than assumed at outset. This was partly because much of the descriptive data/metadata was captured at this stage, greatly facilitating other tasks such as the metadata task and the creation of the web content
- The physical tracking of items during the imaging/digitisation task required a good deal of thought , preparation, and monitoring, despite the fact that items were removed from their original locations and numbered directly and placed in folders, and subsequently put into 'batches' based on their imaging specification. This seems to have been due to the fact that many items required cleaning and flattening, but the treatment techniques required varied slightly and the original documents responded at different rates to flattening. Although items were ordered within each 'batch', in order to maintain work flow items not ready for imaging were bypassed, thus complicating the task of knowing at all times where each item was in the system.
- Although technical issues were expected to be the most dominant concern, many issues which arose were of a practical and administrative nature, for example, setting out the work areas and storage space for the imaging task, allocating time and work flow procedures and monitoring progress.
- The Project relied heavily upon existing Archival knowledge, methods and standards. It was revealing that this professional knowledge was so vital in such a technology-dependent Project, for example, in the area of handling and conservation, and also the degree to which technology can broaden and facilitate the application of this knowledge, for example, in the area of creating metadata and in digital preservation.
- It was difficult to implement exactly all imaging specifications and procedures during the imaging task as very many 'small' technical issues arose which were difficult to anticipate; for example, the range of built-in imaging settings to use with the given equipment and the use of corrective filters such as 'Auto Levels' or sharpening. A good deal of testing and experimentation was needed to finalise procedures in this regard.
- The specific task of imaging/digitisation proved quite speedy once minor issues were solved. In particular, the ability to control camera settings and to release the camera shutter directly

from the PC greatly facilitated work flow when imaging multi-page documents such as bound volumes.

- The existing basic technological knowledge and training of staff proved greatly beneficial in completing the Project. This included basic keyboard skills, knowledge of essential for the successful completion of a Project.
- The overall equipment infrastructure proved more than adequate for the vast majority of the tasks and the imaging requirements of documents selected for digitisation. The high specification PC proved worth the investment and full use was made of its storage capacity and processing speed; for example the browsing of the master image directories through Adobe PhotoShop would have been quite slow on a lower-specification PC.
- The process of judging image quality was more subjective than anticipated, and staff involved in the process did not necessarily agree all the time. The number of variables associated with this task would seem to explain this fact, such as room lighting, equipment settings, software settings, corrective settings, and the nature of the original document.
- The learning curve when using the digital camera was extremely high, and required a good deal of basic photographic knowledge about exposure, lighting, lens settings, and so on. Staff did have some knowledge in the area, however in retrospect a good deal of time was used on getting to know how to use the camera and other photographic equipment competently. The advice of a professional photographer might have been a good idea.
- Editing of descriptive data was a more painstaking and time-consuming task than anticipated. This was in part due to the use of non-professional staff to draft the initial descriptions. In retrospect, more intensive training should have been provided here or the task should have been carried out by an Archivist.
- The capture and preservation of the metadata proved to be a relatively simple task using the digitisation database to record the information and export it to a platform-independent format.
- The non-linear nature of the Project was striking. Although the project was divided into Tasks, these often overlapped greatly in time and were greatly interdependent. An example is the need to have a good knowledge of the technical issues of digital preservation when drafting imaging specifications and purchasing equipment.

Recommendations

Project Management

- Digitisation is not a strictly linear process, as tasks overlap to a great degree. It is best if the Project is planned out as completely as possible, based on best practice, financial and staff resources, and the availability and capability of equipment.
- Well-defined Project objectives and parameters are essential in applying selection criteria, benchmarking imaging specifications, and creating web content.
- It is not feasible to undertake a larger scale project without analysing before hand the necessary procedures to maximise workflow/throughput. Everything from the batching of the documents to the configuration of the equipment must be addressed.
- Awareness of the technical aspects of imaging is important from the start. For example, the physical features of the document as recorded during selection have an influence on the setting of scanning/imaging requirements and the image processing and storage capacity of equipment.
- A detailed estimate of the time involved in producing finished digital images of the types of original documents found in the chosen collections is required at an early stage to facilitate planning and resource allocation.
- A ‘digitisation database’ is a very useful tool for facilitating task workflow and project management. The database may be used to generate reports and queries which are invaluable to each task. The database should be relational and may contain:
 - archival descriptive data (item and collection level)
 - some metadata (technical specifications, equipment settings, etc.)
 - physical data (dimensions)
 - data on imaging specifications
 - a system of allocating each item to a ‘batch’

(Image database management systems are often built into commercial archival database software products.)

- Tasks in order of time required (Greatest First)
 - 1 Selection
 - 2 Scanning/Imaging
 - 3 Project Management
 - 4 Web content, creation, uploading and testing
 - 5 Benchmark Imaging Specifications;
 - 6 Digital Master Preservation; a technological minefield
 - 7 Tagging/Metadata
 - 8 Copyright

- A digitisation Project demands the exclusive attention of a Project Leader as well as project Staff for the entire duration of the Project. The Project Leader and Project Manager should be professional archivists, and it is desirable that other staff would also be professionally trained in Archives. The Project Leader must have a good deal of knowledge and preferably good practical experience in the area of digitisation and information technology. The Project Manager or other staff must not be too heavily committed to other work in the Institution.
- Any prior experience or expertise of staff in the area of information technology or digitisation will be of great benefit in the smooth running of the project.
- The expertise of other professionals in the following areas may prove invaluable:
 - Selection and sourcing of IT Equipment
 - Specialist repair and conservation of archives
 - Data storage systems
 - Web design
 - Digital preservation
 - Metadata
- The IT department of the project institution may provide valuable advice and assistance.
- Tasks in order of Cost (Greatest First)
 - 1 Scanning/Imaging: Equipment and Staff
 - 2 Selection: Staff
 - 3 Digital Master Preservation: Equipment
 - 4 Benchmark Imaging Specifications: Staff
 - 5 Web content, creation, uploading and testing: Staff
 - 6 Tagging/Metadata

Selection

- Well-defined Project objectives and parameters are essential in drafting selection criteria and applying selection criteria
- It is useful to divide selection criteria into those which are essential in order to fulfil the Project objectives and other criteria which might influence the selection decision but were not decisive.
- The conservation and preservation aspects of selection are critical when working with original archival documents. This means the use of appropriate archival staff, equipment, training, and the setting of firm policies and procedures.
- Having a set of existing finding aids, indices or data about collections/documents is a help in the selection process. It was apparent that the digitising of an entire archival collection would require that the collection is fully catalogued, with the creation of a full archival descriptive list.

- Some form of electronic digitisation database is an important tool for the selection task. Its design should be determined by archival standards and the need to manage the task workflow by also recording information about the physical features and later the imaging specifications of the documents. The database may be used to generate reports and queries which may prove valuable in making a final selection decision.
- Awareness of the technical issues of imaging is important during the selection process in order to facilitate the workflow of the project as a whole. For example, the physical features of the document as recorded during selection would have an influence on the setting of scanning/imaging requirements. Capturing as much data about the document as possible during selection also meant that there is less handling of documents.
- A time and motion study for producing finished digital images of the types of original documents found in collection may be required before a final list of selected items can be made.
- The selection task is extremely time consuming may require the input of a number of staff members.

Copyright/ IPR

- It is important to establish, verify and record the copyright status of each item selected for digitisation
- The Accessions Register of the archives service may record copyright ownership details
- Staff should have an up to date knowledge of the legal issues involved. Advice of fellow professionals or legal advice should be sought.
- Where the copyright owner of an item is not known, every effort should be made to establish ownership.
- The level of copyright protection used should be appropriate to the material being digitised, the method of delivery, and the requirements of the end user. Any technology-based solutions, for example encryption, may effect the long term preservation and accessibility of the digital master copies.
- A copyright/legal declaration on the web site may, for example, declare copyright ownership and permissions required for use of copies for other than private research or study
- It is relatively simple to place details of copyright within individual images by inserting text into the 'tags' in an image file and/or by saving a visible copyright notice as text within each image using image editing software.

Imaging Specifications & Equipment

- A networked computer system would be most desirable where more than one Staff member is involved in the Project and where electronic images and files are being shared.
- Equipment and specifications must be powerful and wide ranging enough to be able to cope with a variety of physical and intellectual features and workflow issues.
- A benchmark resolution of 300dpi produces high quality master images in most cases, with relatively small file sizes. 400dpi or higher is required for items with less legible text.
- A digital camera with high digital (6 megapixel) resolution and high quality lens and with the option of interchangeable lenses, will be capable of quality results in most cases, except for documents of a dimension greater than about 500mm x 500mm, which may require professional photography or the use of another type of equipment such as an overhead scanner.
- Higher resolutions produce larger file sizes which require more processing time and more storage capacity
- It is useful to estimate the range of file sizes and then to assess the storage and processing capacity of the equipment in relation to that estimate.
- A digitisation database is a useful tool for allocating imaging specifications, allowing reports and changes to be made quite easily.

Imaging Methodology

- Adequate and well-organised working and storage space is an essential component of the management of items and also of staff health and safety.
- Software which allows immediate download of images and direct control of the digital camera from the computer is useful
- The functionality offered by leading image processing packages is worth the investment, particularly in relation to:
 - range of image adjustments available
 - browsing and saving images and copies of images
 - viewing and controlling tagging/metadata
 - use and changing of filenames
 - creating derivative images for use on the web
 - batch processing and automated features
- A quality assessment needs to be made for every image produced.

- Colour management is required to produce consistent results, which means setting or calibrating all equipment and software to the same colour specification
- The ease of file name saving in modern computer systems make it essential to have a system of file naming that allows errors to be easily identified.
- The conservation and preservation aspects of the digitisation task are of paramount concern when working with original archival documents.
- Documents should not be put at risk. The use of professional archival staff, appropriate archival equipment and techniques, the training of other staff in the use and handling of archives, and the setting of firm policies and procedures according to best archival practice and standards are the best means in this regard.
- Handling of documents should be kept to a minimum. Capturing as much data about the document as possible during the selection task may mean less handling of documents during the digitisation task.
- The use of a professional conservator is required to carry out any specialist paper repair, binding, and to advise on special damage treatment and flattening techniques.

Metadata and Digital Preservation

- A number of metadata types are generated in a digitisation project. These are, descriptive data concerning the intellectual and physical content of the Project, for resource discovery; technical data on the imaging specifications, the equipment settings and file formats used; structural data on data arrangement and the framework within which the descriptive data was created; administrative data for project and preservation/migration management; plus 'meta-metadata' describing the content and structure of all the other metadata files.
- Resource discovery metadata may be based on the Dublin Core Metadata Element Set (DCMES) which is the most widely used schema used to described internet resources to allow for accessible use of the collection on the World Wide Web. The Irish Public Service Metadata Standard (IPSMS) should also be taken into account.
- Technical metadata created includes that generated automatically by the capture process and that created deliberately as part of the digitisation process. It includes imaging specifications, which describe the resolution, pixel dimensions, bit depth, file format, colour settings formed part of the digitisation database. The description of the equipment and software used, make, model may need to be captured, for example, as part of the digitisation database.
- Metadata not embedded in the master images should be converted to a platform independent format (such as ASCII text) and combined to form a single technical metadata file. This file is then preserved with the master images.

- The issue of digital preservation and long term access must be taken into account during all tasks in the digitisation project. Particularly important are the selection of imaging specifications and equipment and the capture of metadata. The types of file formats and file naming systems used are also of vital concern.
- The challenge of keeping data accessible to a constant degree over time is not just a matter of preserving the original data files but also of extensively documenting the technological and administrative context of their creation and of having a strategy for overcoming hardware and software obsolescence. It is likely that any configuration of storage medium, technology for reproducing the data stored and software chosen to preserve digital data will be obsolete at some stage in the near future. It is therefore necessary to have a migration strategy which manages putting the data onto another media and/or converting into accessible file formats.
- An Access and Preservation Strategy must be drawn up to ensure that the chosen media are stored correctly, handled correctly, regularly checked and that advances in technology are monitored and that includes a migration strategy for the physical media and/or file formats.
- File formats chosen must be platform independent with wide industry acceptability. TIFF is an accepted standard for master image files. For metadata preservation and access purposes, flat ASCII text files remain a good choice given its high level of platform independency and the relative ease of creating simple text files. XML is the likely to be used increasingly for data exchange and should also be considered.
- Optical CD-Recordable disks are a relatively inexpensive and reliable way of storing master data, provided multiple copies are made, the masters are stored in appropriate conditions, and once a long term preservation and access strategy is implemented based on the likelihood of migration in the next 5-10 years. It is important that good quality CDs (ideally silver or gold based CDs) from a recognised manufacturer are used and that the CD's are written according to the manufacturer's recommendations
- The lack of international standards concerning all aspects of digital preservation means that the Archivist must rely on the experience and best practice of other projects and de facto technical standards such as TIFF 6.0, and ASCII.

Web Content

- The web content task is facilitated by the capturing of descriptive data during the selection task.
- Web content may be enhanced by noting during the digitisation process any particularly interesting visual or content features of items, and employing these as visual features on the web site.

- Textual descriptions about each item web content may be taken directly from the descriptive archival data (for example information held in the digitisation database, or a Descriptive List of a collection)
- General and introductory information must also be added
- The structure, navigation and presentation of the web site should be maximised for the end-user.
- Images for the web are usually derivatives of master images. Commercially available image processing software has a high level of functionality for producing web-ready images.
- JPEG format is the most widely used image format. PDF / Adobe Acrobat might be considered as an alternative for delivering multi-paged items.
- It is possible to produce web content, such as JPEG images and thumbnail images, and even an entire set of HTML automatically using a built-in tool in Adobe PhotoShop 7.0.
- Larger and less legible items such as deeds may present problems for online delivery. In this case the use of extracts or portions of the full image should be considered.

References

Selection

Archives On-Line Report

National Council On Archives 1998 (UK)

<http://www.archives.org.uk/publications/archivesonline/aolintro.asp>

Arms, C. 'Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress' Part I *D-Lib Magazine* (April, 1996, <http://www.dlib.org/dlib/april96/loc/04c-arms.html>); Part II in *D-Lib Magazine* (May, 1996, <http://www.dlib.org/dlib/may96/loc/05c-arms.html>).

Arts and Humanities Data Service

In particular the Managing Digital Collections section. (<http://ahds.ac.uk/manage/manintro.html>)

Ayris, P. 'Guidelines for Selection' compiled by P. Ayris (UCL) as part of the joint RLG and NPO Preservation Conference, Warwick, 1998

Columbia University 'Selection Criteria for Digital Imaging Projects', <http://www.columbia.edu/cu/libraries/digital/criteria.htm>

Council on Library and Information Resources (<http://www.clir.org/>) - CLIR. Various reports and publications.

Cornell University Library Research Department

Moving Theory into Practice *Digital Imaging Tutorial*.
<http://www.library.cornell.edu/preservation/tutorial/>

European Commission on Preservation and Access (ECPA)

<http://www.knaw.nl/ecpa/>

Forde, H. The education of Staff and Users for the Proper Handling and Care of Archival Materials: A RAMP study with guidelines ,1991, UNESCO, Paris
<http://www.unesco.org/webworld/ramp/html/r9117e/r9117e00.htm>

Harvard University 'Selection for Digitizing: A Decision-Making Matrix',
<http://preserve.harvard.edu/resources/digitization/selection.html>

Hazen, D., Horrell, J., Merrill-Oldham, J.

Selecting Research Collections for Digitization (European Commission on Preservation and Access -report; 08) 1998

University of California 'Selection Criteria for Digitization',
<http://www.library.ucsb.edu/ucppp/digselec.html>

Lee, Stuart D.

Scoping the Future of the University of Oxford's Digital Library Collections

Final Report September 1999

Esp. Appendix G: Assessment Criteria for Digitization

Library of Congress American Memory Project and National Digital Library Program

<http://lcweb2.loc.gov/ammmem/award/docs/docimqr.html>).

National Archives and Records Administration's Electronic Access Project

(<http://www.nara.gov/nara/>), especially *Guidelines for Digitizing Archival Materials for Electronic Access* (<http://www.nara.gov/nara/vision/eap/eapspec.html>).

National Preservation Office, The British Library

Managing the Digitisation of Library, Archive and Museum Materials

www.bl.uk/npo

Smith, A. 'Why Digitize?' and 'The Future of the Past: Preservation in American Research Libraries' (1999 - <http://www.clir.org/pubs/reports/reports.html>).

Technical Advisory Service for Images (TASI - <http://www.tasi.ac.uk/>) esp. Overview and Project Management.

UNESCO Archives Portal

Primary Documents Online : Projects

http://www.unesco.org/webworld/portal_archives/pages/Primary_Sources_Online/Projects/

Yola de Lusenet Preservation of digital heritage

Draft discussion paper prepared for UNESCO

European Commission on Preservation and Access

March 2002

<http://www.knaw.nl/ecpa/PUBL/unesco.html>

Copyright

Copyright and Related Rights Act, 2000 (Stationery Office, Ireland)

Canadian Heritage Information Network

'Protecting Your Interests: A Legal Guide to Negotiating Web Site Development and Virtual Exhibition Agreements'

www.chin.gc.ca/english/intellectual_property/acknowledging_solidifying_ip_rights.htm

'The Virtual Display Case: Making Museum Image Assets Safely Visible'

www.chin.gc.ca/english/intellectual_property/virtual_display_case/index.html

Irish Patents Office

www.patentsoffice.ie/legislation-copyright.html

Library Association of Ireland

www.libraryassociation.ie/policy/copyright.htm

Scoilnet Web Publishing Centre - Copyright
www.ncte.ie/wpc/plan_copyright_details.html

NUI Maynooth Library
www.may.ie/library/services/copyright.shtml

Technical Advisory Service for Images (UK)
<http://www.tasi.ac.uk/advice/managing/pdf/copyright.pdf>
www.tasi.as.uk/advice/managing/copyrights.html

Imaging Specifications

Research Department, Cornell University Library. 'Moving Theory into Practice Digital Imaging Tutorial'

<http://www.library.cornell.edu/preservation/tutorial/>

Digitisation Method (Stuart D. Lee, Mar 1999)
<http://www.bodley.ox.ac.uk/scoping/digitization.html>

Steps in the Digitization Process (Carl Fleischhauer, Library of Congress, National Digital Library Program, Jan 1996)
<http://lcweb2/ammem/award/docs/stepsdig.html>

Quality Review of Document Images (Library of Congress, National Digital Library Program, Sep 1996)
<http://memoryt.loc.gov/ammem/award/docs/docimgr.html>

Digital Imaging (Minnesota Historical Society)
<http://www.mnhs.rog/preserve/records/electronicrecords/erdigitalimaging.html>

NARA Guidelines for Digitizing Archival Materials for Electronic Access (Jan, 1998)
http://www.archives.gov/research_room/arc/Arc_info/guidelines_for_digitizing_archival_materials.pdf

Technical Advisory Service for Images (UK) (TASI - <http://www.tasi.ac.uk/>)

National Preservation Office, The British Library
Managing the Digitisation of Library, Archive and Museum Materials
www.bl.uk/npo

UNESCO Archives Portal

Primary Documents Online : Projects

http://www.unesco.org/webworld/portal_archives/pages/Primary_Sources_Online/Projects/

European Commission on Preservation and Access (ECPA)

<http://www.knaw.nl/ecpa/>

Arts and Humanities Data Service

In particular the Managing Digital Collections section (<http://ahds.ac.uk/manage/manintro.html>) with its series of reports.

Council on Library and Information Resources (<http://www.clir.org/>) - CLIR.
Research Libraries Group (RLG). Incl. Guides to Quality in Visual Resource Imaging.

Mobile Computing Online <http://www.mobilecomputing.com> esp. hardware.

Imaging Resource <http://www.imagingresource.com> esp. reviews of hardware

Nikon USA www.nikonusa.com

PhaseOne - large format digital cameras www.phaseone.com

ICAM Archives Systems (UK) microfilm and imaging equipment
www.icamarchive.co.uk

Photologic Ltd (Kodak Cameras) www.photologic.ie

Lambeth Palace Library Church Plans Online digitisation project
www.lambethpalacelibrary.org

Imaging Methodology

Digital Imaging (Minnesota Historical Society)

<http://www.mnhs.org/preserve/records/electronicrecords/erdigitalimaging.html>

NARA Guidelines for Digitizing Archival Materials for Electronic Access (Jan, 1998)

http://www.archives.gov/research_room/arc/arc_info/guidelines_for_digitizing_archival_materials.pdf

Request Proposals For Digital Images from Original Documents
Text Conversion and SGML-Encoding
(National Digital Library Program, Library of Congress, Jun 1996)

<http://memory.loc.gov/ammem/prpsal/coverpag.html>

Research Department, Cornell University Library. 'Moving Theory into Practice Digital Imaging Tutorial'

<http://www.library.cornell.edu/preservation/tutorial/>

Digitisation Method (Stuart D. Lee, Mar 1999)

<http://www.bodley.ox.ac.uk/scoping/digitization.html>

Steps in the Digitization Process (Carl Fleischhauer, Library of Congress, National Digital Library Program, Jan 1996)

<http://lcweb2/ammem/award/docs/stepsdig.html>

Quality Review of Document Images (Library of Congress, National Digital Library Program, Sep 1996)

<http://memoryt.loc.gov/ammem/award/docs/docimgr.html>

Digital Imaging (Minnesota Historical Society)

<http://www.mnhs.rog/preserve/records/electronicrecords/erdigitalimaging.html>

NARA Guidelines for Digitizing Archival Materials for Electronic Access (Jan, 19980)

http://www.archives.gov.research_room/arc/Arc_info/guidelines_for_digitizing_archival_materials.pdf

Technical Advisory Service for Images (UK) (TASI - <http://www.tasi.ac.uk/>)

National Preservation Office, The British Library

Managing the Digitisation of Library, Archive and Museum Materials

www.bl.uk/npo

European Commission on Preservation and Access (ECPA)

<http://www.knaw.nl/ecpa/>

Arts and Humanities Data Service

In particular the Managing Digital Collections section (<http://ahds.ac.uk/manage/manintro.html>) with its series of reports.

Council on Library and Information Resources (<http://www.clir.org/>) - CLIR.

Research Libraries Group (RLG). Incl. Guides to Quality in Visual Resource Imaging.

Lambeth Palace Library Church Plans Online digitisation project

www.lambethpalacelibrary.org

Metadata and Digital Preservation

An Introduction to Digital Preservation (Technical Advisory Service for Images, 2002)

<http://www.tasi.ac.uk/advice/delivering/digipres.html>

“Preservation and Migration of Electronic Records: The State of the Issue” Kenneth Thibodeau (NARA, 09 Jan 2003)

http://www.archives.gov/electronic_records_archives/papers/print_friendly.html

An Approach to the Preservation of Digital Records

Heslop, Helen et al (National Archives of Australia, Dec 2002)

Preservation of Digital Heritage
(de Lusenet, Yola, UNESCO, Mar 2002)

Digital Data Preservation (National Institute of Standards and Technology, Sep, 2001)
<http://www.itl.nist.gov/div895/datapreservationmain.html>

How to Preserve Authentic Electronic Records
(InterPARES Preservation Task Force, 08 Oct 2002)
Image Formats for Preservation and Access (Lesk, M., Jul 1990)
<http://www.palimpsest.stanford.edu/byauth/lesk/lesk.html>

Image Standards and Implications for Preservation (Battin, P., Jun 1992)
<http://www.palimpsest.stanford.edu/byauth/battin/imagestd.html>

The Unofficial TIFF Home Page
<http://www.earthlink.net/~ritter/tiff>

TIFF – Revision 6 (Adobe Developers Association, 03 Jun 1992)
<http://www.adobe.com/Support/TechNotes.html>

Ms Access Speaks HTML (Webreview, 04 Feb 2002)
http://www.webreview.com/2002/02_04/developers/index01.shtml

Extensible Markup Language (XML) (W3C Architecture Domain, 05 Feb 2003)
<http://www.w3.org/XML>

Cross Platform compatible files – File Naming Issues
<http://www.imagemontage.com/Docs/FileNames.html>

ASCII – American Standard Code for Information Interchange
<http://www.encyclopedia.com/html/A/AS1C11111.asp>

ISO 9660

Storage Media Life Expectancies (John Van Bogart, National Media Laboratory, Jun 22, 1998)
<http://ssdoo.gsfc.nasa.gov/nost/isoas/dads/presentations/VanBogart/VanBogart-Unabridged.ppt>

Metadata And Digital Images (Technical Advisory Service for Images, 2002)
<http://www.tasi.ac.uk/advice/delivering/metadata.html>

Metadata and the Management of Current Records in Digital Form, (ICA/CER – Committee on electronic and other Current Records, Jul 2000)

Creating IPSMS Metadata
<http://ww.gov.ie/webstandards/metastandards/index.htm>

Dublin Core Metadata Element Set, Ver 1.1 Reference Description
<http://dublincore.org/documents/dces>

Dublin Core Metadata Initiative – Type Vocabulary

<http://wwdublincore.org/documents/dcmi-type-vocabulary>

Web Content

Library of Congress; Manuscript Digitization Demonstration Project

Final Report (October 1998)

<http://lcweb2.loc.gov/ammem/pictel/index.html>

FDR Library Digital Archives Research Center <http://www.fdrlibrary.marist.edu/research.html>

Immigration History Research Center - COLLAGE Project

<http://www.umn.edu/ihrccollage.htm>

UNESCO/IFLA Directory of Digitized Collections <http://thoth.bl.uk/ddc/index.html>

Navigational Aids for the History of Science, Technology and the Environment (NAHSTE)

<http://www.nahste.ac.uk/>

RLG Primary Sources Program

<http://www.rlg.org/primary/index.html>

Scottish Archive Network - Virtual Vault

<http://www.scan.org.uk>

Evidence! Europe Reflected in Archives

<http://www.euarchives.org/>

France - "Les archives de guerre"

<http://www.ina.fr/Archives/Guerre/index.fr.html>

Germany - "Kaiser Heinrichs Romfahrt" <http://www.landeshauptarchiv.de/ausstellung/>

Hungary - "Forced Labour Camps"

<http://www.osa.ceu.hu/gulag/>

Library and Archival Exhibitions on the Web

<http://www.sil.si.edu/SILPublications/Online-Exhibitions/>

"Virtual Museum" of the Public Records Office

<http://learningcurve.pro.gov.uk/virtualmuseum/default.htm>

Northern Ireland - "19th Century Emigration to the North Americas"

<http://proni.nics.gov.uk/exhibiti/exhibiti.htm>

Northern Ireland - "PRONI on the Record"

<http://proni.nics.gov.uk/exhibiti/ontherec/ontherec.htm>

Northern Ireland - 'From North to South: Northern Ireland's contribution to the development of Australia

<http://proni.nics.gov.uk/exhibiti/austral/contents.htm>

United Nations - A look at Historic Documents <http://www.un.org/Depts/archives/exhibit.html>

USA - Digital Memories

<http://www.lib.uidaho.edu/special-collections/dm/dgtlcurr.htm>

USA - IIT Archives - Reflection and Revelation: Yearbooks Through the Years

http://archives.iit.edu/exhibits/1_yearbook/

USA - Louisiana State Archives - Exhibitions

<http://www.sec.state.la.us/archives/archives/archives-events.htm>

Ireland - National Archives

<http://www.nationalarchives.ie/>

Denmark - Statens Arkiver

<http://www.sa.dk/>

Estonia - National Archives - Rahvusarhiiv

<http://www.ra.ee/>